DASSAULT SYSTEMES

SIMULIA

fe-safe 2018

FATIGUE THEORY REFERENCE

3DEXPERIENCE®

Volume 2 – Fatigue Theory Reference Manual

Contents

1 Introduction to fatigue

1.1 Introduction

Fatigue cracks are caused by the repeated application of loads which individually would be too small to cause failure. Fatigue cracks usually initiate from the surface of a component *(Figure 1.1)*. This is *crack initiation*. The crack may then propagate in a direction perpendicular to the direct stress. This is *crack propagation*. Finally the component may fracture.

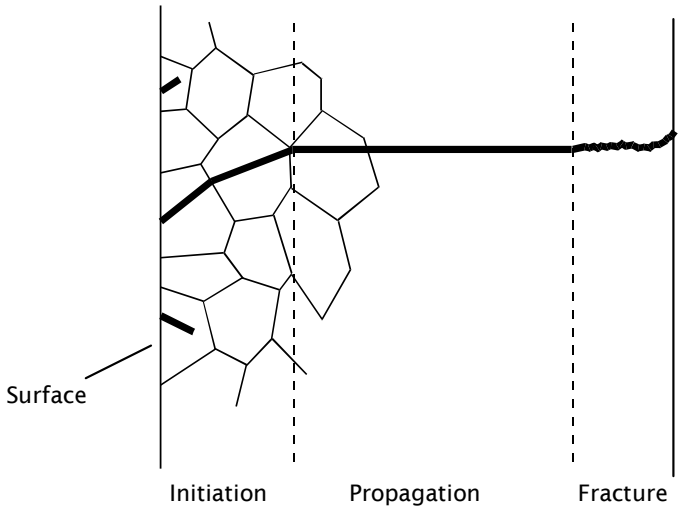


Figure 1.1 The three stages of fatigue failure

Modern fatigue theories provide separate analyses for each phase. Crack initiation theories are based on the assumption that fatigue cracks are initiated by the local strains and stresses on the surface of a component. Crack propagation theories relate crack growth to the stresses in the component. Final fracture is analysed using fracture mechanics. Earlier theories treated the whole of the fatigue life as a single entity, and related fatigue life to the calculated engineering stresses in the component. Much current research is attempting to describe the whole fatigue process by the study of crack propagation from very small initial defects.

Figure 1.2 shows a fatigue fracture from a shaft which was tested in bending. Crack initiation has occurred from the surface (at the top of the picture). The progress of the crack is shown by the bands - called *striations*, or beach marks - which occupy most of the fracture face, with a small area of brittle fracture at the bottom of the picture.

Figure 1.2 Fatigue fracture surface from a shaft in bending

An early example of testing to determine the fatigue strength of a component was reported by W.A.J. Albert in 1838. He investigated the fatigue failure of chains used to haul trucks in mine shafts (Ref 1.1). Examples of the chains were tested using a purpose-built test rig *(Figure 1.2)*. The weight was raised and lowered, applying repeated loads to the links of the chain.

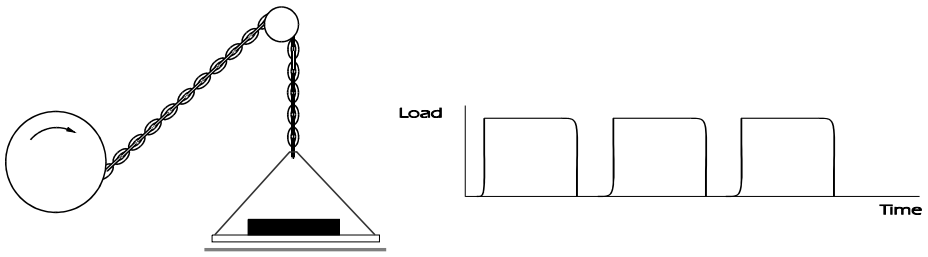


Figure 1.3 Time history of applied loading on the chain

If a new chain is used for each test, and each chain is tested using a different load, the relationship between the magnitude of the applied load and the number of applications to cause fracture of the chain, can be plotted *(Figure 1.4)*. The horizontal axis shows the number of load applications required to cause failure. This is the life of the chain, and is referred to as the *endurance.* It is always plotted on a $\log_{10}$ axis. The vertical axis shows the magnitude of the applied load.

