A Real Time Neural Net Estimator of Fatigue Life

T. Troudet
Sverdrup Technology, Inc.
Lewis Research Center Group
Brook Park, Ohio

and

W. Merrill
National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio

Prepared for the
International Joint Conference on Neural Networks
cosponsored by the IEEE and INNS
San Diego, California, June 17-21, 1990
A REAL TIME NEURAL NET ESTIMATOR
OF FATIGUE LIFE

T. TROUDET
Sverdrup Technology, Inc.
NASA Lewis Research Center Group
Brook Park, Ohio 44142

W. MERRILL
National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135

Abstract

A neural net architecture is proposed to estimate, in real-time, the fatigue life of mechanical components, as part of the Intelligent Control Systems for Reusable Rocket Engines. Arbitrary component loading values were used as input to train a two hidden-layer feedforward neural net to estimate component fatigue damage. The ability of the net to learn, based on a local strain approach, the mapping between load sequence and fatigue damage has been demonstrated for a uniaxial specimen. Because of its demonstrated performance, the neural computation may be extended to complex cases where the loads are biaxial or triaxial, and the geometry of the component is complex (e.g. turbopump blades). The generality of the approach is such that load/damage mappings can be directly extracted from experimental data without requiring any knowledge of the stress/strain profile of the component. In addition, the parallel network architecture allows real-time life calculations even for high frequency vibrations. Owing to its distributed nature, the neural implementation will be robust and reliable, enabling its use in hostile environments such as rocket engines. This neural net estimator of fatigue life is seen as the enabling technology to achieve component life prognosis, and therefore would be an important part of the life extending control for reusable rocket engines.

1 Introduction.

Intelligent Control Systems (ICS) for Reusable Rocket Engines (RRE), such as the Space Shuttle Main Engine (SSME), require expert system based decision-making within the response times of the RRE [1]. For a given level of knowledge of the RRE, and a given amount of computational power allocated to the expert system, accurate decision-making requires a complete and timely description of the dynamical state of the RRE. Such a timely dynamic information can be obtained by using peripheral Application Specific Processors (ASP) interacting with a general-purpose Central Processing Unit (CPU), i.e. the expert system, in the spirit of sixth generation computers [2]. The ASP would include hardwired, parallel distributed architectures, e.g. neural networks, to extract global dynamic state information from a myriad of sensor data (in the time and space domain).

This cybernetics perspective of the ICS is suggestively illustrated by analyzing the visual information processing pathway: eye-ball/retina/brain. The sensor function is performed by a spatial filter (eye-ball), which focuses the individual photons onto the retina (ASP), where global information such as contrast enhancement and motion detection is extracted by an intricate neural network.
of cones, rods, bipolar and horizontal cells. When reaching the brain, (general purpose CPU), the visual information has been compressed in a ratio of 150/1, and therefore can be more efficiently analyzed in relation to patterns or information previously stored in the brain. Likewise, the engineering model stored in the expert system can be best utilized on the premise of an efficient compression of the sensor data.

From this perspective, we propose an integrated architecture which will regularly measure the load values as a function of time, identify the presence of load cycles, and extract in real-time their contribution to the cumulative fatigue damage. Component life prognosis and failure detection will be achieved through such an accurate monitoring of the load history.

2 Architecture of the Real Time Fatigue Life Estimator.

As indicated in Fig.4, the real time fatigue life estimator that is proposed consists of two functional blocks: a preprocessor and a neural network. The role of the preprocessor is to perform simple front-end signal processing in order to improve the computational efficiency of the neural network.

2.1 Preprocessor.

First, the preprocessor extracts the peak-to-peak load transitions by comparing the loads sampled at three successive time intervals. Once an extremal load value has been identified, it is stored in a shift-register type of buffer which is continuously analyzed to detect the occurrence of load cycles. If a load cycle is identified, the content of the shift-register is input to a neural net architecture. If not, the shift-register is ready to receive an additional extremal load value. In order to define a load cycle and understand its impact on the fatigue life, it is best to review briefly the underlying concepts of the uniaxial local strain approach [3] [4] [5].

