

In-situ fatigue damage monitoring using symbolic dynamic filtering of ultrasonic signals

D S Singh, S Gupta, and A Ray*

Department of Mechanical Engineering, Pennsylvania State University, University Park, Pennsylvania, USA

The manuscript was received on 13 February 2009 and was accepted after revision for publication on 17 April 2009.

DOI: 10.1243/09544100JAERO530

Abstract: This article presents a data-driven method of pattern identification for *in-situ* monitoring of fatigue damage in polycrystalline alloys that are commonly used in aerospace structures. The concept is built upon analytic signal space partitioning of ultrasonic data sequences for symbolic dynamic filtering of the underlying information. The statistical patterns of evolving damage are generated for real-time monitoring of the possible structural degradation under fatigue load. The proposed method is capable of detecting small anomalies (i.e. deviations from the nominal condition) in the material microstructure and thereby generating early warnings on damage initiation. The damage monitoring algorithm has been validated on time series data of ultrasonic sensors from a fatigue test apparatus, where the behavioural pattern changes accrue because of the evolving fatigue damage in polycrystalline alloys.

Keywords: aircraft structures, fatigue damage monitorings, statistical pattern recognition

1 INTRODUCTION

Reliability of aircraft structures deteriorates because of gradual evolution of damage over a prolonged period of operation. Fatigue damage is one of the most commonly encountered sources of structural degradation in aerospace structures that are often made of polycrystalline alloys. Many model-based techniques have been reported in recent literature related to structural health monitoring [1]. Apparently, existing models cannot adequately capture the dynamic behaviour of fatigue damage at the grain level, solely based on the fundamental principles of physics [2]. Therefore, it is necessary to develop a sensor-based (i.e. data-driven) damage detection and estimation scheme for real-time damage monitoring and to make appropriate decisions in collaboration with the available models of damage growth.

Several techniques have been proposed in recent literature for fatigue damage monitoring based on various sensing devices (e.g. ultrasonics, acoustic emission, and eddy currents) [3, 4]. The traditional

analysis methods using an acoustic emission technique are used to correlate the signal parameters (such as the acoustic-emission counts, the peak amplitudes and the energy) with the defect formation mechanisms and to provide a quantified estimate of faults [5, 6]. However, the major drawback of acoustic emission technique is poor performance in noisy environments where signal–noise separation becomes a difficult task. Similarly, the presence of faults in the material affect the eddy current flow patterns, which can be detected for prediction and estimation of the structural damage [7–9]. The advantages of eddy current inspection technique include sensitivity to small cracks, portability of sensor equipment, minimum part preparation, and non-contact evaluation. However, the major limitation of the eddy current inspection technique is the depth of penetration; therefore, it can be used to detect only surface and near-surface defects.

Recent studies have shown that ultrasonic signals can be used to capture some of the minute defects and small changes during the early stages of fatigue damage, which may not be possible to detect by an optical microscope [10–12]. Since the evolution of material microstructural characteristics (e.g. dislocations and short cracks) influence the ultrasonic impedance, a small fault in the specimen is likely to change the signature of the ultrasonic signal at the

*Corresponding author: Department of Mechanical Engineering, Pennsylvania State University, 329 Reber Building, University Park, Pennsylvania 16802, USA.
email: axr2@psu.edu

receiver end [10, 13]. The long-wavelength (Rayleigh) scattering of ultrasonic pulse propagating at the surface of a sample under cycling loading has been used to monitor small surface breaking crack [14].

A statistical pattern identification method, called symbolic dynamic filtering (SDF) [15, 16], has been recently formulated for anomaly detection in complex systems. The SDF method is a pattern identification tool, based on the concepts of statistical mechanics [17, 18], and is built upon a fixed-structure, fixed-order Markov chain; it has been validated by comparison with existing pattern recognition techniques such as principal component analysis and artificial neural networks [10]. In this method, time series data of sensors are converted from the time domain to quasi-stationary symbolic sequences by symbolic dynamic encoding [19]. This method of behavioural pattern identification has been used to detect statistical changes in the ultrasonic data generated from a fatigue test apparatus for damage monitoring in polycrystalline alloys. Benefits of the proposed technique for damage monitoring in aircraft structures are summarized below.

1. *In-situ* measurement, detection, and monitoring of fatigue damage.
2. Early detection of fatigue damage for condition-based maintenance [20].
3. Capability for generation of timely information obtained for life extending control [21].

The article is organized into five sections including the present section. Section 2.2 presents the underlying concepts of SDF for identification of behavioural patterns due to slowly evolving fatigue damage in mechanical structures. Section 4 describes the test apparatus that serves as a platform for experimental validation of the proposed damage monitoring method, and section 5 presents the results of experimental validation. The article is concluded in section 6 along with recommendations for future research.

2 REVIEW OF SYMBOLIC DYNAMIC FILTERING

This section presents the underlying concepts of SDF [15–18] that is used to identify statistical patterns in the ultrasonic signals, that represents gradually evolving fatigue damage in mechanical structures. Specifically, the theories of analytic signal space partitioning (ASSP) [22] and construction of a probabilistic finite-state machine (PFSM) [15] are briefly reviewed.

2.1 Two-time-scale formulation for damage monitoring

The pattern identification of a quasi-stationary process is recognized as a two-time-scale problem. The fast-time scale refers to the local behaviour of the

system and is defined as the time scale over which the dynamical behaviour of the system is assumed to remain invariant, i.e. the process has stationary dynamics. The slow-time scale, on the other hand, refers to the long-term behaviour of the system, where the patterns of the process dynamics might deviate from those under the nominal conditions. It is assumed that any observable non-stationary behaviour pattern is associated with changes occurring on the slow-time scale. In general, a long time span in the fast-time scale is a tiny (i.e. several orders of magnitude smaller) interval in the slow-time scale. For example, evolution of fatigue damage in structural materials (causing a detectable change in the dynamics of the system) occurs on the slow-time scale (possibly in the order of months or years); fatigue damage behaviour is essentially invariant on the fast-time scale (approximately in the order of seconds or minutes). Time series data on the fast-time scale are collected at different slow-time scale epochs, simply referred to as epochs in the sequel, to identify the evolving behavioural patterns. The concept of two time scales is illustrated in Fig. 1.