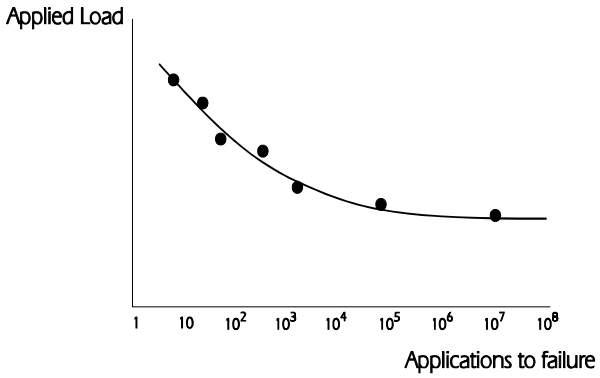


Figure 1.4 Fatigue test results

1.2 Description of the applied loading

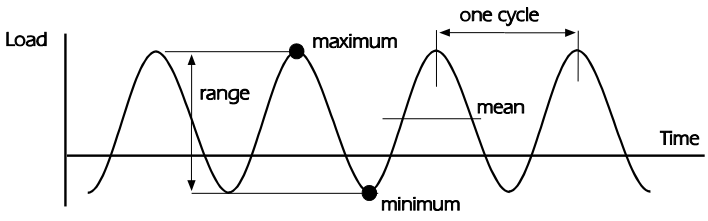


Figure 1.5 Terms used to describe fatigue loading

The following terms are used to describe fatigue loading

maximum load = P_{max}

minimum load = P_{min}

load range $\Delta P = P_{max} - P_{min}$

load amplitude $P_a = \frac{\Delta P}{2}$ (also called alternating load)

mean load $P_m = \frac{P_{max} + P_{min}}{2}$

load ratio $R = \frac{P_{min}}{P_{max}}$ amplitude ratio $A = \frac{P_a}{P_m}$

One application of the load is called a fatigue cycle. The loading used in the chain tests is *constant amplitude loading* because each cycle has the same load amplitude. Commonly used load ratios in fatigue testing are R = 0 where the minimum load is zero, and R = -1 where the minimum and maximum loads are equal and opposite and the mean load is zero.

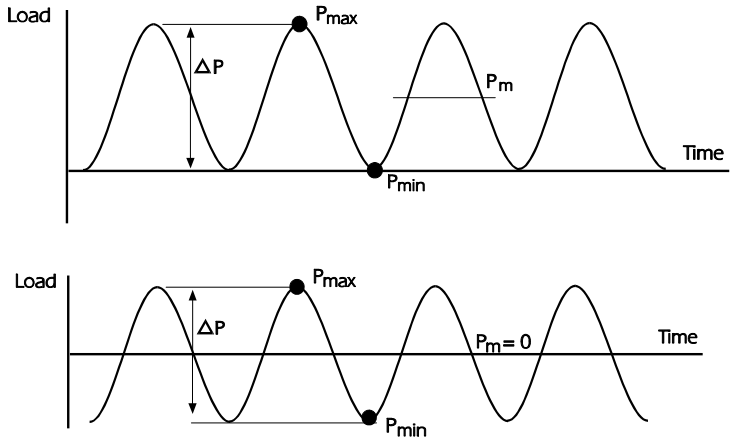


Figure 1.6 Fatigue loading at a load ratio R = 0 And R = -1

1.3 Endurance curves

Endurance curves are plotted with cycles to failure on the horizontal axis and load amplitude on the vertical axis. If $\log_{10}$ scales are used for both axes, the load-life relationship will approximate to a straight line over a large range of endurance, for many components.

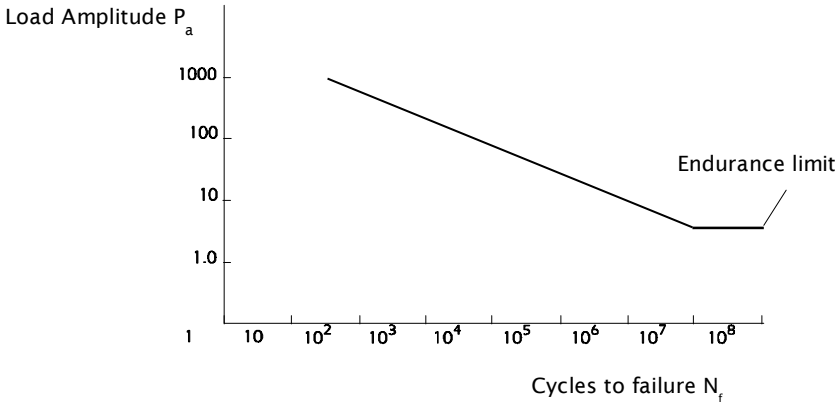


Figure 1.7 Example of an endurance curve

For steels there may be an endurance limit at around 10^7 cycles, implying that cycles with amplitudes smaller than the endurance limit load will not cause fatigue failure, however many times they are applied. Endurance limits are discussed in Chapter 13.

