PERGAMON

International Journal of Solids and Structures 37 (2000) 6397–6417

INTERNATIONAL JOURNAL OF SOLIDS and STRUCTURES

www.elsevier.com/locate/ijsolstr

Steps towards a qualitative dynamics of damage evolution

J.P. Cusumano*, A. Chatterjee

Department of Engineering Science & Mechanics, Penn State University, 227 Hammond Building, University Park, PA 16802, USA

Received 21 April 1998; received in revised form 23 October 1998

Abstract

Basic elements of a qualitative dynamics of systems with slowly accumulating damage are presented. The relationship of averaging to equations relating damage evolution rate with current damage state and load is discussed. The particular case of a scalar damage variable with one load parameter is examined in some detail. Well-known phenomenological features of fatigue tests are used as a guide. Using asymptotic arguments for small initial damage, conditions are obtained under which the time to failure is essentially determined by the initial damage state, and essentially independent of the failure criterion. A suitable form for damage evolution equations is deduced for a class of fatigue problems. Comparisons with published data show that this qualitative dynamics approach produces a model that fits, and predicts, the data as closely as a more physically motivated fatigue model.

Keywords: Damage; Fatigue; Failure prediction; Qualitative dynamics; Nonlinear dynamics

1. Introduction

Theoretical studies of damage evolution and failure prediction often employ differential equations developed with varying degrees of rigor from some underlying theory of material failure, or perhaps empirical damage mechanics (Chaboche, 1988a,b; Suresh, 1991; Lemaitre, 1992). In this paper, we present a more abstract, state-space based framework in which to formulate and evaluate damage evolution laws. We do not specifically tie our models to any particular damage physics, but show how a few reasonable *mathematical* assumptions can lead to models that usefully predict damage evolution. In particular, using qualitative arguments, we develop a simple model for a class of fatigue problems. The parameters in the model can be fitted from experiments. On fitting these parameters to some published

* Corresponding author. Tel.: +814-865-3179; fax: +814-863-7967.
E-mail address: jpc@crash.esm.psu.edu (J.P. Cusumano).

0020-7683/00/$ - see front matter
PII: S0020-7683(99)00042-6

data, we find that the model quantitatively matches the data as closely as a more physically motivated fatigue model.

Models similar in form to those considered here can be found elsewhere in the damage, fatigue and fracture mechanics literature. The primary contribution of this paper, however, is to demonstrate that a study of the qualitative dynamics of damage evolution, as indicated by certain well-known observations from the experimental literature, can shed great light on the necessary mathematical form of the required evolution laws, without being necessarily tied to a particular damage mechanism.

We do not wish, however, to deemphasize the importance of physics: in modeling, it is always useful to know about underlying physical processes. Our aim is only to show that models and methods sometimes thought of as arising *only* from the physics of material failure seem less specialized when viewed from a general dynamical systems perspective, implying that similar modeling approaches might be useful for types of damage that involve significantly different kinds of material failure, and even systems having no material failure per se.

The approach taken here makes explicit several features of the failure prediction problem which we believe should be addressed by condition monitoring programs. In particular, we emphasize the notions of damage state space, failure surface, time to failure and sensitivity to initial damage state. We attempt to make a clear connection between the forms of damage evolution laws and their *qualitative* effects on quantities of interest such as time to failure and sensitivity to initial state.

In this paper, systems with slowly accumulating damage are viewed as hierarchical dynamical systems with two distinct time scales: regular operation of these systems occurs on a fast time scale, while damage evolves, at least initially, on a slow time scale. In section 2 we show how the full state space description of hierarchical systems with evolving damage, via the method of averaging, can yield simple damage evolution laws. In section 3 our basic assumptions are presented, and in section 4 some relevant ideas from engineering practice are discussed. With the basic framework established, section 5 presents a detailed analysis of the scalar damage variable case. Qualitative behaviors of several basic damage evolution laws are examined. An application example is provided in section 6, where a damage rate law developed earlier is shown, on some parameter fitting, to match and predict multi-level fatigue data from the literature. Conclusions are presented in section 7.

2. Averaging and damage evolution laws

2.1. Time scale separation

The observation of time scale separation in hierarchical systems with slow damage is not new. This same idea is, for example, emphasized in Natke and Cempel (1997). It is not usually made explicit, however, that such time scale separation allows one to apply the method of averaging (or a related method, such as multiple scales—Verhulst, 1990; Nayfeh and Mook, 1995) to obtain approximate equations relating damage evolution rate to current damage state and load.

In our study of finite dimensional systems with damage, we consider dynamical systems of the form

$$\dot{x} = f(x, \mu(\phi), F, t) \tag{1a}$$

$$\dot{\phi} = \epsilon g(\phi, x, t) \tag{1b}$$

where: x is a fast (possibly vector-valued) dynamic variable viewed as the 'macroscopic' or directly observable state of the system; ϕ is a slow (also possibly vector-valued) 'hidden' dynamic variable representing the damage state; $0 < \epsilon \ll 1$ is a small positive number interpreted as a rate constant; μ is

a (possibly non-invertible) function of ϕ representing the material parameters in the fast subsystem (1a); F is a vector of parameters representing external loads; and t is time. Overdots denote time derivatives. We remark that Eqs. 1 also include continuum damage evolution if the vector fields f and g are allowed to be partial differential operators in the space variables, though we consider only the finite dimensional case.

Note that if $\epsilon = 0$ in Eq. (1b), the damage ϕ is constant, and $\mu(\phi)$ is a vector of parameters in Eq. (1b). If ϵ is nonzero but small, then ϕ evolves slowly, and the evolution of x over short to intermediate lengths of time may be approximately determined by treating $\mu(\phi)$ as constant in Eq. (1a). This is the heuristic motivation behind the method of averaging (Verhulst, 1990; Sanders and Verhulst, 1985), a perturbation method used to obtain approximate analytical solutions to nonlinear differential equations.

2.2. Averaging procedure

In averaging, the governing equations are usually transformed into 'normal form' using variation of parameters, which implies that explicit analytical solutions are available for the unperturbed case. The implementation and asymptotic performance of the averaging procedure depends on the particular properties of the systems under consideration, especially the nature of solutions to the unperturbed problem.

However, we are not primarily concerned here with averaging as an analytical procedure, but only with the fact that in many cases of practical importance averaging can be carried out *in principle*. Our main interest is to establish the theoretical relationship between the general state space description of Eqs. (1) and damage evolution laws as typically found in the literature. For simplicity and brevity, we present the method of averaging in the most basic form required to determine its effect on the slow time subsystem Eq. (1b), and present an example calculation in the Appendix.

At some initial time $t = t_0$, let $x = x_0$ and $\phi = \phi_0$. Let $X_0(\phi_0, x_0, t_0, F; t)$ be the solution to the unperturbed equation

$$\dot{x} = f(x, \mu(\phi_0), F, t), \tag{2}$$

with $x(t_0) = x_0$. For $\epsilon > 0$, we average the right hand side of Eq. (1b), defining the function $\bar{g}$ by

$$\bar{g} = \lim_{T\to\infty} \frac{1}{T}\int_{t_0}^{t_0+T} g(\phi, X_0(\phi, F; t), t)\, \mathrm{d}t, \tag{3}$$

where, in X_0, ϕ_0 has been replaced by ϕ and the dependence on x_0 and t_0 has been suppressed. If the solution to Eq. (2) has transients which decay rapidly leaving a well defined steady state solution, the average $\bar{g}$ defined in Eq. (3) will be independent of t_0. If Eq. (2) has multiple steady state behaviors, then $\bar{g}$ will be dependent on x_0 only in that it determines which steady state the system approaches. Thus we may explicitly ignore the dependence on x_0, but implicitly recognize that $\bar{g}$ can be multivalued. Therefore, we write the averaged function of (3) as $\bar{g} = \bar{g}(\phi, P(F))$, assuming that if $P(F)$ is multivalued, then we know which solution branch is under consideration.

Up to leading order for small ϵ, the evolution of ϕ is given by the averaged equation

$$\dot{\phi} = \epsilon \bar{g}(\phi, P(F)).$$

We can now rescale time using $\tau = \epsilon t$, to obtain the *slow flow equation*

$$\dot{\phi} = g(\phi, P), \tag{4}$$

where we drop the overbar, suppress the dependence of P on F, and let the overdot denote the

derivative with respect to 'slow time' τ. Eq. (4) demonstrates how averaging will in general take the dynamics of the complete hierarchical system of Eqs. (1) into a 'damage law' relating the damage evolution rate to the current damage state and the load (via the parameter vector P). These ideas are demonstrated using a simple example problem in the Appendix.