2.2 Symbolic dynamic encoding and pattern identification

Let Ω be a compact (i.e. closed and bounded) region in the (finite-dimensional) phase space, which contains the trajectory of the dynamical system, as illustrated in Fig. 2. The region Ω is partitioned as $\{\Phi_1, \dots, \Phi_{|\Sigma|}\}$ consisting of $|\Sigma|$ mutually exclusive (i.e. $\Phi_j \cap \Phi_k = \emptyset \forall j \neq k$) and exhaustive (i.e. $\bigcup_{j=1}^{|\Sigma|} \Phi_j = \Omega$) cells, where Σ is the symbol alphabet that labels the partition cells. A trajectory of the dynamical system is described by the discrete time series data as $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\}$, where each $\mathbf{x}_i \in \Omega$. The trajectory passes through or touches one of the cells of the partition; accordingly, the corresponding symbol is assigned to each point \mathbf{x}_i of the trajectory as defined by the mapping $\mathcal{M} : \Omega \rightarrow \Sigma$. Therefore, a sequence of symbols is generated from the trajectory starting from an initial state $\mathbf{x}_0 \in \Omega$, such that

$$\mathbf{x}_0 \mapsto s_0 s_1 s_2, \dots, s_j, \dots \quad (1)$$

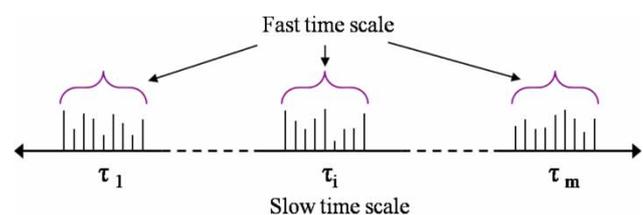


Fig. 1 Pictorial view of the two time scales: (a) slow time scale of anomaly evolution and (b) fast time scale for data acquisition and signal conditioning

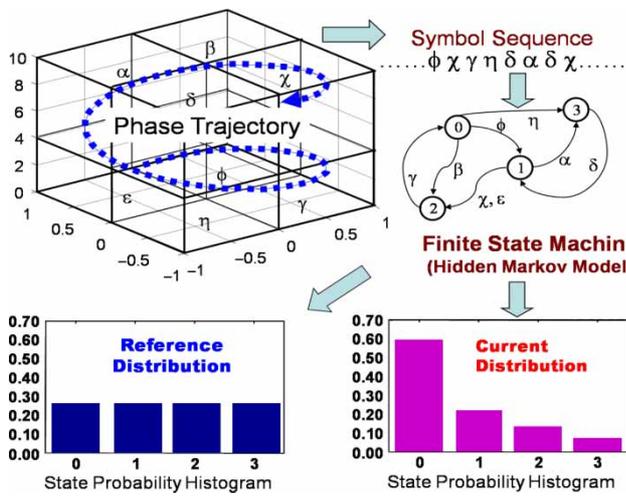


Fig. 2 Concept of phase space partitioning for symbol sequence generation

where $s_k \triangleq \mathcal{M}(x_k)$ is the symbol generated at the (fast scale) instant k . The symbols s_k , $k = 0, 1, \dots$ are identified by an index set $\mathcal{I} : \mathbb{Z} \rightarrow \{0, 1, 2, \dots, |\Sigma| - 1\}$, i.e. $\mathcal{I}(k) = i_k$ and $s_k = \sigma_{i_k}$ where $\sigma_{i_k} \in \Sigma$. Equivalently, equation (1) is expressed as

$$\mathbf{x}_0 \mapsto \sigma_{i_0} \sigma_{i_1} \sigma_{i_2} \dots \sigma_{i_j} \dots \quad (2)$$

Figure 2 pictorially elucidates the concepts of partitioning a finite region of the phase space and the mapping from the partitioned space into the symbol alphabet, where the symbols are indicated by Greek letters (e.g. $\alpha, \beta, \gamma, \delta, \dots$). This represents a spatial and temporal discretization of the system dynamics defined by the trajectories.

The partitioning of phase space for symbol sequence generation is a crucial step in SDF. This article has used ASSP, which is described in section 2.3. Once the partitioning is done with total alphabet size $|\Sigma|$ at the nominal condition (epoch t_0), it is kept constant for all subsequent epochs $\{t_1, t_2, \dots, t_k, \dots\}$, i.e. the structure of the partition is fixed at the nominal condition.

Figure 2 also shows conversion of the symbol sequence into a finite-state machine and generation of the state probability vectors at the current and the reference conditions. The states of the finite-state machine and the histograms in Fig. 2 are indicated by numerics (i.e. 1, 2, 3, and 4); the necessary details are provided in section 2.4. State probability vectors $\mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^k \dots$ are obtained at slow time epochs $t_1, t_2, \dots, t_k \dots$, based on the respective time series data. The state probability vectors are representative of the damage evolving in the system. The quantification of damage accumulation relative to the nominal condition is obtained from a scalar valued function, called damage measure ψ (section 3).

2.3 Analytic signal space partitioning

Symbol sequence generation is a crucial step in SDF that requires partitioning of the phase space of the dynamical system. Several partitioning techniques have been proposed in the literature for symbol sequence generation. Examples are symbolic false nearest neighbour partitioning (SFNNP) [23], wavelet space partitioning (WSP) [24], and ASSP [22]. A major shortcoming of SFNNP is that it may become extremely computation intensive if the dimension of the phase space of the underlying dynamical system is large. Furthermore, if the time series data are noise corrupted, the symbolic false neighbours rapidly grow in number and may erroneously require a large symbol alphabet to capture pertinent information on the system dynamics; WSP was introduced as an alternative to SFNNP as the wavelet transform [25] largely alleviates the above-mentioned shortcomings of SFNNP and is particularly effective with noisy data from high-dimensional dynamical systems. Although significantly computationally faster than SFNNP and thus suitable for real-time applications, WSP has several shortcomings (e.g. selection of a wavelet basis function and identification of appropriate scales). ASSP [22] overcomes the shortcomings of WSP; therefore, ASSP is more appropriate than SFNNP and WSP for online applications.

This article makes use of ASSP for symbol generation as needed in SDF-based damage monitoring in mechanical structures under fatigue load. The role of ASSP is to capture the relevant statistical information for damage monitoring in real-time. The underlying concept of ASSP is built upon the Hilbert transform of observed real-valued data into the corresponding complex-valued analytic signal. The time-dependent analytic signal is represented as a trajectory in the two-dimensional pseudo-phase space, which is partitioned to generate a symbol sequence. The symbol sequences obtained at different time epochs represent evolution of the dynamical behaviour that is identified as damage patterns.