1.4 Generalising fatigue data

An endurance curve which relates load amplitude to endurance for a specific component is not particularly useful - it becomes invalid if the shape of the component is changed, for example. Fatigue endurance curves are generalised to make them more widely applicable. There are three main areas of generalisation :

To allow constant amplitude endurance curves to be used to analyse complex load histories

To allow endurance curves obtained from smooth specimen tests to be used with different shapes of component

To allow an endurance curve obtained by testing one material to be used to calculate fatigue lives for another material - and if possible to estimate fatigue properties of a material without performing fatigue tests.

1.4.1 Analysis of complex signals

The endurance curve in *Figure 1.7* shows the number of applications of a constant amplitude loading which are required to cause failure of the component. Real components are not usually subjected to constant amplitude loading in service. Practical fatigue analysis requires a method of calculating the fatigue life for components which are subjected to variable amplitude loading.

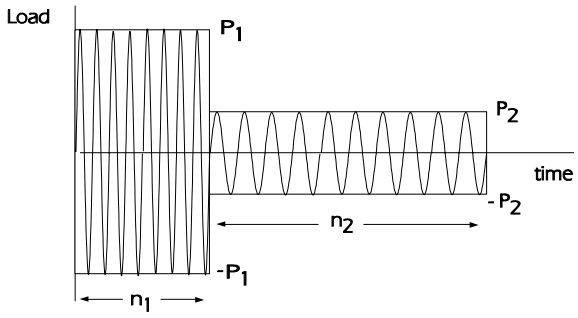


Figure 1.8 Load history with two blocks of constant amplitude loading

Figure 1.8 shows a load history consisting of two blocks of constant amplitude loading. It is assumed that the load history will be repeated until the component fails.

If the loading consists of only the larger amplitude P_{a1}, failure of the component will occur when the number of cycles applied (n_1) equals the number of cycles to failure (N_1) obtained from the endurance curve, ignoring any scatter in test results.

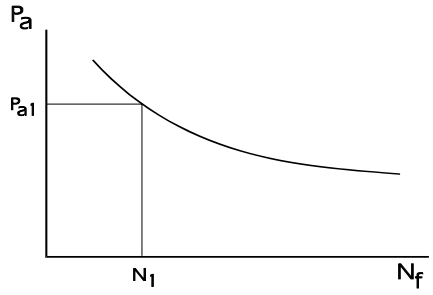


Figure 1.9 Endurance for the larger load amplitude

Clearly, the ratio $\frac{n_1}{N_1}$ has a value of 1.0 at failure.

Similarly, if the loading consists only of the smaller amplitude P_{a2}, failure will occur when the number of cycles applied (n_2) equalled the number of cycles to failure (N_2) obtained from the endurance curve for this amplitude.

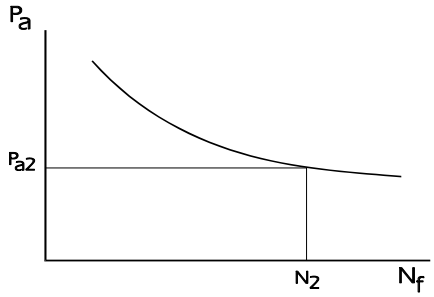


Figure 1.10 Endurance for the smaller load amplitude

The ratio $\frac{n_2}{N_2}$ also has a value of 1.0 at failure.