We remark that a basic consequence of averaging is that the right hand side of the slow flow equation developed above is not explicitly dependent on the slow time τ for constant F. Furthermore, if in Eq. (1a) F is a prescribed function of the slow time $\tau = \epsilon t$, the functional form of g in Eq. (4) is unchanged except for the fact that the vector P becomes a known function of τ as well. Thus, we do not consider damage evolution equations where the rate of increase of damage is explicitly dependent on the time even for the case when the load parameter is constant, though such equations have been proposed in the literature (Kujawski and Ellyin, 1988).

3. Fundamental assumptions

We now outline the fundamental assumptions and definitions used in our study of the slow flow system of Eq. (4). We focus on simple cases; our aim here is to identify issues important in establishing a state-space framework for damage evolution. If the physics of an application demands it, these assumptions may be modified.

1. *State of Zero Damage*: We assume that a well-defined state of zero damage can be identified, and that we may take this state as $\phi = 0$
2. *Smoothness*: We assume that the function g is as smooth as needed over the domain of interest.
3. *Failure Surface*: We assume that a well-defined failure surface exists, defined by the equation $h(\phi) = 0$: if $h(\phi) > 0$ the system has failed, and if $h(\phi) < 0$ it has not.
4. *Monotonicity*: We assume that damage variables are nondecreasing. In particular, for the special case of scalar damage, $\dot{\phi}$ is assumed to be nonnegative for all ϕ and P. Thus, we do not consider systems with 'healing'.

Note that item 4 above *is not* the same as assuming that $\mu(\phi)$ is nondecreasing. For example, depending on details of its material behavior, a vibrating cracked beam might initially stiffen in the vicinity of the crack, and then weaken until failure. Thus the stiffness (if represented by $\mu(\phi)$) is not a monotonic function of ϕ or of time. However, the beam steadily and irreversibly acquires damage, and ϕ may reasonably be expected to be a monotonic function of time. Monotonicity is often assumed in other work as well, and may be thought of as being related to irreversibility.

The foregoing assumptions are not tied to any particular physics of damage. For example, ϕ is not necessarily the length of a crack. In fact, it is known that the growth rate of small cracks is not a smooth function of crack length, if the crack length is comparable to microstructural features (James and de los Rios, 1996; Hussain et al., 1993); however, for systems where damage and failure occur through fracture, we assume that there is a suitable internal variable that need not be the crack length, and that evolves monotonically and smoothly.

We will show how under assumptions 1–4 above a qualitative study of candidate damage laws can identify one that is suitable for a particular application.

4. Relationship to engineering practice

Given the mathematical framework presented above, we now discuss some observations and ideas from the field of fatigue. These provide guidelines for the rest of this paper. The idea is to examine the

dynamical features of the physical phenomena being studied, so as to deduce restrictions on, and appropriate forms for, Eq. (4). The facts discussed here can be found in engineering design handbooks (Rothbart, 1996), or books on fatigue (Suresh, 1991).

The oldest and most widely used approach to the study of fatigue failures is through S–N curves, which are curves of load or stress amplitude S (during cyclic loading) versus the number N of cycles to failure. One might similarly construct W–N curves, where W is the positive work of deformation during a loading cycle. This paper does not explicitly distinguish between the two, since in either case the loading quantity (S or W) is identified with the parameter P.

From a state space perspective, damage evolution is a deterministic process described by Eq. (4), a failure surface $h(\phi)=0$, and damage initial conditions $\phi(0)=\phi_i$. The time to failure T is determined by the solution $\phi(\tau)$ to Eq. (4) with initial conditions ϕ_i, via the failure condition $h(\phi(T))=0$. T will clearly depend on both the load parameter P as well as the initial condition ϕ_i. Given ϕ_i, P can be plotted against T, giving our theory's equivalent of an S–N curve.

From this perspective, the statistical variability observed in fatigue experiments arises from the initial statistical distribution of ϕ_i in a sample, and the way this distribution evolves under Eq. (4). In fatigue experiments, substantial statistical variability is usually observed even for carefully prepared specimens with apparently little initial variation between them. This may be interpreted as sensitivity to initial conditions of the time to failure. Furthermore, the variability in fatigue life is known to be high for high cycle fatigue and somewhat lower for low cycle fatigue (Rothbart, 1996).

Many materials (including steel) have a *fatigue limit*: if loaded cyclically below this limit, they have essentially infinite life. Loading a specimen partially to failure, above the fatigue limit, results in a lowering of the fatigue limit. Conversely, loading a specimen below the fatigue limit can *increase* its fatigue limit to some extent.

Such qualitative observations provide guidelines for evaluating candidate forms for Eq. 4, as mentioned above.

Practical applications of fatigue models involve loads with fluctuating amplitudes. For such applications, the approach of S–N curves has been extended to 'cumulative damage' approaches, the earliest and best known of which is the Palmgren–Miner rule (Rothbart, 1996; Suresh, 1991, and references therein). Twenty different approaches to cumulative damage are discussed in Hwang and Han (1986), which may be an indication of both the importance as well as unresolved nature of this issue.

A basic difficulty encountered with most cumulative damage formulas of the kind discussed in Hwang and Han (1986) is that they ignore the dependence on initial conditions ϕ_i, something that we feel is of primary importance. These various formulas attempt to guess the functional dependence of the time to failure T on load parameters P, or perhaps postulate simple, directly integrable forms for $g(\phi, P)$ while always starting from the same initial conditions, which is the same thing. The scatter observed in experiments, which in our treatment is attributed to variations in ϕ_i, must then be directly incorporated as an ad hoc statistical quantity.[1]

In contrast, the approach taken here is to postulate the form of $g(\phi, P)$ based on a combination of mathematical considerations and physical observations, and to determine the time to failure T by integrating Eq. (4) with appropriate initial conditions. Since numerical integration of nonlinear differential equations has become easy with widespread access to computers, the loss of some of the graphical interpretations of older techniques (such as the double-linear damage rule of Manson and Halford, 1986, or the rotation of S–N curves about suitable points, as discussed in Kujawski and Ellyin,

[1] Sun (1994) goes further and proposes that even the evolution of damage from specimen to specimen in the population should be treated as a statistical quantity.

1988) is not serious, and is compensated by greater generality, in that a wider variety of systems may be studied.

Finally, we remark that papers treating the damage evolution problem often scale time τ with respect to the time to failure T (more precisely, they use n/N, or fraction of cycles to failure or 'fraction of life'). In a dynamical systems approach it is natural to view T as being determined by the function $g(\phi, P)$ and the initial condition ϕ_i. The fraction of life concept involves an awkward scaling of time that is dependent on initial conditions.[2] However, to enable comparison with published experimental data (Golos and Ellyin, 1988; Kujawski and Ellyin, 1988), in section 6 we express our results in the usual fraction of life form.

5. The case of a scalar damage variable

In general, the damage variable ϕ and parameter P in Eq. (4) are vector-valued. However, we now consider in some detail the case where both the damage variable and load parameter are scalars, since many basic ideas can be described easily in this context. Note that for this case, the damage 'surface' is a point, $\phi = \phi_f$, assumed finite and nonzero. We assume that $g(\phi_f, P) > 0$, i.e. the point of failure is not an equilibrium point.

The different qualitative cases considered here are shown schematically in Fig. 1.

5.1. No equilibria

Consider Eq. (4), and the possibility that there are no equilibria for nonzero load, i.e. $g(\phi, P) \neq 0$ for any $\phi \geq 0$ and $P > 0$ (see Fig. 1(a)). Then the time to failure T is given by

$$T = \int_{\phi_i}^{\phi_f} \frac{\mathrm{d}\phi}{g(\phi, P)}. \tag{5}$$

Since $g > 0$, the integral is finite and the system must eventually fail. In fact, defining

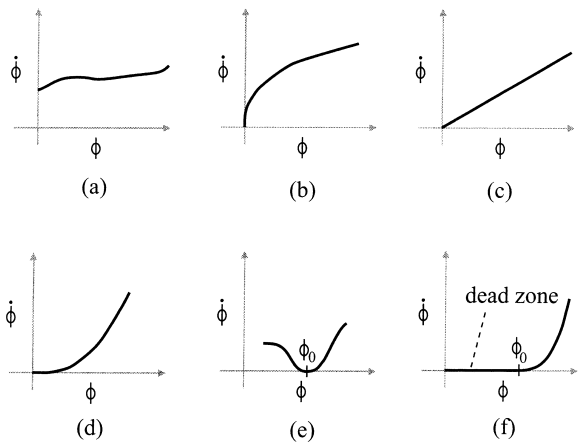

Fig. 1. Various cases considered for single damage variable—(a) no equilibria; (b), (c) and (d) single equilibrium at $\phi = 0$; (e) isolated equilibrium at $\phi_0 \neq 0$; (f) family of equilibria with a 'dead zone' for $0 \leq \phi \leq \phi_0$.

$$g_{\min} := \min_{0 \le \phi \le \phi_f} g(\phi, P),$$

the time to failure is bounded by $0 \le T \le \phi_f / g_{\min}$.