Hilbert transform \mathcal{H} of a real-valued signal $x(t)$ is defined as

$$\tilde{x}(t) = (\mathcal{H}x)(t) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{x(\tau)}{t - \tau} d\tau \quad (3)$$

That is, $\tilde{x}(t)$ is the convolution of $x(t)$ with $1/\pi t$ over the real field \mathbb{R} , whose Fourier transform is given as

$$(\mathcal{F}\tilde{x})(\xi) = -i \operatorname{sgn}(\xi) (\mathcal{F}x)(\xi) \quad (4)$$

$$\text{where } \operatorname{sgn}(\xi) = \begin{cases} +1 & \text{if } \xi > 0 \\ -1 & \text{if } \xi < 0 \end{cases}$$

Given the Hilbert transform of a real-valued signal $x(t)$, the corresponding complex-valued analytic

signal \mathcal{A} is defined as

$$(\mathcal{A}x)(t) = x(t) + i\tilde{x}(t) \quad (5)$$

The construction of equation (5) is based on the fact that the values of Fourier transform of a real-valued function at negative frequencies are redundant due to their Hermitian symmetry imposed by the transform. Thus, the phase of the Hilbert transform $\tilde{x}(t)$ is in quadrature to the phase of $x(t)$. That is, the analytic signal can be expressed as

$$(\mathcal{A}x)(t) = a_x(t) \exp[i\varphi_x(t)] \quad (6)$$

where $a_x(t)$ and $\varphi_x(t)$ are called the instantaneous amplitude and instantaneous phase of $(\mathcal{A}x)(t)$, respectively. The amplitude and phase of an analytic signal satisfy the following three physical properties.

1. Amplitude continuity: a small perturbation in $x(t)$ induces a small change in $a_x(t)$.
2. Phase independence of scale: scaling $x(t)$ by a constant $c > 0$ has no effects on $\varphi_x(t)$ and multiplies $a_x(t)$ by c .
3. Harmonic correspondence: a mono-frequency signal (i.e. a pure sinusoid $a_0 \cos(\omega_0 t + \varphi_0)$) yields $a_x(t) = a_0$ and $\varphi_x(t) = \omega_0 t + \varphi_0$ for all t .

Thus, for a mono-frequency signal, which is embedded in a two-dimensional (2D) state space, a direct parallel can be drawn between the phase plot and the Hilbert transform plot. The procedure for ASSP is formulated next. Let the observed signal be available as a real-valued time series of N data points. Upon Hilbert transformation of this data sequence, a pseudo-phase plot is constructed from the resulting analytic signal by a bijective mapping of the complex field onto \mathbb{R}^2 , i.e. by plotting the real and the imaginary parts of the analytic signal on the x_1 and x_2 axes, respectively. It is important to note that the pseudo-phase space is always 2D, whereas the phase space of the dynamical system is a representation of the n -dimensional manifold, where n could be an arbitrarily large positive integer.

The time-dependent analytic signal in equation (5) is now represented as a (one-dimensional) trajectory in the 2D pseudo-phase space. Let Ξ be a compact region in the pseudo-phase space, which encloses the trajectory. The objective is to partition Ξ into finitely many mutually exclusive and exhaustive segments, where each segment is labelled with a symbol or letter. The segments are conveniently determined by the magnitude and phase of the analytic signal as well as based on the density of data points in these segments, i.e. if the magnitude and phase of a data point of the analytic signal lie within a segment or on its boundary, then the data point is labelled with the corresponding symbol. Thus, a symbol sequence is derived from the

(complex-valued) sequence of the analytic signal. The set of (finitely many) symbols is called the alphabet Σ .

The analytic signal is partitioned at nominal condition using the principles of maximum entropy partitioning [24], which maximizes the entropy of the partition that is characterized by the alphabet size $|\Sigma|$, thereby imposing a uniform probability distribution on the symbols. The maximum entropy partitioning is generated by maximizing the Shannon entropy, which is defined as

$$S = - \sum_{i=0}^{|\Sigma|-1} p_i \log(p_i) \quad (7)$$

where p_i is the probability of a data point to be in the i th partition segment. In this partitioning, regions with rich information are partitioned into finer segments than those with sparse information. Computationally, the maximum entropy partition can be obtained by sorting the data sequence in an ascending order. This sorted data sequence is then partitioned into $|\Sigma|$ equal segments of length $\lfloor N/|\Sigma| \rfloor$, where N is the length of the data sequence and $\lfloor x \rfloor$ is the greatest integer $\leq x$. Each of these segments is assigned a symbol and all data points in a given segment are assigned the corresponding symbol.

The magnitude and phase of the analytic signal in equation (5) are partitioned separately according to uniform partitioning, maximum entropy partitioning, or any other type of partitioning; the type of partitioning may depend on the characteristics of the physical process. In essence, each point in the data set is represented by a pair of symbols – one belonging to the alphabet Σ_R based on the magnitude (i.e. in the radial direction) and the other belonging to the alphabet Σ_A based on the phase (i.e. in the angular direction). The analytic signal is converted into a one-dimensional symbol sequence p by associating each pair of symbols into a symbol from a new alphabet Σ as

$$\Sigma \triangleq \{(\sigma_i, \sigma_j) : \sigma_i \in \Sigma_R, \sigma_j \in \Sigma_A\} \text{ and } |\Sigma| = |\Sigma_R| \cdot |\Sigma_A|$$

2.4 Construction of a probabilistic finite-state machine

Once the symbol sequence is obtained, the next step is the construction of a PFMSM and calculation of the respective state probability vector as depicted in the lower part of Fig. 2 by the histograms. The partitioning is performed at the reference condition.

A PFMSM is then constructed, where the states of the machine are defined corresponding to a given alphabet set Σ and window length D . The alphabet size $|\Sigma|$ is the total number of partition segments, while the window length D is the length of consecutive symbol words [15], which are chosen as all possible words of length D from the symbol sequence. Each state belongs to an equivalence class of symbol words of

length D , which is characterized by a word of length D at the leading edge. Therefore, the number n of such equivalence classes (i.e. states) is less than or equal to the total permutations of the alphabet symbols within words of length D . That is, $n \leq |\Sigma|^D$; some of the states may be forbidden, i.e. these states have zero probability of occurrence. For example, if $\Sigma = \{\alpha, \beta\}$, i.e. $|\Sigma| = 2$ and if $D = 2$, then the number of states is $n \leq |\Sigma|^D = 4$, and the possible states are words of length $D = 2$, i.e. $\alpha\alpha, \alpha\beta, \beta\alpha$, and $\beta\beta$, as shown in Fig. 3.

The choice of $|\Sigma|$ and D depends on specific applications and the noise level in the time series data as well as on the available computation power and memory availability. As stated earlier, a large alphabet may be noise sensitive and a small alphabet could miss the details of signal dynamics. Similarly, while a larger value of D is more sensitive to signal distortion, it would create a much larger number of states requiring more computation power and increased length of the data sets. In the results section of this article, the analysis of time series data sets is done using the window length equal to $D = 1$; consequently, the set of states of finite-state machine is equivalent to the symbol alphabet Σ . For ASSP with the selection of the parameters $|\Sigma_R| = 6$ and $|\Sigma_A| = 3$, the PFSM has $n = 18$ states if $D = 1$. With this choice of parameters, the SDF algorithm is shown to be capable of detecting fatigue damage evolution. However, other applications, such as 2D image processing, may require larger values of the parameter D and hence possibly larger number of states in the PFSM.