2.1.1 Uniaxial Local Strain Approach.

To demonstrate the concepts of the local strain approach, it is suggestive to interpret the mechanical effects of load changes in terms of dynamical trajectories in the stress/strain plane. It assumes a uniaxial behavior of the component, i.e. a stress \( \sigma \) and a strain \( \epsilon \), in the direction of the applied load \( P \). Changes in stress are related to changes in strain according to the standard elastic/plastic decomposition of the total strain range \( \Delta \epsilon \)

\[
\Delta \epsilon = \frac{\Delta \sigma}{E} + \Delta \epsilon_p
\]

where the elastic strain range, \( \frac{\Delta \sigma}{E} \), is the ratio of the total stress range, \( \Delta \sigma \), over the elastic modulus, \( E \); \( \Delta \epsilon_p \) is the plastic strain range. This non-linearity of the stress/strain relationship is responsible for the occurrence of hysteresis loops in the \( \{ \sigma, \epsilon \} \) plane which characterize the degree of fatigue damage of the component caused by the corresponding load variations. Examples of trajectories in the \( \{ \sigma, \epsilon \} \) plane are given in Fig.1 for an irregular load sequence applied to the fatigue specimen of Ref[3]. Properties of RQC100 steel are used in this paper to illustrate the correlations between load sequences and fatigue damage. A peak value of the load corresponds to a reversal in the loading direction, e.g. compression to tension, or vice-versa. As a result of such reversals, small cracks are initiated and enlarged, leading ultimately to mechanical rupture. The fatigue damage of a load cycle is therefore characterized by the number of times, \( N_{\text{cycle}} \), that such a cycle can be repeatedly applied to a component before a crack develops in the material. For this reason, we will refer to \( N_{\text{cycle}} \) as cycle number.
According to the Linear Damage Rule (LDR) [6]-[7], the fatigue damage $\rho_{\text{cycle}}$ of a hysteresis loop (such as one of those in Fig.1b) is given by the ratio

$$1 \geq \frac{1}{\rho_{\text{cycle}}} = \frac{1}{N_{\text{cycle}}} > 0,$$  \hspace{1cm} (2)

An estimation of the fatigue life is obtained by cumulating the contributions $\rho_{\text{cycle}}$ of the individual cycles which occurred during the load history. In the LDR approach, the fatigue damage of a load cycle is independent of the fatigue damage accumulated by the component. This cycle-counting method is useful for the treatment of random load sequences, where the load history consists of a succession of cycles randomly distributed in the stress/strain plane. Yet, its predictions are non-conservative when the interaction of loadings at different strain/stress levels has to be taken into account. Such is the case of high-low load sequences, where the fatigue damage of a load cycle not only depends on its cycle number, $N_{\text{cycle}}$, but also on the cumulative fatigue damage of the component. This necessitates the use of Damage Curve Approaches (DCA) [8]-[9] to reflect the effects of such non-linear interactions on component fatigue life. In spite of these non-linearities, component fatigue life can be estimated from the sole knowledge of the mapping between load cycle and cycle number, whether the Fatigue Damage Accumulator is implemented using LDR or DCA (Fig.4).

It is the purpose of this paper to demonstrate the possibility of learning, based on a local strain approach, the mapping of random load cycles with their fatigue damage $\rho_{\text{cycle}} = \frac{1}{N_{\text{cycle}}}$.

The way the preprocessor detects the existence of a load cycle from a set of four consecutive extremal values is indicated in Fig.2, together with a diagram of the corresponding hysteresis loop in the $\{\sigma, \epsilon\}$ plane. In Fig.2a, the compression sequence AB is interrupted in B by the hysteresis loop BCB (tension sequence BC, followed by compression sequence CB), before resuming to BD. Fig.2b is obtained from Fig.2a by exchanging compression and tension. The existence conditions for hysteresis loops are $P_A \geq P_C$ and $P_B \geq P_D$ (Fig.2a), or $P_A \leq P_C$ and $P_B \leq P_D$ (Fig.2b). Compressive and tensile load cycles are shown in Fig.3a for a load sequence resulting from the superposition of low-frequency and high-frequency vibrations. In contrast, several load cycles may be initiated without being completed, as is the case for the damped vibrations shown in Fig.3b.

Whenever load reversals associated with a load cycle have been identified, the preprocessor passes the result of the neural computation to the Fatigue Damage Accumulator as shown in Fig.4; and then removes the load cycle from the content of the shift-register.