Clearly, there can be no fatigue limit for such systems. Furthermore, for carefully prepared, nearly identical specimens (that is, when the variation in ϕ_i is small) the variation in time to failure is given by

$$\Delta T \approx \frac{\mathrm{d}T}{\mathrm{d}\phi_i}\Delta\phi_i = -\frac{1}{g(\phi_i, P)}\Delta\phi_i. \tag{6}$$

Thus, the sensitivity of time to failure to variations in initial damage state will be small if initial conditions are such that $g(\phi_i, P)$ is large. This suggests, for example, that functions g which stay bounded well away from zero are not appropriate for typical fatigue problems because they do not show sensitivity to initial conditions nor the possibility of fatigue limits. Such functions might, of course, be relevant to other problems with slowly accumulating damage.

5.2. Single equilibrium at zero damage

If $\phi = 0$ is an equilibrium, i.e. $g(0, P) = 0$ for $P > 0$ (see Fig. 1(b)–(d)), then it is interesting to examine the behavior for small ϕ. We will examine this case in some detail.

Let us assume that g permits a power series representation in the neighborhood of zero, i.e.

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = g(\phi, P) = a_0(P)\phi^{m(P)} + \cdots, \tag{7}$$

and consider what happens if we retain only the first term of this series.

5.2.1. Time to failure

Taking only one term and solving Eq. 7 gives the time to failure:

$$T = \frac{1}{a_0(P)}\left(\frac{\phi_f^{1-m(P)} - \phi_i^{1-m(P)}}{1 - m(P)}\right) \quad \text{for} \quad m(P) \neq 1, \quad \text{and} \tag{8a}$$

$$T = \frac{1}{a_0(P)} \ln\left(\frac{\phi_f}{\phi_i}\right) \quad \text{for} \quad m(P) = 1. \tag{8b}$$

The above indicates the following possibilities: if $m(P) < 1$ (Fig. 1(b)), then the time to failure T is essentially determined by ϕ_f for sufficiently small initial damage ϕ_i; if $m(P) = 1$ (Fig. 1(c)), then T is determined by the ratio of ϕ_f to ϕ_i; finally, if $m(P) > 1$ (Fig. 1(d)) then T is essentially determined by ϕ_i for small initial damage.

5.2.2. Insensitivity to higher order corrections

The case of $m(P) > 1$ is interesting for the following reason. In this case, if ϕ_i is small, the time to

[2] Scaling of variables in a way that is dependent on initial conditions is not incorrect, just awkward. In dynamics, one usually views time and the system equations as separate from initial conditions. The evolution of the system with time is studied qualitatively and simultaneously for all initial conditions of interest. This is difficult if time or coordinates are transformed in ways dependent on initial conditions.

failure T is insensitive to higher order corrections in the power series representation of Eq. (7). To see this, consider an equation of the form

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = g(\phi, P) = a_0(P)\frac{\phi^{m(P)}}{s(\phi)} \tag{9}$$

where $s(\phi)$ is a strictly positive function of ϕ which is bounded everywhere, satisfies $s(0)=1$, and is smooth in some neighborhood of $\phi=0$. Let the maximum value of $s(\phi)$ in the interval $(0 \le \phi \le \phi_f)$ be $s_{\max}$. We assume that ϕ_i is small, and that $\phi_f = \mathcal{O}(1)$ with respect to ϕ_i. Then using Eq. (9), one finds

$$T = \frac{1}{a_0(P)}\int_{\phi_i}^{\phi_f} \frac{s(\phi)\,\mathrm{d}\phi}{\phi^{m(P)}}$$

which can be split into two parts, $T = I_1 + I_2$, where

$$I_1 := \frac{1}{a_0(P)}\int_{\phi_i}^{\sqrt{\phi_i}} \frac{s(\phi)\,\mathrm{d}\phi}{\phi^{m(P)}} \quad \text{and} \quad I_2 := \frac{1}{a_0(P)}\int_{\sqrt{\phi_i}}^{\phi_f} \frac{s(\phi)\,\mathrm{d}\phi}{\phi^{m(P)}}.$$

The integral I_1 is over a small domain in which $s(\phi)$ may be expanded in a power series since it is smooth near zero: in the interval $(0, \sqrt{\phi_i})$, let $s(\phi) = 1 + s_1\phi + s_2\phi^2 + \cdots$ for suitable constants $s_1, s_2, \ldots$. Assuming $m(P)$ is not an integer for simplicity (the case for integers involves logarithms, but leads to similar conclusions), we find that

$$I_1 = \frac{1}{a_0(P)}\left(\frac{\phi_i^{1-m(P)}}{m(P)-1} + s_1\frac{\phi_i^{2-m(P)}}{m(P)-2} + \cdots\right) + \frac{1}{a_0(P)}\left(\frac{\phi_i^{\{1-m(P)\}/2}}{1-m(P)} + s_1\frac{\phi_i^{\{2-m(P)\}/2}}{2-m(P)} + \cdots\right),$$

while I_2 is bounded by

$$0 < I_2 \le \frac{s_{\max}}{a_0(P)}\int_{\sqrt{\phi_i}}^{\phi_f} \frac{\mathrm{d}\phi}{\phi^{m(P)}} < \frac{s_{\max}}{a_0(P)}\int_{\sqrt{\phi_i}}^{\infty} \frac{\mathrm{d}\phi}{\phi^{m(P)}} = \frac{s_{\max}}{a_0(P)}\,\frac{\phi_i^{\{1-m(P)\}/2}}{m(P)-1}.$$

Thus, for $m(P) > 1$, the leading order term for I_1 provides a good approximation to T for small ϕ_i:

$$T \approx \frac{1}{a_0(P)}\,\frac{\phi_i^{1-m(P)}}{m(P)-1} \tag{10}$$

This shows that for $m(P) > 1$, taking a single term in the power series is a valid approximation. In fitting parameters to the model, only the leading order behavior near zero damage (i.e. the exponent $m(P)$) needs to be determined. Higher order terms have a negligible effect on the time to failure if ϕ_i is small.

We emphasize that our analysis is not intended to apply to the case when ϕ_i is allowed to be large. For example, we will show later that in fatigue problems $m(P)$ will typically be a decreasing function of the load P. In that case, it is easy to show (Chaboche, 1977) that if a first load is held for a sufficiently long time (i.e. ϕ is allowed to become sufficiently large), and then followed by a reduction in load, single-term power law models like the one presented above can incorrectly predict a reduction in the small amount of remaining life. However, the single-term power law model discussed here is *asymptotically valid* for systems with small initial damage (i.e. significant remaining life). Thus the contradictory qualitative predictions of such models *near to* failure are not relevant to our analysis.

5.2.3. Sensitivity to initial damage state

The time to failure, for $m(P) > 1$, will be strongly sensitive to variations in the initial damage state for small ϕ_i since

$$\frac{\mathrm{d}T}{\mathrm{d}\phi_i} = -\frac{1}{a_0(P)}\,\frac{1}{\phi_i^{m(P)}}. \tag{11}$$

In fact, the sensitivity indicated by Eq. (11) can be easily distinguished from the case when $m(P) < 1$ since

$$\left|\frac{\mathrm{d}T}{\mathrm{d}\phi_i}\phi_i\right| \to \infty \quad \text{as} \quad \phi_i \to 0 \quad \text{for} \quad m(P) > 1, \quad \text{whereas}$$

$$\left|\frac{\mathrm{d}T}{\mathrm{d}\phi_i}\phi_i\right| \to 0 \quad \text{as} \quad \phi_i \to 0 \quad \text{for} \quad m(P) < 1.$$

If we confine our attention to systems with $m(P) > 1$, and consider only small initial damage, we can ignore the small relative error and treat Eq. 10 as exact, especially in view of the high variability typically observed in fatigue data (as much as a factor of five or more). We remark that treating Eq. 10 as exact is equivalent to treating ϕ_f as infinity.

The implication is that in using such a model, the failure point ϕ_f need not be determined since it has a negligible effect on the time to failure if ϕ_i is small.

5.2.4. The suitability of $m(P) > 1$ for fatigue problems

Thus, $m(P) > 1$ seems appropriate for fatigue problems for several reasons. In a wide variety of fatigue experiments, it is observed that

1. damage (in the form of macroscopic cracks) becomes visible fairly late in the process, and
2. small differences in specimens cause large variations in time to failure.