Using the symbol sequence generated from the time series data, the state machine is constructed on the principle of sliding block codes [19]. The window of length D on a symbol sequence is shifted to the right by one symbol, such that it retains the most recent $(D - 1)$ symbols of the previous state and appends it with the new symbol at the extreme right. The symbolic permutation in the current window gives rise to a new state. The PFSM constructed in this fashion is called the D -Markov machine [15], because of its Markov properties.

Definition 1

A symbolic stationary process is called D -Markov if the probability of the next symbol depends only on the

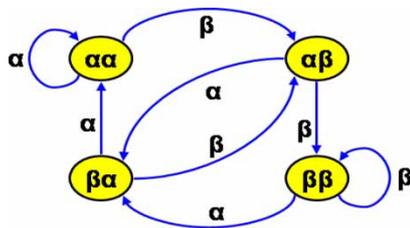


Fig. 3 Example of finite-state machine with $D = 2$ and $\Sigma = \{\alpha, \beta\}$

previous D symbols, i.e. $P(s_j|s_{j-1} \dots, s_{j-D}s_{j-D-1} \dots) = P(s_j|s_{j-1} \dots, s_{j-D})$.

The finite-state machine constructed above has D -Markov properties because the probability of occurrence of symbol $\sigma \in \Sigma$ on a particular state depends only on the configuration of that state, i.e. the previous D symbols. The states of the machine are marked with the corresponding symbolic word permutation and the edges joining the states indicate the occurrence of a symbol σ . The occurrence of a symbol at a state may keep the machine in the same state or move it to a new state.

Definition 2

Let Ξ be the set of all states of the finite-state machine. Then, the probability of occurrence of symbols that cause a transition from state ξ_j to state ξ_k under the mapping $\delta : \Xi \times \Sigma \rightarrow \Xi$ is defined as

$$\pi_{jk} = P(\sigma \in \Sigma \mid \delta(\xi_j, \sigma) \rightarrow \xi_k); \quad \sum_k \pi_{jk} = 1 \quad (8)$$

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 definition above is equivalent to an alternative representation such that

$$\pi_{jk} \equiv P(\xi_k|\xi_j) = \frac{P(\xi_j, \xi_k)}{P(\xi_j)} = \frac{P(\sigma_{i_0} \dots \sigma_{i_{D-1}} \sigma_{i_D})}{P(\sigma_{i_0} \dots \sigma_{i_{D-1}})} \quad (9)$$

where the corresponding states are denoted by $\xi_j \equiv \sigma_{i_0} \dots \sigma_{i_{D-1}}$ and $\xi_k \equiv \sigma_{i_1} \dots \sigma_{i_D}$. This phenomenon is a consequence of the PFSM construction based on the principle of sliding block codes described above, where the occurrence of a new symbol causes a transition to another state or possibly the same state.

For computation of the state transition probabilities from a given symbol sequence at a particular slow time epoch, a D -block (i.e. a window of length D) is moved by counting occurrences of symbol blocks $\sigma_{i_0} \dots \sigma_{i_{D-1}} \sigma_{i_D}$ and $\sigma_{i_0} \dots \sigma_{i_{D-1}}$, which are, respectively, denoted by $N(\sigma_{i_0} \dots \sigma_{i_{D-1}} \sigma_{i_D})$ and $N(\sigma_{i_0} \dots \sigma_{i_{D-1}})$. Note that if $N(\sigma_{i_0} \dots \sigma_{i_{D-1}}) = 0$, then the state $\sigma_{i_0} \dots \sigma_{i_{D-1}} \in \Xi$ has zero probability of occurrence. For $N(\sigma_{i_0} \dots \sigma_{i_{D-1}}) \neq 0$, the estimates of the transitions probabilities are then obtained by these frequency counts as follows

$$\pi_{jk} \approx \frac{N(\sigma_{i_0} \dots \sigma_{i_{D-1}} \sigma_{i_D})}{N(\sigma_{i_0} \dots \sigma_{i_{D-1}})} \quad (10)$$

The symbol sequence generated from the time series data at the reference condition, set as a benchmark, is used to compute the state transition matrix Π^0 using equation (10). The left eigenvector \mathbf{p}^0 corresponding to the unique unit eigenvalue of the irreducible stochastic matrix Π^0 is the probability vector whose elements

are the stationary probabilities of the states belonging to Ξ [15]. Similarly, the state probability vector \mathbf{p}^k is obtained from time series data at a (possibly) anomalous condition. The partitioning of time series data and the state machine structure should be the same in both cases but the respective state transition matrices could be different. The probability vectors \mathbf{p}^0 and \mathbf{p}^k are estimates of the respective true probability vectors and are treated as statistical patterns.

Behavioural pattern changes, if any, are detected at time epoch t_k with respect to the reference condition at time epoch t_0 by defining a scalar-valued damage measure ψ , such that

$$\psi^k = f(\mathbf{p}^k, \mathbf{p}^{k-1}, \dots, \mathbf{p}^0) \quad (11)$$

where $f(\bullet, \bullet, \dots, \bullet)$ is a path function of the evolution of the pattern vector \mathbf{p} on the slow scale; further details are provided in section 3. Another choice of ψ is a distance function $d(\mathbf{p}^k, \mathbf{p}^0)$ (e.g. the standard Euclidean norm).

2.5 Summary of the SDF Method

1. Time series data acquisition from appropriate sensor(s) at the epoch t_0 , i.e. the nominal condition, when the system is assumed to be in the healthy state (i.e. zero damage measure).
2. Generation of the analytic signal via the Hilbert transform of the (real-valued) time series data.
3. Maximum entropy partitioning in the domain of (complex-valued) analytic signal at the nominal condition and generation of the corresponding symbol sequence. The partitioning is fixed for subsequent time epochs.
4. Construction of the PFSM and generation of the state probability vector \mathbf{p}^0 at time epoch t_0 . Note that \mathbf{p}^0 is a uniform distribution because of maximum entropy partitioning.
5. Collection of time series data at epochs $t_1, t_2, \dots, t_k, \dots$, and their conversion to analytic signals to generate respective symbolic sequences based on the partitioning at time epoch t_0 .
6. Calculation of the corresponding probability vectors $\mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^k, \dots$, at epochs $t_1, t_2, \dots, t_k, \dots$, from the respective symbolic sequences using the finite-state machine constructed at time epoch t_0 .
7. Computation of the scalar damage measures $\psi^1, \psi^2, \dots, \psi^k, \dots$, at epochs $t_1, t_2, \dots, t_k, \dots$.