Although the local strain approach assumes the existence of a uniaxial state of stress, its underlying concepts of hysteresis looping and memory effects are sufficiently general to be valid under more complex conditions [3], e.g. biaxial/triaxial states of strain/stress.

2.2 Neural Network Architecture.

Once a load cycle has been identified by the preprocessor, the contents of the shift-register are input to a feedforward neural net. As shown in Fig.4, each input unit of the neural net is in one-to-one correspondence with each buffer of the shift-register of the preprocessor. This means that a peak value of the load sequence is fanned out to each neuron of the first hidden layer. The activation function of a neuron is chosen to be the exponential type of I/O response function, $\text{output} = \frac{1}{1 + e^{-\text{input}}}$, where the input is the sum of the weighted outputs of all the connected neurons and a threshold term. The neural net output contributes to the fatigue damage only if the presence of a load-cycle is identified by the preprocessor.

An important issue which affects the design of the neural architecture is the accuracy with which the fatigue damage needs to be estimated. To estimate the cumulative fatigue damage of a component, it is necessary to estimate accurately the fatigue damage due to each individual cycle since the load history consists of a multitude of load reversals of small amplitude. We propose a neural architecture that can be expanded laterally in order to adaptively increase the precision.
of the neural computation. The architecture is an arrangement of neural net building blocks that is functionally reminiscent of a successive approximation register, operating in parallel, and it is described as follows.

If $p_{\text{min}}$ is an acceptable lower bound for the fatigue damage of an individual hysteresis loop, a feed-forward net $F_1$ is trained to map individual load cycles over the interval $[0, -\log(p_{\text{min}})]$. If $F_1$ has learned the mapping with a sufficient accuracy, it can be used to predict the fatigue life. If not, $F_1$ is used as a classifier to separate the cycle contributions that are larger than $-0.5\log(p_{\text{min}})$ from those that are smaller. Two feed-forward nets, $F_2$ and $F_3$, are then trained separately to map the load cycles over the intervals $[0, -\log(p_{\text{min}}) + \delta]$ and $[-\log(p_{\text{min}}) - \delta, -\log(p_{\text{min}})]$ respectively, where $\delta$ is the average resolution of $F_1$. After training, $F_1$, $F_2$, and $F_3$ are operated in parallel, with the output node of $F_1$ inhibiting the output node of $F_2$ or $F_3$ which does not correspond to the class defined by $F_1$. As shown in Figure 5, the remaining active node of $F_2$, or $F_3$, provides an estimation of the fatigue damage caused by the load-cycle which has been identified by the preprocessor. If this estimation is accurate, the parallel arrangement of $F_1$, $F_2$, $F_3$ can be used for prognosis/diagnosis of the mechanical fatigue. If not, $F_2$ and $F_3$ are used as classifiers, and the architecture is expanded laterally with four additional nets whose outputs are activated/inhibited by the output nodes of $F_2$ or $F_3$. This procedure can be reiterated until desired accuracy is reached.

To illustrate the practicability of this approach and demonstrate the possibility of extracting the highly non-linear features which govern the material behavior in complex situations, an architecture of two-hidden layer neural building blocks has been trained in supervised learning mode [10]-[11] to predict the fatigue damage of a specimen of RQC100 steel in the uniaxial local strain approach. When learning from experimental data, the choice between a back-error-propagation network (BEP) or a counter-propagation network will depend upon the degree of representativeness of the training data set. When learning from an analytical model, e.g. the local strain approach, it is possible to generate a training data set that is highly representative of the load/damage relationship. In that case, a good internal representation of the mapping can be developed by a BEP network [12]-[13] known to be more efficient as the degree of representativeness of the training data set increases.