To calculate the life for the combined loading, it was proposed that if, for either amplitude on its own, failure occurs when $\frac{n}{N} = 1$, then for the combined signal, failure will occur when

$$\frac{n_1}{N_1} + \frac{n_2}{N_2} = 1$$

and for more complex loading, with many different amplitudes, failure will occur when

$$\boxed{\Sigma \frac{n}{N} = 1} \qquad (1.1)$$

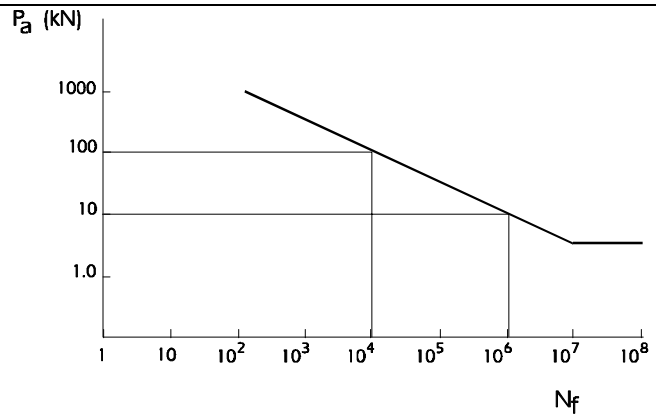
This relationship was first proposed by Palmgren and later by Miner, and is the Palmgren-Miner Cumulative Damage Hypothesis, or Miner's Rule (Ref 1.2).

Example 1.1

The loading on a component consists of

10 cycles of a load amplitude 100 kN
2000 cycles of a load amplitude 10 kN

How many times can the loading be repeated before the component fails ?



It is useful to set up a table for the calculation

Load amplitude P_a kN	Applied cycles n	Endurance N cycles	$\frac{n}{N}$
100	10	10^4	0.001
10	2000	10^6	0.002
			0.003

The total $\Sigma\frac{n}{N}$ = 0.003 . As Miner's Rule states that failure will occur when the summation $\Sigma\frac{n}{N}$ = 1, then failure will occur when the loading has been repeated $\frac{1}{0.003}$ = 333 times. The calculated life is therefore 333 repeats of the load history.

The individual values of $\frac{n}{N}$ are referred to as the *damage* caused by the particular load amplitude, and the summation $\Sigma\frac{n}{N}$ is the *total damage* from one repeat of the load history.

In this example it was a simple matter to determine the number of cycles in the signal, and their amplitudes. Real measured signals are much more complex, and a method of determining the cycles present in real signals is required. This method, *Rainflow cycle counting*, is described in Chapter 4.

The total damage for a complex signal can be calculated from the endurance curve using Miner's Rule. However, recent research has shown that although constant amplitude tests indicate the presence of an endurance limit for many steels when tested in air, in a complex load history containing large and small cycles the cycles with amplitudes below the endurance limit can cause significant fatigue damage. The endurance limit should really be called the *constant amplitude endurance limit*. This is discussed in Chapter 13, together with a more comprehensive treatment of Miner's Rule. Aluminium alloys generally do not have a constant amplitude endurance limit, nor do steels when tested in corrosive environments.

1.4.2 Applying endurance curves to different shapes of component

Instead of expressing the endurance curve in terms of load, it is usual to use a derived parameter such as the stress or strain at some location in the component. Fatigue failures in a component often occur from cracks which initiate at changes of section such as holes, grooves, and fillet radii. This is because these design features produce local stresses and strains which are very much higher than would be calculated by conventional engineering stress analysis. *(Figure 1.11)*

If a fatigue crack is to initiate, some local plasticity must occur during some part of the load history. Modern fatigue theories attempt to relate fatigue endurance to the local stresses and strains which occur at the point of crack initiation, and the effects of local plasticity are included in these theories. These are *critical location*, or *local stress-strain* theories, and are introduced in Chapter 2.

Earlier methods of analysis use a relationship between fatigue life and the engineering stress calculated using conventional stress analysis. The endurance curve shows the relationship between stress amplitude and the number of cycles to failure. This is a stress-life curve, or *S-N curve*. Fatigue analysis based on engineering stresses is introduced in Chapter 5.

For both methods of analysis, materials data is obtained from constant amplitude tests on smooth cylindrical specimens. This data can be applied to components of complex shape using *stress concentration factors*.