Major advantages of SDF for damage detection are listed below:

- (a) robustness to measurement noise and spurious disturbances [10];
- (b) adaptability to low-resolution sensing because of coarse graining in symbolization [15];
- (c) capability for early detection of damage because of sensitivity to signal distortion;

- (d) real-time execution on commercially available inexpensive platforms [10].

3 DAMAGE EVOLUTION AND MONITORING ALGORITHM

The pattern changes are quantified as deviations from the nominal behaviour (i.e. the probability distribution at the nominal condition). The resulting anomalies (i.e. deviations of the evolving patterns from the nominal pattern) are characterized by a scalar-valued function, called damage measure ψ that is quasi-static in the fast-time scale and is monotonically non-decreasing in the slow-time scale. The state probability vector at any epoch corresponds to a singleton point on the unity-radius hypersphere, which is a rhomboid in the topology of sum-normalization (i.e. ℓ^1 -norm [26]). Since all elements of a probability vector are positive, the movements of the tip of the probability vector are restricted on a simplex plane. During fatigue damage evolution, the tip of the state probability vector moves along a path on the surface of this hypersphere. The initial starting point of the path is the state probability vector \mathbf{p}^0 with uniform distribution because of maximum entropy partitioning [24]. As the damage progresses, the probability distribution changes; eventually when a very large crack is formed, complete attenuation of the time series of ultrasonic data occurs (see section 4) and consequently the tip of the probability vector reaches a point where all states have zero probabilities of occurrence except the ones that correspond to least energy states. In the context of an irreversible process, such as fatigue crack growth phenomena, the damage measure is computed based on the following assumptions.

1. Assumption 1: Damage evolution is an irreversible process (i.e. with zero probability of self healing) and implies the following conditions

$$\psi(t) \geq 0, \quad \psi(t + \delta) - \psi(t) \geq 0 \quad \forall t \geq t_0 \quad \forall \delta > 0 \quad (12)$$

2. Assumption 2: Damage accumulation between two epochs is a path function, i.e. dependent on the path traversed to reach the target state from the initial state.

The damage measure is derived in terms of the following distance function between the state probability vectors at two epochs

$$d(\mathbf{p}, \tilde{\mathbf{p}}) = \sqrt{(\mathbf{p} - \tilde{\mathbf{p}})^T (\mathbf{p} - \tilde{\mathbf{p}})} \quad (13)$$

A real positive parameter ϵ , associated with robustness of the damage measure ψ for compensation of measurement and computation noise, is identified

by performing an experiment with a sample with no notch, where \mathbf{q}^k is the state probability vector at the k th observation. In the absence of any notch, there is practically no stress augmentation and therefore negligible fatigue damage increment. Any difference between the state probability vectors at consecutive observations is attributed to noise only. Accordingly, the parameter ϵ is estimated as

$$\epsilon \approx \left(\frac{1}{N} \sum_{k=1}^N d^2(\mathbf{q}^{k+1}, \mathbf{q}^k) \right)^{1/2} \quad (14)$$

from N consecutive observations with $N \gg 1$.

The algorithm for computation of the damage measure ψ is presented below:

- (a) $\psi^0 = 0, \delta\psi^1 = 0, \tilde{\mathbf{p}} = \mathbf{p}^0, k = 1$;
- (b) if $d(\mathbf{p}^k, \tilde{\mathbf{p}}) > \epsilon$, then $\delta\psi^k = \sqrt{d^2(\mathbf{p}^k, \tilde{\mathbf{p}}) - \epsilon^2}$ and $\tilde{\mathbf{p}} = \mathbf{p}^k$;
- (c) $\psi^k = \psi^{k-1} + \delta\psi^k$;
- (d) $k \leftarrow k + 1, \delta\psi^k = 0$, go to step (b).

The above algorithm is executed in the following way. The reference point $\tilde{\mathbf{p}}$ is initialized to the starting point \mathbf{p}^0 and the initial damage measure ψ^0 is set to 0. At an epoch t_k , if the state probability vector changes such that, if the distance $d(\mathbf{p}^k, \tilde{\mathbf{p}})$ is $> \epsilon$ as specified in step (b), then the damage measure is incremented by the amount $\sqrt{d^2(\mathbf{p}^k, \tilde{\mathbf{p}}) - \epsilon^2}$ because the signal of damage increment and the noise are assumed to be statistically independent. Consequently, the reference point is shifted to the current point \mathbf{p}^k . The procedure is repeated at subsequent epochs and the path length travelled by the tip of the state probability vector represents the accumulated damage.

4 EXPERIMENTAL VALIDATION

The damage monitoring algorithm has been validated on time series data of ultrasonic sensors from a fatigue test apparatus [13], where the behavioural pattern changes accrue because of the evolving fatigue damage in polycrystalline alloys.

The experimental apparatus, shown in Fig. 4, is a special-purpose uniaxial fatigue testing machine [13], which is operated under load control or strain control at speeds up to 12.5 Hz. A typical specimen, made of 7075-T6 aluminium alloy, is shown in Fig. 5. The specimen is 3 mm thick and 50 mm wide with a slot on one side of 1.58 mm diameter and 4.57 mm length. The notch is made to increase the stress concentration factor in that region and it guarantees crack propagation at the notch end. The fatigue test apparatus is equipped with an ultrasonic flaw detector and a travelling optical microscope for fault monitoring.

The travelling optical microscope, shown as part of the test apparatus in Fig. 4, provides direct measurements of the visible portion of a crack. The resolution