## 3 Neural Network Training.

The training data set was generated on the basis of the uniaxial local strain approach for cyclic loading within fixed limits [3]-[5]. The initial value of strain, $\epsilon_0$, is calculated according to the load strain calibration of the loading path

$$\epsilon_0 = \text{sign}(P_0) \times \left( \frac{|P_0|}{C_1} + \left( \frac{|P_0|}{C_2} \right)^{1/d} \right)$$

(3)

where $P_0$ is the final value of the loading path, and the fitting constants are $C_1 = 1410$ ksi (kilopounds per square inch), $C_2 = 65.5$ ksi, and $d = 0.31$ for the geometry of the uniaxial RQC100 specimen given in Ref.[3]. The initial value of the stress $\sigma_0$ corresponding to $P_0$ is calculated following the non-linear stress/strain relationship given in Eq.(1)

$$|\epsilon_0| = \frac{|\sigma_0|}{E} + \left( \frac{|\sigma_0|}{A} \right)^{1/s} \; ; \; \text{sign}(\sigma_0) = \text{sign}(\epsilon_0)$$

(4)

where the elastic modulus is $E = 29500$ksi, $A = 167$ksi and $s = 0.10$ for RQC100 steel. The initial loading creates a cristallographic slip in the material, which modifies its mechanical response under subsequent loadings. The irreversible effect of this cristallographic slip on component behavior is accounted for by expanding the calibration curve and the stress/strain curve, Eqs.(3)-(4), with a factor of 2 and by shifting their origins.
After loading, the "image" of a load sequence in the \( \{ \sigma, \epsilon \} \) plane is therefore obtained as the solution of the set of equations

\[
\frac{\epsilon - \epsilon_r}{2} = \frac{|P - P_r|}{2C_1} + \left( \frac{|P - P_r|}{2C_2} \right)^{1/d} = \frac{|\sigma - \sigma_r|}{2E} + \left( \frac{|\sigma - \sigma_r|}{2A} \right)^{1/s};
\]

\[
\text{sign}(\sigma - \sigma_r) = \text{sign}(\epsilon - \epsilon_r) = \text{sign}(P - P_r)
\]

(5)

where \( (P, \epsilon, \sigma) \) corresponds to the origin of the stress-strain hysteresis loop curve of the current "point" \( (P, \epsilon, \sigma) \). For a hysteresis loop \( H_{12} \) starting at \( (P_1, \epsilon_1, \sigma_1) \) and ending at \( (P_2, \epsilon_2, \sigma_2) \), the number \( N_{12} \) of loops \( H_{12} \) that would lead to the initiation of a crack is obtained from the fitted equation \[3\]

\[
\frac{|\epsilon_2 - \epsilon_1|}{2} = \frac{\sigma_{tf} - \sigma_m}{E} (2N_{12})^b + \epsilon_{tf} (2N_{12})^c
\]

(6)

with \( \sigma_m = \frac{\sigma_1 + \sigma_2}{2} \) if it is a tensile mean-stress, i.e. \( \sigma_m > 0 \), and \( \sigma_m = 0 \) if it is a compressive mean-stress. For RQC100 steel, the fatigue strength coefficient is \( \sigma_{tf} = 168 \text{ksi} \), and the fatigue strength exponent is \( b = -0.075 \); the fatigue ductility coefficient is \( \epsilon_{tf} = 1.06 \), and the fatigue ductility exponent is \( c = -0.75 \).

It is well-known that the performance of the neural computation can be significantly enhanced by embedding explicitly any a-priori knowledge or constraints of the system in the preprocessing or the neural architecture itself. The proposed neural estimation of the fatigue life is actually subject to the detection of load cycles by the preprocessor. Similarly, any a-priori knowledge about the gross dynamic properties of the system is likely to improve the performance of the neural computation. For example, load sequences that are periodic or quasi-periodic, such as in Fig.3a, lead to a maximum of two or three cycles that remain simultaneously open during the entire load history. This implies that the mechanical behavior of components subjected to load patterns such as the one of Fig.3a can be predicted by reducing the number of input units to a maximum of 8-10. In addition, the regularity of such load patterns facilitates the feature extraction, and improves the ability of the net to predict component behavior for such load patterns. In contrast, for the most irregular sequences where the loads are randomly distributed, load patterns such as in Fig.3b can occur, so that more buffers are needed in the shift-register of the preprocessor. In the latter case, training/learning is more difficult because of the great irregularities of the load variations.