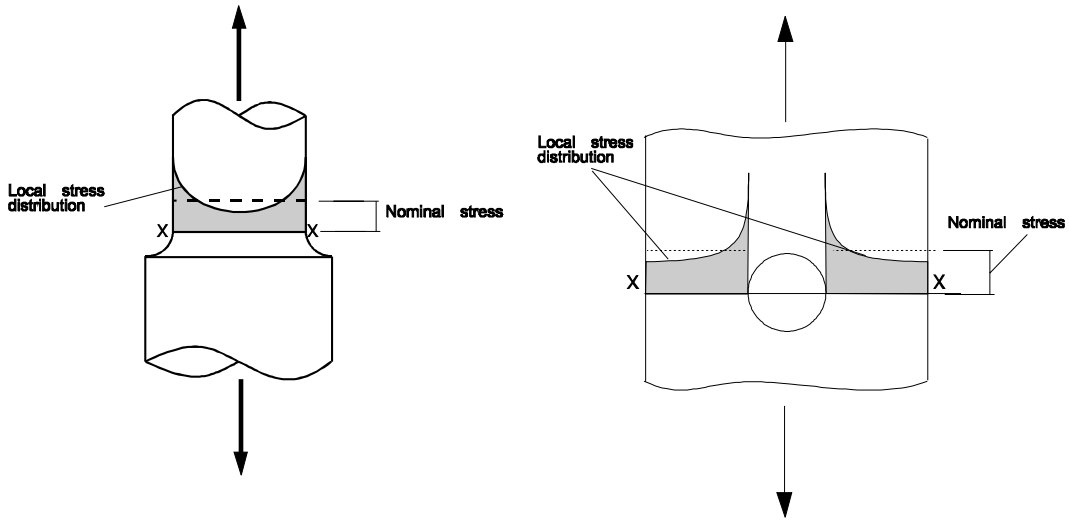


Figure 1.11 Local stresses at typical design details

The section XX in *Figure 1.11* represents a stress concentration, or notch. The average stress across section XX is the reference stress or *nominal stress*. The ratio of the maximum local stress to the nominal stress, if all the stresses are elastic, is the elastic stress concentration factor, K_t .

$$K_t = \frac{\text{local stress at notch}}{\text{nominal stress}}$$

Stress concentration factors have been published for a large number of engineering details (see Chapter 6) and an example for a semicircular groove in a solid shaft in tension/compression is shown in *Figure 1.12*.

In *Figure 1.12* the method of calculating the nominal is defined. Stress concentration factors must always be related to the nominal stress at a specific location - they are factors to be applied to the calculated stress or measured strain at this location. Stress concentrations are described in Chapter 6.

At an endurance close to the constant amplitude endurance limit there will be little plasticity at the notch, and the elastic stress concentration factor can be used to calculate the fatigue strength of a notched component. Using an S-N curve as an example, the endurance limit stress amplitude S_a for the smooth specimen can be divided by the stress concentration factor K_t to provide an estimate of the fatigue strength of the notched component at the same endurance *(Figure 1.13)*.

For many materials the actual fatigue strength of the notched component will be a little higher than this. Determining the actual fatigue strength requires knowledge of the material's *notch sensitivity*. This is discussed in Chapter 6.

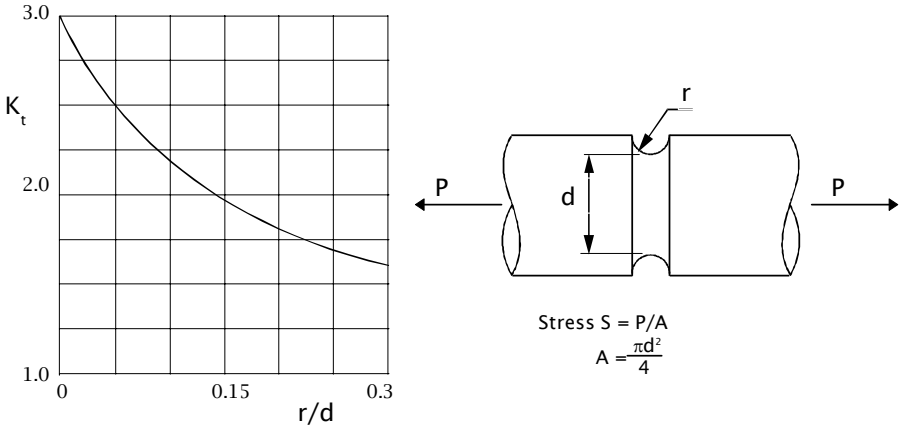


Figure 1.12 Stress concentration factor for a semicircular groove in a solid shaft

At a lower endurance, stresses at the notch will exhibit plasticity. Calculation of the notched fatigue strength at lower endurance is more complex because K_t is an elastic stress concentration factor, and the stresses at the notch will exhibit plasticity. Both local strain and engineering stress fatigue methods contain equations for calculating the effect of notches when the stresses are elastic-plastic.