In the context of our discussion, note that if $m(P) > 1$ and ϕ_i is small, then the damage stays small for a long time, followed by rapid growth near the end. This is qualitatively consistent with observation 1 above. The fact that small changes in ϕ_i can cause large variations in T, qualitatively matches observation 2 above. That T is relatively insensitive to the precise form of $g(\phi, P)$ except in the neighborhood of $\phi = 0$, and also insensitive to the failure point ϕ_f so long as it is much larger than ϕ_i, is consistent with the observation that items 1 and 2 above seem to hold for a wide variety of fatigue experiments on components of different materials, shapes, etc.

While keeping in mind that ϕ is not to be thought of as necessarily being the length of a crack, we note in passing that in Paris's Law for crack growth, the exponent of the stress intensity factor range is typically around 3, i.e. greater than 2 (Suresh, 1991). The stress intensity factor for a given load is roughly proportional to the square root of the crack length, for cracks that are small compared to the dimensions of the component, under linear elastic fracture mechanics. This gives an exponent $m(P)$ that is typically around 1.5, i.e. greater than unity.

5.2.5. The forms of the functions $a_0(P)$ and $m(P)$

Ideally, one would use physical theories from materials science and mechanics or some other branch of physics to propose suitable forms for the unknown functions $a_0(P)$ and $m(P)$. However, qualitative reasoning based on experimental data can also be used to deduce fundamental properties of these functions. For example, in this one-variable formulation, the time to failure T is a monotonically

decreasing function of the initial damage ϕ_i, so if ϕ_i has some statistical variation, then the median value of ϕ_i gets mapped to the median value of T. Noting that T is actually a function of both ϕ_i and P in this case, and considering the median value of ϕ_i, Eq. (10) defines the general form of an 'S–N' curve (more properly, a P–T curve) for a system with $m(P) > 1$. This fact allows experimental data to be used to draw further conclusions about $a_0(P)$ and $m(P)$.

To demonstrate the basic idea, suppose that a linear fit to S–N data on a semilog scale is considered appropriate for some range of S. Then in that range (identifying P with S and T with N) we have

$$P = P_0 - \alpha \ln T, \quad \text{or} \quad T = \mathrm{e}^{\left(\frac{P_0 - P}{\alpha}\right)} \tag{13}$$

where $P_0 > 0$ and $\alpha > 0$ are experimentally determined constants for a given specimen type that depend on ϕ_i only, and P is in some suitable range $P_{\mathrm{A}} \leq P \leq P_{\mathrm{B}}$. Using Eqs. (10) and (13), we obtain (explicitly writing the functional dependence of α and P_0 on ϕ_i)

$$a_0(P) = \mathrm{e}^{-\left(\frac{P_0(\phi_i) - P}{\alpha(\phi_i)}\right)} \frac{\phi_i^{1-m(P)}}{m(P) - 1} \tag{14}$$

Since the left hand side of Eq. (14) is independent of ϕ_i, the right hand side must be independent of ϕ_i also. Temporarily introducing the variable $r = \ln \phi_i$ and writing P_0 and α as functions of r instead of ϕ_i, we obtain

$$a_0(P) = \frac{\mathrm{e}^{-\left(\frac{P_0(r) - P}{\alpha(r)}\right) + r\{1 - m(P)\}}}{m(P) - 1}$$

which implies that the exponent in the numerator of $a_0(P)$,

$$-\left(\frac{P_0(r) - P}{\alpha(r)}\right) + r\{1 - m(P)\} \equiv \eta(P), \tag{15}$$

must be independent of r. Differentiating Eq. (15) with respect to P and then r, one finds

$$\frac{\alpha'(r)}{\alpha^2(r)} = -m'(P). \tag{16}$$

Since the left hand side of Eq. (16) is a function of r only, while the right hand side is a function of P only, it must be that both of these are the same constant, say μ_1. Therefore,

$$m(P) = m_0 - \mu_1 P \quad \text{and} \quad \frac{1}{\alpha(r)} = C_0 - \mu_1 r$$

where m_0 and C_0 are constants.[3] Substituting into Eq. (15) and differentiating with respect to P, we find

[3] Note that $r = \ln \phi_i$ is an increasing function of ϕ_i. Assuming $C_0 > 0$ and $\mu_1 > 0$, we see that α is an increasing function of ϕ_i (assuming $\alpha > 0$). Thus, for a batch of several specimens with some statistical variability in ϕ_i, the smaller values of ϕ_i correspond to a shallower line and the larger values of ϕ_i correspond to a steeper line. This leads to greater relative variability, i.e. on a log scale, at high cycle fatigue than at low cycle fatigue, as is observed in practice (Rothbart, 1996). This observation provides some qualitative support for taking $m(P) > 1$.

$$\eta'(P) = C_0 \Longrightarrow \eta(P) = C_0 P + \ln \mu_2,$$

where μ_2 is a constant of integration. Substitution for $\eta(P)$ back into Eq. (14) yields the general form of the governing damage evolution equation for this class of system (scalar damage variable with equilibrium at zero and local power law exponent $m(P) > 1$):

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \frac{\mu_2 \, \mathrm{e}^{C_0 P}}{m_0 - 1 - \mu_1 P} \phi^{m_0 - \mu_1 P}, \tag{17}$$

assumed to be valid for some range of the load parameter $P_A \leq P \leq P_B$. In Eq. (17), m_0, μ_1, μ_2 and C_0 may be looked upon as material constants, in that they are properties of a given type of specimen under a given loading mode, but not dependent on the magnitude of the load or the specimen's initial damage state. These constants can in principle be determined experimentally (using regression) from several sets of multi-level fatigue tests on a batch of 'identical' specimens. In the presence of statistical variability in ϕ_i, median values should be used, as discussed earlier.

The foregoing analysis provides a good example of how qualitative dynamical features of basic quantities of interest can sometimes guide modeling, without recourse to detailed underlying physics (such as considerations of microstructure evolution).

The model developed here can be used for systems that have some of the many features observed experimentally, namely statistical variability in T (which here corresponds to sensitivity to initial damage state), with more scatter at lower loads. However, fatigue limits cannot be modeled. We remark that if the S–N curve for a system is known to be linear for some range of loads on a log–log plot, instead of a semi-log plot as assumed above, then the foregoing analysis still applies if we consider P to be log S.

We close this Section with a remark about the so-called linear damage summation rule, or the Palmgren–Miner rule (Suresh, 1991). For systems that obey Eq. 17 above, with $\mu_1 = 0$, it is easy to show that the Palmgren–Miner rule will hold. If $\mu_1 > 0$, then the rule does not hold.

5.3. Isolated equilibria at nonzero ϕ

Next we consider the case of an isolated equilibrium at a nonzero value of damage $\phi = \phi_0(P)$, as shown in Fig. 1(e). We will show why this case does not seem useful in the context of fatigue damage problems.

Note that if we are only interested in initial damage states $\phi_i > \phi_0(P)$, then we can shift the origin to $\phi_0(P)$ or beyond and use the analyses in Sections 5.1 or 5.2. Thus, the case of an isolated equilibrium at $\phi_0(P)$ is only relevant if we expect $\phi_i < \phi_0(P)$ for at least some cases of interest.

If P is held fixed and the initial damage $\phi_i \leq \phi_0$, then the system never fails. Thus, these models can predict fatigue limits, unlike the cases considered previously in this section. Again considering only the leading term in the power series expansion, we write the damage evolution law near the isolated equilibrium of Fig. 1(e) as

$$\dot{\phi} = a_0(P) \mid \phi - \phi_0(P) \mid^{m(P)} . \tag{18}$$

We expect a family of equilibria $\phi_0(P)$, for $0 \leq \mathrm{P} \leq P_{\max}$. That is, we expect the isolated equilibrium to persist under small changes in P. To see why, let us briefly suppose that the equilibrium disappears under small changes in P, i.e. that $g(\phi_0, P_0) = 0$, but there is no other point (ϕ, P) in some neighborhood of (ϕ_0, P_0) such that $g(\phi, P) = 0$. This means that the system will fail for P both less than as well as more than P_0, yet will not fail for P_0. Such a model is not useful in fatigue, where we expect the system to not fail at loads below P_0. We conclude that, for fatigue applications, the isolated equilibrium ϕ_0 persists for some range of P.

Now we show that such systems predict behavior that contradicts experimental observations. Let there be two loads P_1 and P_2, with $P_2 > P_1$, and with corresponding equilibria ϕ_{01} and ϕ_{02}, respectively.

If ϕ_0 is an *increasing* function of P, then $\phi_{02} > \phi_{01}$. If the system starts with an initial damage ϕ_i such that $\phi_{01} < \phi_i < \phi_{02}$, then for the load P_1 the system fails, while for the *larger* load P_2 the system approaches an equilibrium and does not fail. This feature contradicts fatigue studies. Therefore, $\phi_0(P)$ is not an increasing function.