The behavior of RQC100 steel under random load sequences has been simulated for cyclic loading within fixed limits [3] [4] [5] in the dynamic range \( D_p = [-P_{max}, +P_{max}] \). The initial load \( P_i \) is arbitrarily chosen over \( D_p \), and defines the domain of variations of the subsequent loads \( P(t) \), \( |P(t)| \leq |P_i| \). (The neural estimation of cumulative damage resulting from cyclic loading with variable boundaries, for which there exists \( t \) such that \( |P(t)| \geq |P_i| \), is the object of an on-going analysis.) The random load sequences used to generate the training data set are obtained as follows.

A uniform set \( \{P_i\} \) of initial loading values is generated over \( D_p \). For each \( P_i \), a uniform set \( \{P_{av}\} \) of average values of the load sequence is generated with the condition that \( |P_{av}| \geq |P_i| \). Given an average value \( P_{av} \), a uniform set \( \{P_a\} \) of maximum amplitudes of the load variations is generated over \( [0, \max(|P_i - P_{av}|, |P_i + P_{av}|)] \). For each triplet \( (P_i, P_{av}, P_a) \), a load sequence is generated as a set of \( K \) points randomly distributed over \( [P_{av} - P_a, P_{av} + P_a] \), from where only individual load cycles such that \( 1 \geq \rho_{cycle} \geq \rho_{min} \) are retained for training. The interval to be mapped, \( [0, -\log(\rho_{min})] \), is divided in 10 equal subintervals \( \{I_k, k = 1, 10\} \) which are equally populated by data generated from random load sequences. A training iteration consists of choosing randomly a value \(-\log(\rho_{cycle})\) in one of the randomly picked subintervals \( I_k \). The load values of the corresponding sequence are scaled as \( P_{cycle} = \frac{P_{cycle}}{P_{max}} \cdot [-1, +1] \), and the fatigue damage is scaled over the active range \([0.1, 0.9]\) of the neuron output of the last layer:

\[
\rho_{target} = \frac{\log(\rho_{cycle})}{\log(\rho_{min})} \times 0.8 + 0.1
\]

(7)
To favor homogeneous learning of the mapping load/damage over \([0, -\log(\rho_{\text{min}})]\), the weights and thresholds were updated by batching the steepest-descent contributions of 10 consecutive I/O pairs \((\hat{P}_{\text{cycle}}, \hat{\rho}_{\text{cycle}})\).

In this simulation of the behavior of a uniaxial RQC100 steel specimen [3], the dynamic range of load variations has been bounded by \(P_{\text{max}} = 90\text{ksi}\), i.e., approximately at half the value of the fatigue strength coefficient \(\sigma_f = 167\text{ksi}\). With this value of \(P_{\text{max}}\), it was found that load sequences corresponding to \(K = 200\) random values were long enough to indicate that a probable maximum number of 7 initiated cycles can remain simultaneously open during a random load history. On the basis of this statistical result, 15 buffers were allocated to the shift-register of the preprocessor. Cycles contributing less than \(\rho_{\text{min}} = 10^{-10}\) were neglected, and 10,000 elements were used for training.

### 4 Neural Network Performance.

Training was done on a PC-based accelerator coprocessor board (ANZA+) and the elementary neural building block was a two-layer feedforward net with 15 input units, 100 neurons in the first hidden layer, 50 neurons in the second hidden layer, and a single neuron in the output layer. The performance of the neural computation was evaluated by averaging the absolute value of the error \(\varepsilon_{\text{cycle}} = |\hat{P}_{\text{cycle}} - P_{\text{cycle}}|\) over a set of 10,000 I/O pairs \((\hat{P}_{\text{cycle}}, \hat{\rho}_{\text{cycle}})\) that were not used to train the net, and with the scaled fatigue damage \(\hat{\rho}_{\text{cycle}}\) uniformly distributed over \([0.1, 0.9]\).