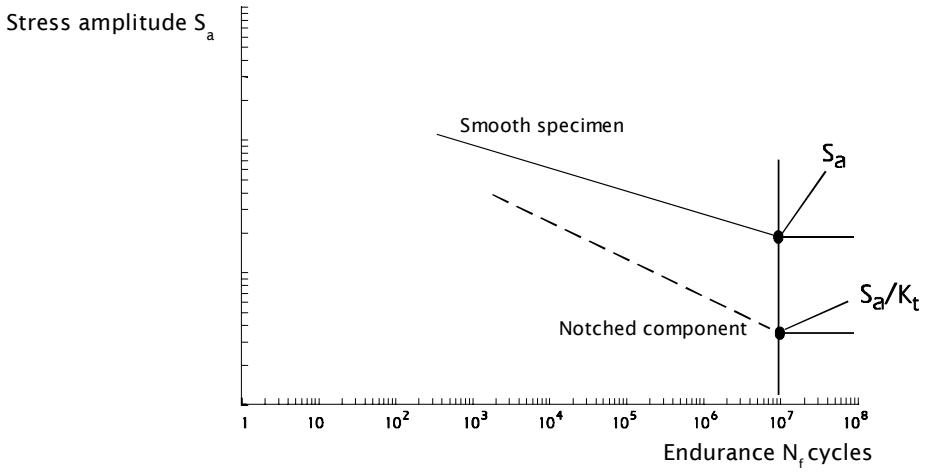


Figure 1.13 Endurance limit for a notched component

1.4.3 Effects of mean stress

Example 1.1 used a fatigue loading consisting of blocks of constant amplitude loading, at a mean load of zero. The mean stress also influences the fatigue life, as tests at a high mean stress will have a shorter fatigue life for the same amplitude than tests at a lower mean stress. Methods of correcting for mean stress effects (*mean stress correction*) are described in Chapters 2 and 5.

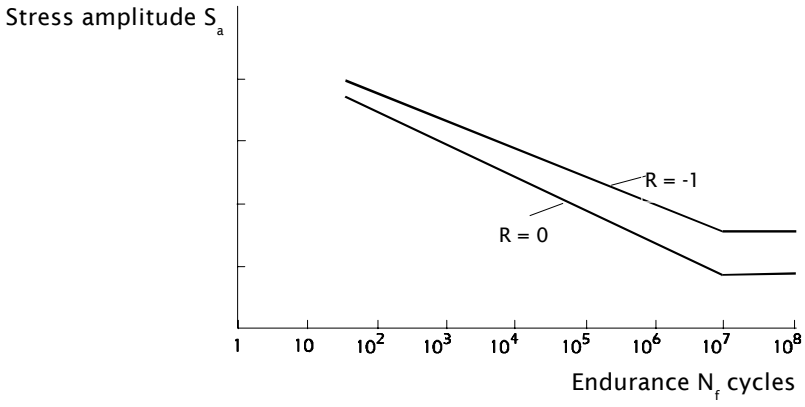


Figure 1.14 Effect of mean stress on fatigue life

1.5 Other factors which influence fatigue strength

Fatigue cracks usually initiate from the surface of a component, where fatigue damage initiates as shear cracks on crystallographic slip planes.

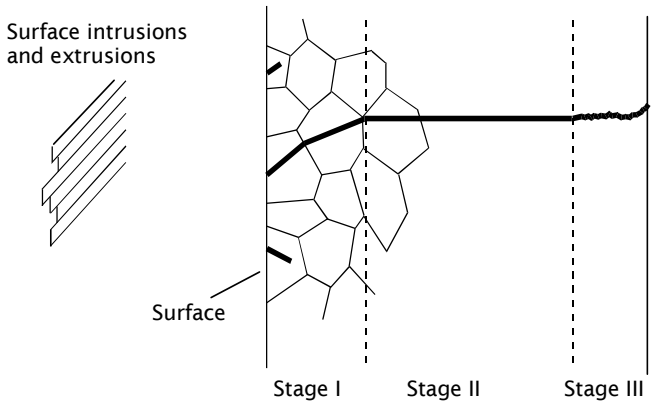


Figure 1.15 The three stages of fatigue crack growth

The surface shows the slip planes as intrusions and extrusions. This is Stage I crack growth. Stage II crack growth takes place in a direction normal to the applied stress. Finally the crack becomes unstable and fracture occurs.