On the other hand, if ϕ_0 is a *decreasing* function of P, then $\phi_{01} > \phi_{02}$. If the system starts with an initial damage $\phi_i < \phi_{02}$, then the loads P_1 and P_2 are both below the fatigue limit for that system. However, if the system is loaded at load P_1 for a long time, then it approaches the equilibrium ϕ_{01} which is greater than ϕ_{02}. Now, loading at P_2 will cause failure. Thus the fatigue limit for this system is *lowered* by initial loads below the fatigue limit. This possibility is again contrary to what is observed in practice.

This just leaves the possibility that ϕ_0 is a constant independent of P—but in that case no system starting with initial damage less than ϕ_0 ever fails *for any load*. Recall that for systems that start with $\phi_i > \phi_0$ we may as well change variables to $\hat{\phi} = \phi - \phi_0$, which becomes the case in Fig. 1(d) already discussed in the previous Section.

We conclude that isolated equilibria as in Fig. 1(e) are not useful in damage laws, at least for fatigue models.

5.4. A dead zone

Finally, we consider the case where a family of equilibria exists which forms a 'dead zone', that is an entire interval $0 \le \phi \le \phi_0$ for which $g(\phi, P) = 0$ for P in some range, as shown in Fig. 1(f). We will consider this case in some detail, and use a model of this type in section 6 to fit and predict some experimental data.

We mention in passing that such dead zones are used, for example, in crack growth models with experimentally determined 'threshold' stress intensity factors (Suresh, 1991; Liu-Nash et al., 1997). As before, however, we refrain from considering specific physical processes and merely view Fig. 1(f) as a natural candidate to examine in the context of scalar damage laws.

As with the isolated equilibrium case (Section 5.3), it is clear that if P is held fixed and the initial damage $\phi_i \le \phi_0$, then the system never fails. Thus, models with a dead zone can predict fatigue limits. However, in this case the damage remains constant for $\phi_i \le \phi_0$, and does not approach ϕ_0.

5.4.1. The form of $\Phi_0(P)$

If $\phi_0(P)$ is an *increasing* function of P, then there are states of initial damage ϕ_i for which failure occurs for lower P (or lower load) but not higher P (or higher load). Thus, for fatigue applications, we expect $\phi_0(P)$ to be a *decreasing* function.

As with our discussion of the functions $a_0(P)$ and $m(P)$ in section 5.2, one would like to use physical theories to propose suitable forms for $\phi_0(P)$, the unknown fatigue limit function. For example, if we explicitly consider the damage ϕ to be the size of a small crack in a specimen, and use linear elastic fracture mechanics, then the stress intensity factor K is roughly proportional to $\sqrt{\phi}$ and the load (say P), i.e. $K \approx CP\sqrt{\phi}$ for some constant C (Suresh, 1991 and the references therein). If we knew that a minimum 'threshold' stress intensity factor K_{th} is needed for crack growth, then we would use $K_{th} = CP\sqrt{\phi_0}$, or $\phi_0(P) \propto P^{-2}$.

In this paper, we avoid this particular form for several reasons. We do not want the damage variable ϕ to be necessarily taken as a crack length. Moreover, crack length may not be a suitable variable in any case. Crack growth in the initial stages is nonsmooth due to microstructural interactions (James and de los Rios, 1996; Hussain et al., 1993) and use of the stress intensity factor to predict the growth of

small cracks can be problematic even without considerations of microstructure (Kfouri, 1997). Our intention in this paper is to see to what extent purely abstract considerations of qualitative dynamic behavior can lead to models that are useful for fatigue problems.

Accordingly, for $\phi_0(P)$ we tentatively select a simple function with the required monotonicity with respect to P:

$$\phi_0(P) = \Phi_0 \, e^{-\gamma P} + \Phi_1, \tag{19}$$

which as before is assumed to be a reasonable approximation in some range of interest, $P_A \leq P \leq P_B$.

The analysis that follows is easily adapted to different forms of $\phi_0(P)$, if needed.

5.4.2. The form of the damage law

Once again considering only the leading term in the power series expansion, we write the damage evolution law for ϕ near ϕ_0 as

$$\dot{\phi} = a_0(P)(\phi - \phi_0(P))^{m(P)} \quad \text{for} \quad \phi > \phi_0 \quad \text{and 0 otherwise} \tag{20}$$

We now show that such a model can capture several features of systems commonly encountered in the context of fatigue. Consider the plot in the (ϕ, P) plane of a system with a dead zone shown in Fig. 2.

In the figure, the shaded region is the dead zone: in it, all points are equilibria. If the system has an initial damage state of ϕ_1, then so long as the load $P < P_1$, the system is in equilibrium (that is, the damage state is constant). Thus, P_1 is the fatigue limit for the system. However, if the system is loaded at a load level $P_A > P_1$, then damage accumulates. If loading is stopped before failure, at a damage level $\phi_2 > \phi_1$, then the new fatigue limit is $P_B < P_1$.

Just as with the previous cases, the power-law behavior with exponent $m(P) > 1$, for ϕ_i greater than but sufficiently near $\phi_0(P)$, gives the system with a dead zone a time to failure T that is sensitive to initial conditions, but relatively insensitive to the failure point ϕ_f. The variability in T is greater for high-cycle fatigue than for low-cycle fatigue. Moreover, due to the presence of the dead zone, the system has a fatigue limit: loading above the fatigue limit, part of the way to failure, damages the system further and gives a new (lower) fatigue limit.

Directly using Eq. (10) and Eq. (20), one can write the time to failure as

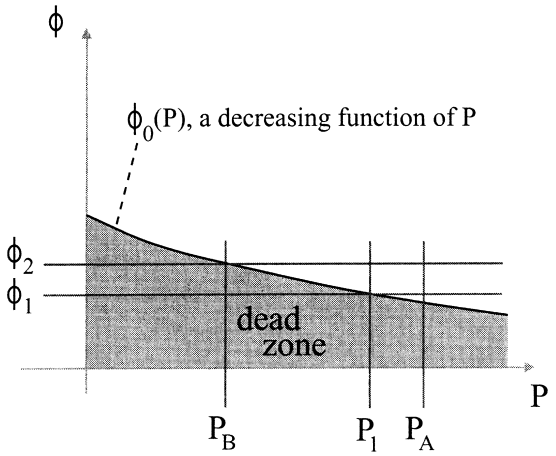

Fig. 2. Plot of the system in Fig. 1(f) showing the variation in $\phi_0(P)$. $\dot{\phi}$ is strictly zero in the shaded 'dead zone', and positive elsewhere. Trajectories of the system for $\phi > \phi_0(P)$ with P fixed move up, parallel to the ϕ axis.

$$T \approx \frac{1}{a_0(P)} \frac{\{\phi_i - \phi_0(P)\}^{1-m(P)}}{m(P)-1} \quad \text{for} \quad \phi_i > \phi_0(P), \quad \text{and } \infty \text{ otherwise.} \tag{21}$$

Using the fatigue limit function of Eq. (19) with Eqs. (17) and (20) as motivation, we now propose

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \frac{\mu_2 \, \mathrm{e}^{C_0 P}}{m_0 - 1 - \mu_1 P} (\phi - \Phi_0 \, \mathrm{e}^{-\gamma P} - \Phi_1)^{m_0 - \mu_1 P} \quad \text{for} \quad \phi > \Phi_0 \, \mathrm{e}^{-\gamma P} + \Phi_1 \quad \text{and 0 otherwise.} \tag{22}$$

The form of Eq. (22) (e.g. the exponent $m_0 - \mu_1 P$) is justified as follows. Much of the analysis in this paper is based on asymptotic considerations of small ϕ_i or small $\phi_i - \phi_0(P)$, where $\phi_0(P)$ may be considered comparable to some ϕ_i of interest, and thus also small. In this case, however, if a specimen is initially loaded part of the way to failure so that the damage ϕ reaches some intermediate range of magnitude, $\phi_0(P) \ll \phi_{\text{int}} \ll 1$, then for the rest of the life of the specimen the effect of $\phi_0(P)$ is negligible. However, since $\phi_{\text{int}} \ll 1$, the local power law behavior still dominates. If we now consider a batch of such specimens, and assume linearity of the S vs log N curves for a range of small initial damage ϕ_{int}, then by the previous analysis the exponent $m(P)$ and the coefficient $a_0(P)$ must have the same forms as before.

The damage evolution model given by Eq. (22) has seven free parameters which must be fitted to experimental data. In addition, the fact that the damage state is (by hypothesis) not directly observable means that the initial condition ϕ_i must also be fitted as an unknown quantity. The model can predict fatigue limits as well as the lowering of fatigue limits on partial loading to failure; shows sensitivity to initial conditions and thus explains the scatter observed in fatigue experiments; shows relative insensitivity to the termination or failure criterion; and, for loads well above the fatigue limit, predicts an approximately linear S-log N (that is, P-log T) curve.