The average absolute error of this net \(F_1\) is \(\varepsilon = \langle \varepsilon_{\text{cycle}} \rangle = 0.021\) with a standard deviation \(\delta\varepsilon = 0.029\) (versus the average \(\varepsilon_{\text{learning}} = 0.020\) and the standard deviation \(\delta\varepsilon_{\text{learning}} = 0.026\) calculated for the training data set); and the distribution of the estimated values is symmetric with respect to the exact values. This represents accurate learning (within 2\%) of the uniaxial strain approach. To interpret the effect of this error on fatigue damage estimation, recall that a logarithmic relationship exists between the network output and the fatigue damage estimate. Thus, the fatigue damage predicted by the network has the same probability of being larger or smaller than that predicted by the teacher. If it is larger, it will be in the average \(r = 1.8\) times the value predicted by the teacher; and if it is smaller, it will be in the average \(\frac{1}{r} = 0.55\) times the value predicted by the teacher. Said another way, 90\% of the estimated values will be within a factor 4 and 1/4 (standard deviation upper-bound) of the exact data. In order to demonstrate the ability of the architecture to improve the accuracy of the neural computation, two additional neural building blocks, \(F_2\) and \(F_3\), have been trained to map load-cycle/fatigue-damage for \(\hat{P}_{\text{cycle}}\) over the intervals \([4.5, 5.5]\) and \([0, 4.5, 10]\) respectively. As shown in Fig.5, net \(F_1\) is used as a classifier for the output nodes of \(F_2\) and \(F_3\). With this two-stage neural architecture, the value of the average ratio \(r\) defined above is lowered to \(r = 1.5\), and 90\% of the estimated values are within a factor 2.5 and 1/2.5 (standard deviation upper-bound) of the exact data. It is also noted that 75\% of the estimated values are already within a factor 1.7 and 1/1.7 of the exact data.

It is finally emphasized that no attempt has been made to look for the minimum number of neurons and synaptic connections that provide the same degree of accuracy as the two-hidden layer building block chosen in this simulation.

### 5 Conclusion.

A neural net architecture has been proposed to extract the correlations between random load sequences and their contributions to cumulative damage in order to predict component fatigue damage. The efficiency of the neural computation to map load sequences with fatigue damage has been demonstrated on a uniaxial RQC100 notched member for the most irregular load histories of cyclic loading with fixed boundaries. The real-time capability of this parallel processing architecture, together with...
the simplicity of its practical implementation, makes it computationally very attractive to enhance the performance of complex ICS. Because of the complexity of the component couplings of the RRE, the ICS has to be able to anticipate in real-time the effect of global changes of conditions on individual components, e.g. the effect of a gas flow variation on the fatigue life of a turbopump blade. In that context, it would be natural to extend the concepts of neural computation developed in this work to the fatigue life prediction of multiaxial components subjected to temperature variations [14]-[15].

Acknowledgments.

We would like to thank C. Lorenzo, Dr. S. Kalluri and M. Krasowski for their critical review of the manuscript; Dr. M. McGaw for informative discussions on Materials and Structures; and Dr. F. Troudet for inspiring discussions on Cybernetics.

References.

(a) Loads distributed between -10 and 10 kips. Initial loading value of 10 kips.

(b) Note the closed hysteresis loops 5-8-5' and 6-7-6', and the gradual formation of the outer cycles.

Figure 1. - Open load sequence and its image in the \( (\sigma, \varepsilon) \) plane.

(a) Compressive.

(b) Tensile.

Figure 2. - Load cycles and their image in the \( (\sigma, \varepsilon) \) plane.

(a) Reducible load sequence.

(b) Irreducible load sequence.

Figure 3. - Extreme cases of load sequences.
Figure 4. - Neural Net based fatigue life estimation.

Figure 5. - A two-stage successive approximation neural net architecture for fatigue life estimation.
A neural net architecture is proposed to estimate, in real-time, the fatigue life of mechanical components, as part of the Intelligent Control Systems for Reusable Rocket Engines. Arbitrary component loading values were used as input to train a two hidden-layer feedforward neural net to estimate component fatigue damage. The ability of the net to learn, based on a local strain approach, the mapping between load sequence and fatigue damage has been demonstrated for a uniaxial specimen. Because of its demonstrated performance, the neural computation may be extended to complex cases where the loads are biaxial or triaxial, and the geometry of the component is complex (e.g., turbopump blades). The generality of the approach is such that load/damage mappings can be directly extracted from experimental data without requiring any knowledge of the stress/strain profile of the component. In addition, the parallel network architecture allows real-time life calculations even for high frequency vibrations. Owing to its distributed nature, the neural implementation will be robust and reliable, enabling its use in hostile environments such as rocket engines. This neural net estimator of fatigue life is seen as the enabling technology to achieve component life prognosis, and therefore would be an important part of the life extending control for reusable rocket engines.