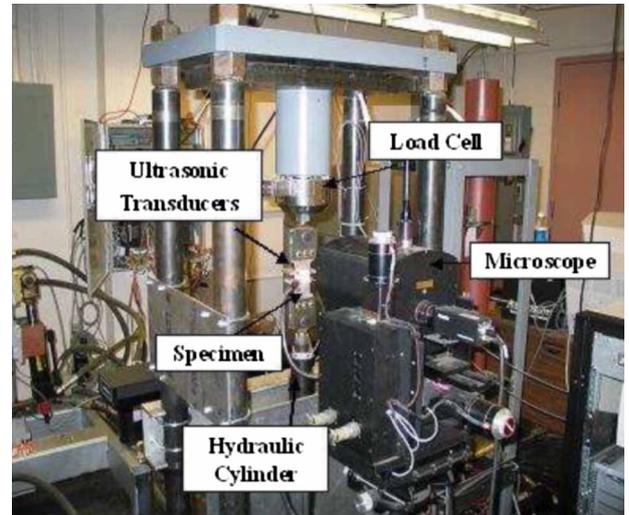


Fig. 4 Special-purpose fatigue test apparatus

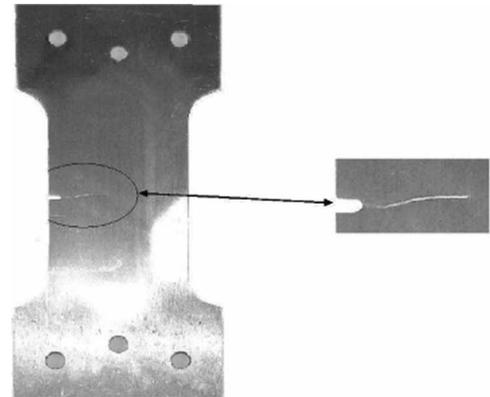


Fig. 5 Cracked specimen with a side notch

of the optical microscope is about $2 \mu\text{m}$ at a working distance of 10–35 cm and the images are taken at a magnification of $75\times$. The microscope is used to monitor and measure crack length in the crack propagation stage. Images of the specimen are taken by the microscope at constant intervals. However, small cracks are not detectable from the microscope. This article analyses ultrasonic data to detect the small cracks or the changes in material characteristics before the formation of a single large crack that appears on the surface.

The ultrasonic flaw detector functions by emitting high-frequency (i.e. 5 MHz) sinusoidal ultrasonic waves that travel through the specimen and return back through the receiver transducers. The signal is sent through the region of crack propagation and received on the other side. Two angle beam transducers are placed at a distance of 10 mm from each side of the notch, as shown in Fig. 6. The transducers are glued to the specimen surface with a thin layer of grease and a clamp is used to maintain a constant pressure between the transducer and the specimen surface. The

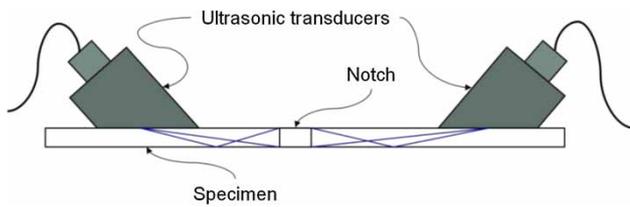


Fig. 6 Ultrasonic flaw detection scheme

transducers are made of piezoelectric material that sends refracted shear wave to the specimen at an angle of 45° . Ultrasonic pulsar is synchronized with the load cycle frequency (i.e. 12.5 Hz) such that it generates one pulse at peak stress in each load cycle, which enables the transducer to send a 5 MHz sine wave through the specimen. The ultrasonic transducer at the other side of the notch receives the transmitted ultrasonic wave and sends it to the data acquisition board. The signal is recorded at 50 MHz sampling frequency. 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. A large number of internal defects develop before they coalesce to form a single crack, which leads to the transition from the crack initiation phase to the crack propagation phase. In the crack initiation phase, plastically strained zones that consist of several small internal defects are formed. A zone whose diameter is greater than half of the ultrasonic wavelength is very likely to interact with the ultrasonic wave; however, the average size of the individual defects in this zone could be an order of magnitude smaller than the zone diameter.

The fatigue tests were conducted at a constant amplitude sinusoidal load for low-cycle fatigue, where the maximum and minimum loads were kept constant at 87 MPa and 4.85 MPa, respectively. 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 caused by multiple small cracks, dislocations, and microstructural defects alters the ultrasonic impedance, which results in signal distortion and attenuation at the receiver end. The crack propagation stage starts when this internal damage eventually develops into a single large crack. The slow-time epochs for data analysis were chosen to be ~ 1000 load cycles (i.e. ~ 80 s) apart. At the onset of each slow-time epoch, the ultrasonic data points were collected on the fast-time scale of 50 cycles (i.e. ~ 4 s at 12.5 Hz frequency), which produced a string of 30 000 data points. It is assumed that during this fast-time scale, the system remains in a stationary condition and no major changes occur in the fatigue crack behaviour. The sets of time series data collected in this manner at different epochs were analysed to calculate the damage measures at those epochs.

5 RESULTS AND DISCUSSION

This section presents results of fatigue damage monitoring using SDF of ultrasonic data collected at different slow-time epochs as described in section 4. Following the SDF procedure, the symbol states are obtained by maximum-entropy partitioning the analytic signal, which results in uniform probability distribution at the nominal condition, as shown in Fig. 7(a). The analytic signal space is partitioned into six segments in the radial direction and three segments in the angular direction, which amounts to having cardinality of the symbol alphabet $|\Sigma| = 6 \times 3 = 18$. For $D = 1$, the total number of states is $n = 18$. Figures 7(a) to (f) show the received ultrasonic data and histograms of probability distribution at six different time epochs, exhibiting gradual evolution of fatigue damage. The top plate in each part of Fig. 7 shows a small segment (i.e. 300 points) of the complete set of ultrasonic data containing 30 000 points that are collected at the receiver end at a particular slow-time epoch. This set of 30 000 points data points is used for state probability vector generation using SDF. The corresponding bottom plates show the evolution of the histograms because of fatigue damage growth at different slow time epochs, signifying how the probability distribution gradually changes from uniform distribution at the nominal condition. As seen in Fig. 7, the visual inspection of ultrasonic data does not reveal much information during early stages of fatigue damage; however, the SDF method is able to capture small statistical changes in the signal that are visible in the corresponding histograms. These statistical changes are a result of the interactions of the ultrasonic signal with the internally damaged region of the specimen. As the number of load cycles increases and fatigue damage accumulates, the probability histograms evolve with respect to the nominal condition.

The top plate, as seen in Fig. 7(a), shows the ultrasonic data at the nominal condition, which is chosen after the start of the experiment when electro-hydraulic response of the apparatus has come to a steady state. At this condition, the damage 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). The top plates of Figs 7(b) and (c) at ~ 15 and ~ 22 kilocycles, respectively, do not show conspicuous changes in the ultrasonic profiles by visual inspection; however, close inspection does indicate that some distortions have appeared in the signal. The corresponding bottom plates do exhibit deviations from the uniform probability distribution in both radial and angular directions. This is an evidence that the analytical measurements, based on ultrasonic sensor data, produce damage information during crack initiation.