6. Comparison with data from the fatigue literature

In the previous section, we have deduced a suitable form for a model for damage evolution by considering some well known qualitative features of damage dynamics. The damage variable has not been explicitly identified with physical phenomena such as crack growth, nor the load parameter P with physical quantities such as stress, strain, or work of deformation. The qualitative dynamics perspective used in this work is quite general, and is expected to be applicable to a variety of systems with slowly accumulating damage, not just those with fatigue.

Since the formulation is guided by qualitative features of fatigue phenomena, we now present an example of fitting our abstractly-derived model to experimental data. In particular, we compare the data of Golos and Ellyin (1988) to the predictions of our model as given by Eq. (23). Golos and Ellyin's data pertains to multi-level, fully reversed, strain-controlled cyclic loading of carefully prepared low-alloy steel specimens with a circular cross section. For each test, the fraction of life as measured in the final stage of loading was compared to predictions of their model. Here, using some fitted parameters, we compare our predictions to theirs.

6.1. Some remarks on fitting parameters

First of all, we observe that in the fatigue tests of Golos and Ellyin, the time to failure is monitored but the damage variable is never explicitly identified, nor monitored. In other words, there is no transitional data. In such cases, a choice of a damage variable ϕ is indistinguishable from a monotonically increasing function of ϕ, such as (say) $a\phi^2 + b$ for constants a and b.

Given this lack of transitional data, the damage variable is indeterminate to some extent. This is reflected in the fact that the problem of fitting the unknown parameters to experimental data is numerically ill-conditioned.

The implications of the indeterminacy mentioned above, for the design of fatigue experiments and the interpretation of fatigue data, deserve further investigation. These issues, and their interpretation from a physical perspective, will need to be studied in future work. For this work, the indeterminacy in the parameters was removed in two ways that are described in the next two Sections.

We mention that we also separately carried out the fitting procedure *without* eliminating any of these parameters, and the prediction errors obtained were about the same as those presented here (within the precision of available data), though the numerical calculations for fitting the parameters were more delicate. For brevity, those results are not presented here.

6.2. Shifting and scaling the damage variable

We can eliminate two constants from Eq. (22), as follows. In the absence of an explicitly identified damage variable, it is natural to let the decreasing function $\phi_0(P)$ go from unity to zero over the interval of interest.

Let us, as before, confine our attention to a load range $P_A \le P \le P_B$. The new variable ψ is defined via $\Phi_2\,\psi + \Phi_3 = \phi$, where $\Phi_2 := \Phi_0(e^{-\gamma P_A} - e^{-\gamma P_B})$ and $\Phi_3 := \Phi_0\, e^{-\gamma P_B} + \Phi_1$. This gives

$$\frac{d\psi}{d\tau} = \frac{\mu_3 e^{C_1 P}}{m_0 - 1 - \mu_1 P}(\psi - \psi_0(P))^{m_0 - \mu_1 P} \quad \text{for} \quad \psi > \psi_0(P) \quad \text{and 0 otherwise,} \tag{23}$$

where

$$\psi_0(P) = \frac{e^{-\gamma P} - e^{-\gamma P_B}}{e^{-\gamma P_A} - e^{-\gamma P_B}}, \tag{24}$$

and where $\mu_3 = \mu_2 \Phi_2^{m_0 - 1}$ and $C_1 = C_0 - \mu_1 \ln \Phi_2$.

Note that now the scaled initial damage ψ_i is expected to be comparable in magnitude to unity, and the failure point $\psi_f \gg 1$ is treated as infinity.

The fatigue limit function $\phi_0(P)$ now satisfies $\psi_0(P_A) = 1$ and $\psi_0(P_B) = 0$.

6.3. Choice of m_0 in the absence of transitional data

Changes of variables of the form $\hat{\phi} := A\,\phi^b$ for suitable positive constants A, b can be used in Eq. 17, e.g. to obtain a similar equation but with $m_0 > 1$ changed to the new value $1 + (m_0 - 1)/b$ (also greater than 1). If the damage variable ϕ is observable, directly or indirectly, then there may be a physically motivated reason to consider or retain a certain m_0. However, if such observations or *transitional* data are not available, then the parameter m_0 is indeterminate; we can change variables in Eq. 17 so as to choose a specific value for m_0, say $m_0 = 2$.

Recall that for values of P and ϕ such that $\phi_0(P) \ll \phi \ll 1$, the power law behavior as in Eq. 17 is a good approximation to Eq. 23. Based on this reasoning, in this section, we will use $m_0 = 2$.

6.4. Results of parameter fitting and life prediction

In Eq. (23), the load parameter P is identified with % strain, after Golos and Ellyin (1988). Since the data is given in terms of fraction of life, i.e. time scaled with respect to time to failure at each load, the parameters μ_3 and C_1 drop out, and cannot be identified from the fraction-of-life data. As explained

above, we selected $m_0 = 2$. Since the loading range was between 0.15% and 0.80% strain amplitude, we set $P_A = 0.15$ and $P_B = 0.80$. Using a gradient search technique for least squares fitting, the following numerical values were obtained for the remaining three parameters, using 23 out of the 24 two-level tests (see Table 1):

$$\psi_i = 1.17178 \quad \text{(scaled initial damage)}$$

$$\gamma = 9.51244$$

$$\mu_1 = 0.13035$$

Table 1
Comparison with fatigue data of Golos and Ellyin (1988). In the second column, each ordered pair represents a load and time: e.g. '(0.80, 0.25)' represents cyclic loading at 0.80% strain amplitude for 25% of the total life at that load. Two stage load-time sequences are represented by two ordered pairs, three stage by three, etc. The second number in the last of the ordered pairs represents the fraction of life at the last loading stage. The third column shows the prediction of this same number as given in Golos and Ellyin. The fourth column shows the prediction of this same number as given by Eq. (23), with parameters fitted to the two-level test data. See text for details

Test no.	Load-time sequence (% strain, fraction of life)	Predictions of Golos and Ellyin	From Eq. (23)
1	(0.80, 0.80), (0.60, 0.16)	0.18	0.19
2	(0.80, 0.60), (0.60, 0.38)	0.36	0.39
3	(0.80, 0.45), (0.60, 0.57)	0.50	0.54
4	(0.80, 0.25), (0.60, 0.69)	0.70	0.74
5	(0.60, 0.10), (0.80, 0.96)	0.93	0.90
6	(0.60, 0.22), (0.80, 0.82)	0.82	0.79
7	(0.60, 0.50), (0.80, 0.57)	0.54	0.51
8	(0.60, 0.74), (0.80, 0.20)	0.29	0.27
9	(0.60, 0.92), (0.80, 0.30)	0.29	0.09
10	(0.50, 0.80), (0.20, 0.20)	0.11	0.10
11	(0.50, 0.78), (0.20, 0.15)	0.12	0.11
12	(0.50, 0.44), (0.20, 0.30)	0.35	0.37
13	(0.50, 0.34), (0.20, 0.60)	0.44	0.47
14	(0.50, 0.25), (0.20, 0.45)	0.52	0.58
15	(0.50, 0.17), (0.20, 0.66)	0.61	0.69
16	(0.20, 0.27), (0.50, 0.96)	0.92	0.85
17	(0.20, 0.54), (0.50, 0.53)	0.69	0.65
18	(0.20, 0.81), (0.50, 0.34)	0.33	0.34
19	(0.50, 0.88), (0.15, 0.06)	0.04	0.02
20	(0.50, 0.66), (0.15, 0.12)	0.12	0.07
21	(0.50, 0.22), (0.15, 0.34)	0.36	0.34
22	(0.15, 0.56), (0.50, 0.88)	0.86	0.85
23	(0.15, 0.37), (0.50, 0.92)	0.96	0.92
24	(0.15, 0.19), (0.50, 0.97)	0.99	0.97
25	(0.80, 0.50), (0.60, 0.30), (0.20, 0.07)	0.10	0.09
26	(0.80, 0.50), (0.20, 0.18), (0.60, 0.30)	0.16	0.23
27	(0.20, 0.18), (0.60, 0.50), (0.80, 0.62)	0.51	0.42
28	(0.20, 0.18), (0.80, 0.62), (0.60, 0.54)	0.33	0.28
29	(0.80, 0.30), (0.60, 0.19), (0.40, 0.21), (0.20, 0.10)	0.12	0.15
30	(0.20, 0.09), (0.80, 0.10), (0.40, 0.54), (0.60, 0.59)	0.32	0.33
31	(0.80, 0.10), (0.20, 0.50), (0.60, 0.10), (0.40, 0.39)	0.13	0.36
32	(0.20, 0.27), (0.40, 0.27), (0.60, 0.37), (0.80, 0.44)	0.34	0.24

In fitting these parameters, test ♯9 was not used because it appears to contain an error: though the loads were the same as in test ♯8, column 2 of the table states that the specimen for test ♯9 lasted 50% longer at the final loading stage than the specimen for test ♯8, even though it was loaded 24% longer at the first loading stage.