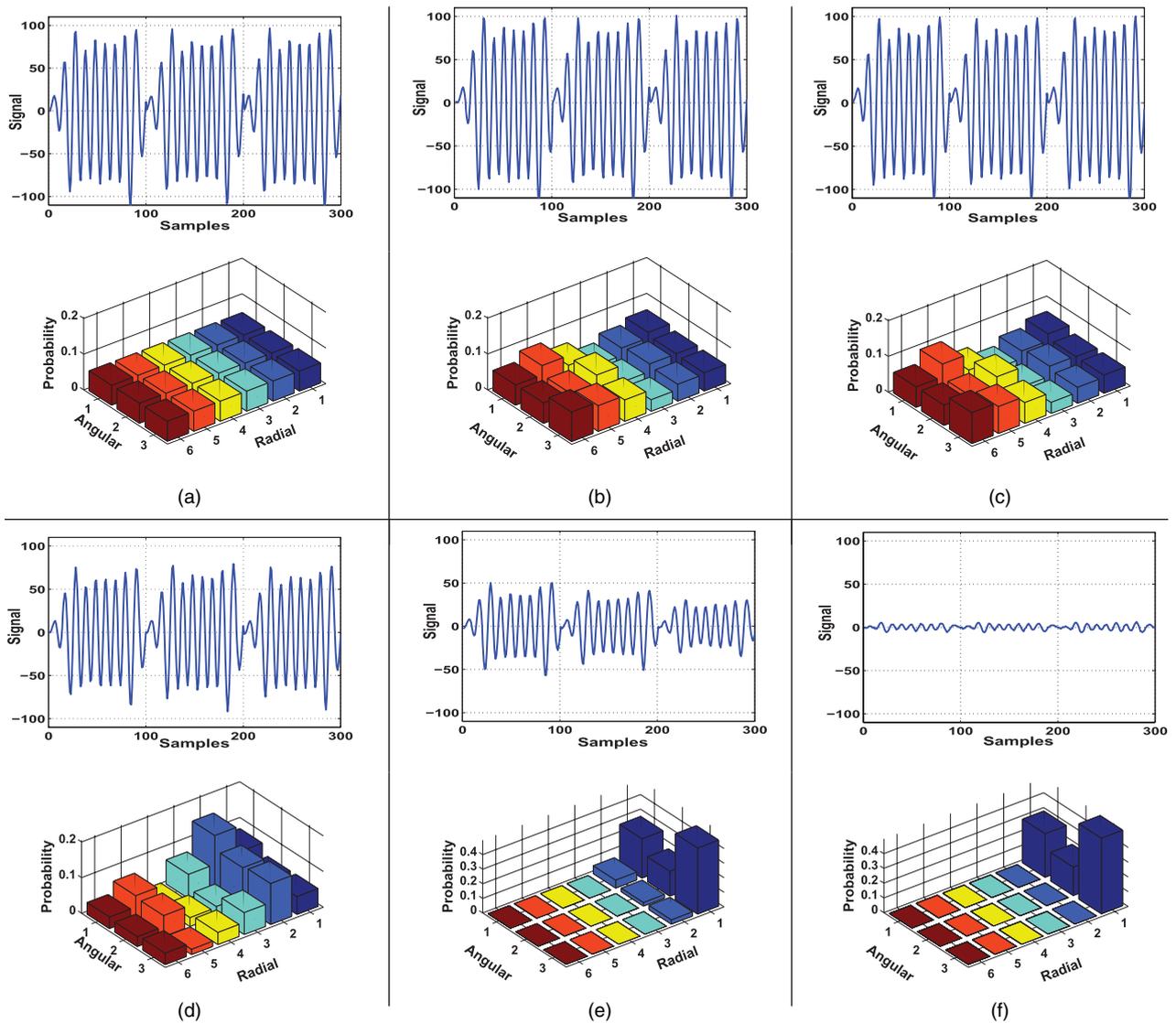


Fig. 7 Received ultrasonic signal and corresponding probability distribution of symbolic states at different epochs: (a) nominal condition after 0 kilocycles; (b) internally damaged after 15 kilocycles; (c) internally Damaged after 22 kilocycles; (d) crack appearance at surface after 32 kilocycles; (e) developed crack after 40 kilocycles; and (f) broken specimen after 45 kilocycles

The top plate of Fig. 7(d) at ~32 kilocycles exhibits noticeable difference in the ultrasonic profile with respect to the nominal condition. This is due to the appearance of ~200 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. The top plate of Fig. 7(e) at ~40 kilocycles exhibits significant distortion of the ultrasonic signal. This is because of the development of a single large crack

in its propagation phase. The corresponding bottom plate shows the histogram of the probability distribution which is significantly different from those in Figs 7(a) to (d). The top plate of Fig. 7(f) at ~45 kilocycles exhibits complete attenuation of the ultrasonic signals because of the development of a significantly large crack. The corresponding bottom plate confirms this finding.

In this article, the fatigue damage phenomenon in polycrystalline alloys is broadly classified into two phases: (a) crack initiation, where the dislocations combine to initiate a crack and (b) crack propagation, where the (possibly widely coalesced) crack is propagated. The crack propagation phase is relatively easier to monitor than the crack initiation phase, because the

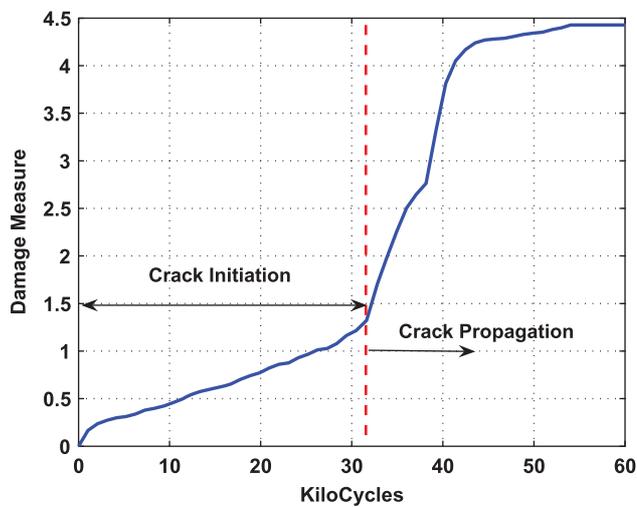


Fig. 8 Evolution of fatigue damage – crack initiation and crack propagation

surface damage could be clearly observed. In contrast, the crack initiation phase may not be detectable easily under an optical microscope; hence, it needs sensor-based monitoring with the aid of efficient and reliable signal processing tools. Figures 7(a) to (c) correspond to the crack initiation phase, (d) and (e) correspond to the crack propagation phase, and (f) corresponds to the fractured specimen. In the crack initiation phase, changes in the ultrasonic signal are mainly due to microstructural changes within the material. Before appearance of optically visible crack the areas at the tip of the notch goes through significant elasto-plastic deformation. The changes in probability distribution are quantified as described in section 3. The damage measure versus the number of load cycles is shown in Fig. 8. The plot is divided into two phases, crack initiation and crack propagation before and after 32 kilocycles, respectively. The damage rate in the crack initiation phase is low because the changes inside the material are at the microstructure level, while the damage rate is much higher in the crack propagation phase. The curve in Fig. 8 shows that damage evolution at an early stage can be detected to enable maintenance and take necessary control actions. It is emphasized 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. However, inferring fatigue damage from the observed anomaly measure is an inverse problem that is a topic of future research.