Observe that ψ_i is positive and comparable to 1, $\gamma > 0$, and $m_0 - \mu_1 P > 1$ in the range of interest, as deduced in the previous section. Using these fitted parameters, the fatigue limit $P_{\lim}$ is predicted, by equating $\psi_0(P)$ of Eq. (24) to the fitted value of ψ_i given above and then solving for P, as

$$P_{\lim} = 0.133\% \text{ strain.}$$

Having fitted the parameters to some data, it is necessary to check these fitted values against some *other* data. To this end, we used the fitted parameters to predict the outcomes of the three- and four-level tests (25 through 32). The results are shown in Table 1. We summarize our results as follows.

1. For the fitting data, i.e. tests 1 through 8 and 10 through 24, the r.m.s. value of the difference between experiments and the predictions of Golos and Ellyin (1988) is 0.063. For these same tests, the r.m.s. difference between experiments and predictions from Eq. (23) is 0.066. Also for these same tests, the r.m.s. difference between the two predictions, Eq. (23) and Golos and Ellyin (1988) is 0.037. Finally, we note that the r.m.s. error for these same tests, on using the Palmgren–Miner 'linear damage summation rule', would be 0.192.
2. For the remainder of the data, i.e. the three- and four-level tests, the r.m.s. difference between experiments and predictions of Golos and Ellyin (1988) is 0.169. The r.m.s. difference between experiments and predictions from Eq. (23) is 0.167. The r.m.s. difference between the two predictions, Eq. (23) and Golos and Ellyin (1988), is 0.099. And the r.m.s. error from using the Palmgren–Miner rule would have been 0.249.

Thus, the model of Eq. (23) performs about as well as the model of Golos and Ellyin. The predictions of the two models are closer to each other than they are to the experimental results. Both models (Eq. (23) and Golos and Ellyin) seem to provide significant improvements over the Palmgren–Miner rule, at least for the two-level tests. For the three- and four-level tests, the improvement over the Palmgren–Miner rule is not as marked, but still significant.

We can go further with the data of Golos and Ellyin by using the model of Eq. (23) to estimate the S–N curve for the system, or the 'P–T' curve in our formulation. Using the additional information in Golos and Ellyin (1988), that the number of cycles to failure for 0.15% and 0.80% strain were 535,920 and 1000, respectively, the two remaining parameters μ_3 and C_1 can be estimated, resulting in[4]

$$\mu_3 = 3.78873 \times 10^{-6}, \quad \text{and } C_1 = 6.79216.$$

Using these numerical values, the P–T curve can be generated. Since the initial state of damage ψ_i has been explicitly identified, the effect of uncertainty in initial conditions can also be plotted. Fig. 3 shows the result: the P–T curve itself is shown by the solid line; the dotted line shows the fatigue limit; the chain-dashed lines show the P–T curves for states of initial damage that are 20% higher and lower than the estimated ψ_i.

We remark that the assumed 20% variation in initial damage does not produce as much variation in time to failure as typically seen in fatigue experiments. Within the present scalar damage variable

[4] Observe that μ_3 is a small number: this is because these parameters are fitted to time in loading cycles, which corresponds to fast time. The rate constant ϵ discussed in section 2 can be chosen, for example, as $\epsilon = 10^{-6}$. Then the value of μ_3 corresponding to the slow time will be the constant 3.78873.

formulation, this suggests that the distribution of initial damage states must be larger than 20%, which may contradict one's intuition about the scatter of initial damage states in carefully prepared specimens. However, it must be noted that the initial damage state is *not directly measurable*, and therefore one must acknowledge that any intuition in the matter cannot at this time be based on experimental fact. Furthermore, potential damage nucleation sites in any system are widely distributed, and thus the nucleation process is itself very high-dimensional. It is only after nucleation that the damage localizes and permits a low-dimensional description. Thus the eventual low-dimensional behavior exhibited by the system starts with an effective initial condition ϕ_i which is itself generated by a high-dimensional nucleation process. This nucleation process can be extremely sensitive to initial defects and microstructural details, and therefore it is not inconceivable that the initial scatter in ϕ_i is larger than intuition might initially suggest. However, further discussion of this issue is outside the scope of this paper, and is left for future work.

7. Discussion and conclusions

In this paper, we have developed a vocabulary and mathematical framework within which to model and analyze the nonlinear dynamics of damage evolution. The relationship between the full state space

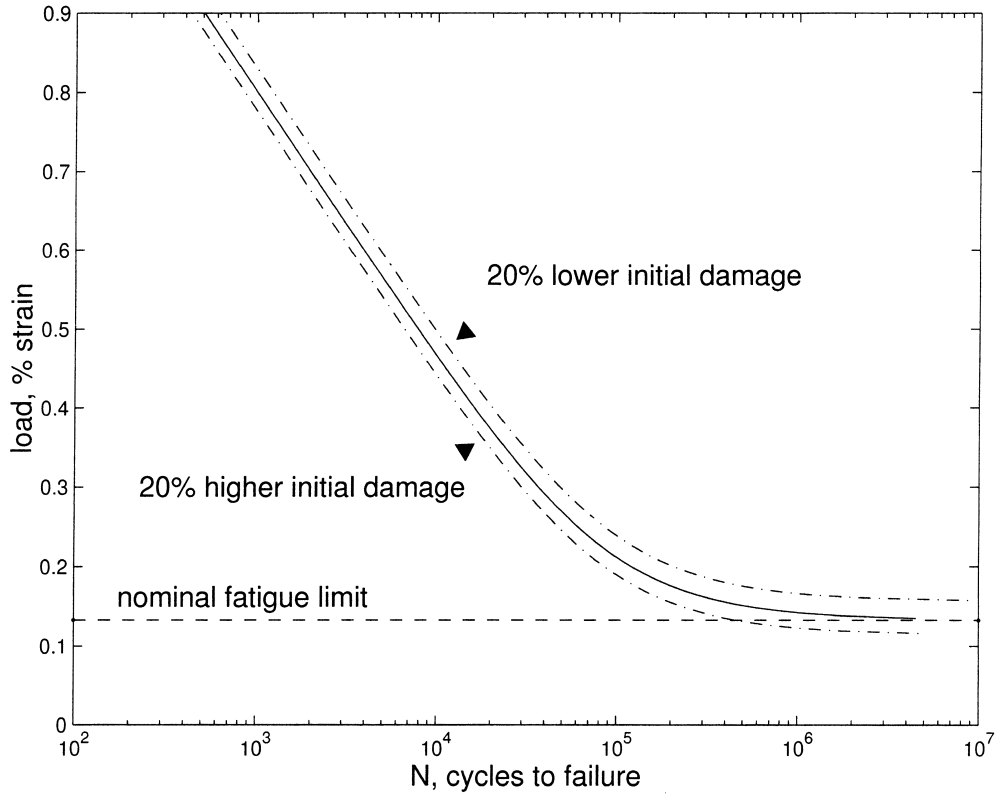

Fig. 3. Predicted P–T curve along with scatter estimates based on $\pm 20\%$ variation in initial damage state. Here, P is % strain, and $T \equiv N$ where N is the number of cycles to failure. Somewhat higher scatter is seen in the high cycle fatigue range, as expected (see footnote 3).

description of hierarchical systems with evolving damage, and damage evolution laws relating the rate of damage accumulation to current damage state and load, has been established using the notion of averaging. The basic elements of a qualitative theory of damage evolution have been presented. This approach emphasizes the relationship between *model structure* and the *qualitative features* of damage dynamics relevant to failure prediction (such as failure surfaces, time to failure T, and variability of T as a function of uncertainty in initial damage state).

We have shown that much can be deduced about the models needed to predict damage evolution using this qualitative approach: in particular, we have performed a detailed analysis of the scalar damage variable case as it pertains to fatigue problems. The notions of a failure surface and initial state sensitivity, together with phenomenological observations from the literature, have been used to analyze the dynamical characteristics of several basic damage evolution laws. A specific model obtained as the result of this analysis, where the damage rate law permits a fatigue limit, was developed and successfully applied to the multi-level fatigue data of Golos and Ellyin (1988). It was found that the prediction using the model of Eq. (23) and the prediction of Golos and Ellyin, which is based on the work of deformation, have nearly identical r.m.s. errors (to experimental precision).