6 CONCLUSIONS AND FUTURE WORK

This article presents an application of ASSP [22] for SDF-based fatigue damage monitoring in aerospace structures, made of polycrystalline alloys. The damage monitoring method is validated on a fatigue test

apparatus for time series of ultrasonic data from a specimen of 7075-T6 aluminium alloy. The results demonstrate that the proposed method is suitable for early detection of fatigue damage. It is also capable of quantifying the damage rate in real time and of revealing certain key features of the fatigue damage evolution mechanism in polycrystalline materials.

The proposed method is a step towards building a reliable instrumentation and control system for early detection of fatigue damage and real-time estimation of remaining useful fatigue life in aerospace structures. Although there are many other research issues, the following tasks are recommended for future work.

1. Estimation of the remaining life of the structure by formulating the problem in a stochastic setting [11].
2. Development of real-time control strategies for damage mitigation based on the statistical patterns inferred from sensor data by SDF.

ACKNOWLEDGEMENTS

This work has been supported in part by NASA under Cooperative Agreement no. NNX07AK49A and by the US Army Research Office under Grant no. W911NF-07-1-0376. Any opinions, findings and conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

REFERENCES

- 1 Meggiolaro, M. and Castro, J. Statistical evaluation of strain-life fatigue crack initiation predictions. *Int. J. Fatigue*, 2004, **26**, 463–476.
- 2 Pathria, R. K. *Statistical mechanics*, 1996 (Elsevier Science and Technology Books, Amsterdam, The Netherlands).
- 3 Grondel, S., Delebarre, C., Assaad, J., Dupuis, J., and Reithler, L. Fatigue crack monitoring of riveted aluminium strap joints by lamb wave analysis and acoustic emission measurement techniques. *NDT&E Int.*, 2002, **35**, 137–146.
- 4 Zilberstein, V., Walrath, K., Grundy, D., Schlicker, D., Goldfine, N., Abramovici, E., and Yentzer, T. MWM, eddy-current arrays for crack initiation and growth monitoring. *Int. J. Fatigue*, 2003, **25**, 1147–1155.
- 5 Baram, J. Fatigue-life prediction by an order statistics treatment of acoustic-emission signals. *Expl Mech.*, 1993, **33**, 189–194.
- 6 Harris, D. O. and Dunegan, H. L. Continuous monitoring of fatigue-crack growth by acoustic-emission techniques. *Expl Mech.*, 1974, **14**, 71–80.
- 7 Zilberstein, V., Grundy, D., Weiss, V., Goldfine, N., Abramovici, E., Newman, J., and Yentzer, T., Early detection and monitoring of fatigue in high strength steels with MWM-arrays. *Int. J. Fatigue*, 2005, **27**, 1644–1652.

- 8 **Bai, H. S., Yu, L. Y., and He (Ho), J. W.** A monitoring system for contact fatigue crack testing. *NDT Int.*, 1989, **22**(3), 162–167.
- 9 **Yusa, N., Janousek, L., Rebican, M., Chen, Z., Miya, K., N. Chigusa, and Ito, H.** Detection of embedded fatigue cracks in inconel weld overlay and the evaluation of the minimum thickness of the weld overlay using eddy current testing. *Nucl. Eng. Des.*, 2006, **236**(18), 1852–1859.
- 10 **Gupta, S., Ray, A., and Keller, E.** Symbolic time series analysis of ultrasonic data for early detection of fatigue damage. *Mech. Syst. Signal Process.*, 2007, **21**(2), 866–884.
- 11 **Gupta, S. and Ray, A.** Real-time fatigue life estimation in mechanical systems. *Meas. Sci. Technol.*, 2007, **18**(7), 1947–1957.
- 12 **Rokhlin, S. and Kim, J.-Y.** In situ ultrasonic monitoring of surface fatigue crack initiation and growth from surface cavity. *Int. J. Fatigue*, 2003, **25**, 41–49.
- 13 **Gupta, S., Ray, A., and Keller, E.** Online fatigue damage monitoring by ultrasonic measurements: a symbolic dynamics approach. *Int. J. Fatigue*, 2007, **29**(6), 1100–1114.
- 14 **Cook, D. A. and Berthelot, Y. H.** Detection of small surface breaking fatigue cracks in steel using scattering of rayleigh waves. *NDT&E Int.*, 2001, **34**, 483–492.
- 15 **A. Ray,** Symbolic dynamic analysis of complex systems for anomaly detection. *Signal Process.*, 2004, **84**(7), 1115–1130.
- 16 **Gupta, S. and Ray, A.** Symbolic dynamic filtering for data-driven pattern recognition. In *Pattern recognition: theory and application* (Ed. E. A. Zoeller), 2007, ch. 2 (Nova Science Publisher, Hauppauge, New York).
- 17 **Gupta, S. and Ray, A.** Pattern identification using lattice spin systems: a thermodynamic formalism. *Appl. Phys. Lett.*, 2007, **91**(19), 194105.
- 18 **Gupta, S. and Ray, A.** Statistical mechanics of complex systems for pattern identification. *J. Stat. Phys.*, 2009, **134**(2), 337–364.
- 19 **Lind, D. and Marcus, M.** *An introduction to symbolic dynamics and coding*, 1995 (Cambridge University Press, UK).
- 20 **Ozekici, S.** *Reliability and maintenance of complex systems*. In NATO Advanced Science Institutes (ASI) Series F: Computer and Systems Sciences, vol. 154, 1996 (Springer, Berlin, Germany).
- 21 **Khatkhate, A., Gupta, S., Ray, A., and Keller, E.** Life extending control of mechanical structures using symbolic time series analysis. In Proceedings of the American Control Conference, Minnesota, USA, 2006.
- 22 **Subbu, A. and Ray, A.** Space partitioning via Hilbert transform for symbolic time series analysis. *Appl. Phys. Lett.*, 2008, **92**, 084107.
- 23 **Buhl, M. and Kennel, M.** Statistically relaxing to generating partitions for observed time-series data. *Phys. Rev. E*, 2005, **71**(4), 046213.
- 24 **Rajagopalan, V. and Ray, A.** Symbolic time series analysis via wavelet-based partitioning. *Signal Process.*, 2006, **86**(11), 3309–3320.
- 25 **Mallat, S.** *A wavelet tour of signal processing*, 2nd edition, 1998 (Academic Press, Boston, MA).
- 26 **Naylor, A. W. and Sell, G. R.** *Linear operator theory in engineering and science*, 1982 (Springer-Verlag, New York).