As the title of this paper suggests, the work presented here is only a first step in what is potentially a very general program. There are many issues which need to be addressed in future work. One issue in particular relates to the required dimensionality of the damage state variable. Some physical observations in even simple, uniaxial fatigue tests can apparently be captured in a simple way only if more than one damage variable is introduced. For example, the increase in the fatigue limit of a specimen that is loaded cyclically below its fatigue limit cannot be captured by the model presented here. In fatigue problems where the load is described by more than one parameter, it will also be necessary to consider two or more damage variables (Zhang and Miller, 1996).

In the scalar damage variable model developed here, not all parameters (including the initial damage state) can be determined from a single S–N curve. More information is needed, which can take the form of *transitional* data during the course of an experiment at a single load-level, failure times of multi-level tests (as in this paper), or some combination thereof. More traditional approaches to fatigue failure prediction aim to develop models for which all parameters can be estimated from a single S–N curve (Chaboche and Lesne, 1988). In contrast with our formulation, these conventional approaches restrict the form of the model using additional physical or empirical information. The larger number of free parameters in our model is an inescapable consequence of making no a priori assumptions about the damage physics, and of basing our development *only* on qualitative aspects of the dynamical behavior, as explained in Section 5.2.5.

The explicit formulation of the damage evolution problem in a state space setting is, we believe, ultimately required for true failure prediction of *individual* components or machines, as opposed to statistical or population based failure prediction schemes. The perspective presented here should be interpreted as being complementary to physics, not antithetical to it. On the one hand, qualitative dynamical analysis is a powerful tool for validation of proposed physics-based damage models. On the other hand, physics-based formulations are needed to clarify the physical origins and/or interpretations of the various parameters in our abstractly-derived dynamical models.

Acknowledgements

This work was supported by the Office of Naval Research MURI on Integrated Predictive Diagnostics, Grant N0014-95-0461.

Appendix A

A sample averaging calculation

In general, it will not be possible to analytically carry out the averaging procedure discussed in section 2: all that is required from our point of view is that such an approximation exists in principle. However, in some cases slow flow equations can be explicitly calculated, and to better illustrate the averaging procedure we have selected the following simple example of a forced, damped, harmonic oscillator with a slowly weakening spring:

$$\ddot{y} + c\dot{y} + k(\phi)y = F \sin \omega t \tag{25a}$$

$$\dot{\phi} = \epsilon \phi^2 y^2 \dot{y}^2 \tag{25b}$$

where $k(\phi) := 1 + e^{-\alpha\phi}$ for some positive constant α. Eq. (25a) can be rewritten in first order form to match Eq. (1a), but we proceed directly.

As with Eq. (2), we note that Eq. (25a) with ϕ constant has a well defined steady state behavior regardless of initial conditions, which we write as

$$Y_0(t) = A \sin \omega t + B \cos \omega t$$

for suitable A and B (not reproduced here). Next, as in Eq. (3), the average of the right hand side of Eq. (25b) with respect to the fast time t is found:[5]

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \phi^2 Y_0^2 \dot{Y}_0^2 \, dt,$$

which gives the slow flow equation

$$\frac{d\phi}{d\tau} = \frac{\phi^2 \omega^2 F^4}{8\{(k(\phi) - \omega^2)^2 + \omega^2 c^2\}^2}, \tag{26}$$

where $k(\phi) = 1 + e^{-\alpha\phi}$. The right hand side of the above equation is the $g(\phi, P)$ of Eq. (4). In this case we can take $P \equiv F^4$ as our load parameter.

Numerical solutions to Eqs. (25a) and (25b) for small ϵ and reasonable choices of c, F, ω, and α show good agreement with the predictions of the averaged Eq. (26). In particular, the numerical solutions show that for this case, over time intervals of $\mathcal{O}(1/\epsilon)$, the difference between the averaged solution and the exact solution is $\mathcal{O}(\epsilon)$.

Finally, expanding Eq. (26) in a Taylor series in ϕ results in

$$\frac{d\phi}{d\tau} = \frac{\phi^2 \omega^2 F^4}{8\{(4 - \omega^2)^2 + \omega^2 c^2\}^2} + \mathcal{O}(\phi^3). \tag{27}$$

Thus in this simple example the exponent $m(P)$, discussed in section 5, is a constant independent of P (that is, $m = 2$).

[5] Actually, since the steady-state response is periodic in this case, we might average over one period, but the limit indicated here would be required if the forcing were, say, the quasiperiodic sum of two parts oscillating at incommensurate frequencies, or if transients were included.

References

Chaboche, J.L. 1977. Une loi différentielle d'endommagement de fatique avec cumulation non-linéaire, Revue Francaise de Mécanique, 50–51, 71–82.

Chaboche, J.L., 1988a. Continuum damage mechanics: Part I—General concepts. ASME Journal of Applied Mechanics 55, 59–64.

Chaboche, J.L., 1988b. Continuum damage mechanics: Part II—Damage growth, crack initiation, and crack growth. ASME Journal of Applied Mechanics 55, 65–72.

Chaboche, J.L., Lesne, P.M., 1988. A non-linear continuous fatigue damage model. Fatigue and Fracture of Engineering Materials and Structures 11 (1), 1–17.

Golos, K., Ellyin, F., 1988. A total strain energy density theory for cumulative fatigue damage. Journal of Pressure Vessel Technology 110, 36–41.

Hussain, K., de los Rios, E.R., Navarro, A., 1993. A two-stage micromechanics model for short fatigue cracks. Fatigue and Fracture of Engineering Materials and Structures 44 (3), 425–436.

Hwang, W., Han, K.S., 1986. Cumulative damage models and multi-stress fatigue life prediction. Journal of Composite Materials 20, 125–153.

James, M.N., de los Rios, E.R., 1996. Variable amplitude loading of small fatigue cracks in 6261-T6 aluminium alloy. Fatigue and Fracture of Engineering Materials and Structures 19 (4), 413–426.

Kfouri, A.P., 1997. Limitations on the use of the stress intensity factor, K, as a fracture parameter in the fatigue propagation of short cracks. Fatigue and Fracture of Engineering Materials and Structures 20 (12), 1687–1698.

Kujawski, D., Ellyin, F., 1988. On the concept of cumulative fatigue damage. International Journal of Fracture 37, 263–278.

Lemaitre, J., 1992. A Course on Damage Mechanics. Springer-Verlag, Berlin.

Liu-Nash, G., Todd, J.A., Mostovoy, S., 1997. An accelerated test method for determining near-threshold stress intensity values in HSLA steels. Fatigue and Fracture of Engineering Materials and Structures 20 (12), 1657–1664.

Manson, S.S., Halford, G.R., 1986. Re-examination of cumulative fatigue damage analysis—an engineering perspective. Engineering Fracture Mechanics 25 (56), 539–571.

Natke, H.G., Cempel, C., 1997. Model-Aided Diagnosis of Mechanical Systems: Fundamentals, Detection, Localization, Assessment. Springer-Verlag, Berlin.

Nayfeh, A., Mook, D., 1995. Nonlinear Oscillations. Wiley, New York.

Rothbart, H.A. (Ed.), 1996. Mechanical Design Handbook. McGraw-Hill, New York.

Sanders, J.A., Verhulst, F., 1985. Averaging Methods in Nonlinear Dynamical Systems. Springer-Verlag, New York.

Sun, Y.S., 1994. Revised Miner's Rule and its application in calculating equivalent loads for components. Reliability Engineering and System Safety 43, 319–324.

Suresh, S., 1991. Fatigue of Materials. Cambridge University Press, New York Cambridge solid state science series.

Verhulst, F., 1990. Nonlinear Differential Equations and Dynamical Systems. Springer-Verlag, Berlin.

Zhang, W., Miller, K.J., 1996. A study of cumulative fatigue damage under variable loading-mode conditions. Fatigue and Fracture of Engineering Materials and Structures 19 (2/3), 229–239.

Joseph Cusumano is an Associate Professor in the Department of Engineering Science and Mechanics at Penn State University, where he directs the Nonlinear Dynamics Laboratory. His research focuses on developing methods to model, characterize and predict the behavior of nonlinear dynamic systems. He currently has research projects in the areas of hydrological modeling, vehicle dynamics, machinery failure prediction, and human locomotion. He is the co-founder of *Vis Viva*, an interdisciplinary research group in engineering and the arts.

Anindya Chatterjee got a B.Tech. in Mechanical Engineering at the Indian Institute of Technology at Kharagpur (1989), an M.S. in Engineering Mechanics and an M.S. in Applied Mathematics at the University of Florida at Gainesville (1993), and a Ph.D. in Theoretical and Applied Mechanics at Cornell University (1997). He is presently a postdoctoral scholar at Penn State University. His research interests include a variety of applied problems in nonlinear dynamics.