For many components Stage I growth occupies most of the fatigue life. This has led to the study of fatigue as a continuous process of crack growth from very short cracks, replacing endurance curves with crack growth calculations. Some of the methods are now very advanced (see for example Ref 1.3) and it is possible that in the future all fatigue calculations will be made this way. However the methods are not yet in a form which allows them to be used for every-day fatigue design and assessment, and so they are not included in this book.

The life to fatigue crack initiation has been found to be a function of the surface stress or strain combined with a critical feature such a weak grain boundary. The volume of highly stressed material near the surface influences the probability of the stresses interacting with a critical feature. For components subject to bending stresses, a large component will have a less steep stress gradient than a small component, and therefore a greater volume of material will be highly stressed *(Figure 1.16)*. Large components have been shown to have shorter fatigue lives than smaller components subject to the same bending stress at the surface. Axially loaded components will have a uniform stress distribution, and a shorter fatigue life than components in bending. The difference is quite significant at high endurance. The constant amplitude fatigue limit stress for axially loaded steel specimens may be only 70% of the endurance limit stresses for specimens tested in bending.

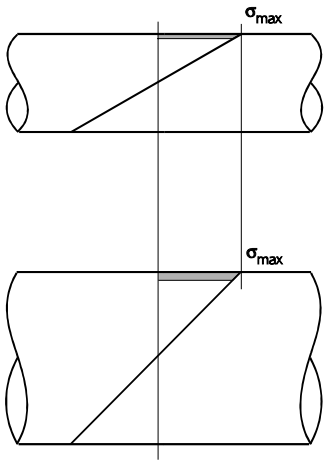


Figure 1.16 Volume of stressed material near the surface
of small and large specimens tested in bending

Because crack initiation is a surface phenomenon, any process which effects the surface may have a significant influence on fatigue strength. Important factors are:

The quality of the surface finish-ground, for example machined, rolled, as-cast, as-forged.

Surface treatments such as plating or cladding.

Residual stresses introduced during manufacture either by the fabrication process or by special treatments such as shot peening or rolling.

The operating environment, for example corrosive gases, salt sprays, and operating temperature.

Fretting between adjacent surfaces.

The effect of surface finish is shown in *Figure 1.17*.

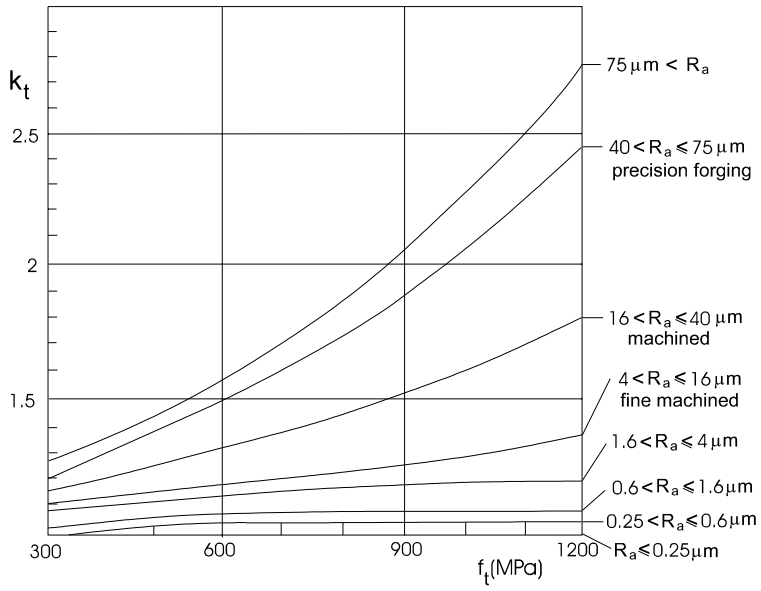


Figure 1.17 Surface finish correction factor

$K_{surface}$ may be used to multiply the geometric stress concentration factor K_t to give the combined effect of geometry and surface finish.

The effect of environmental factors on fatigue strength is very much dependent on the material and on the precise nature of the surface treatment or the operating environment. For practical fatigue calculations where environmental effects are to be included, reference must be made to research papers. To illustrate their importance, examples of the effect of environment and surface treatment on fatigue strength are shown in *Figure 1.18* and *Figure 1.19*.

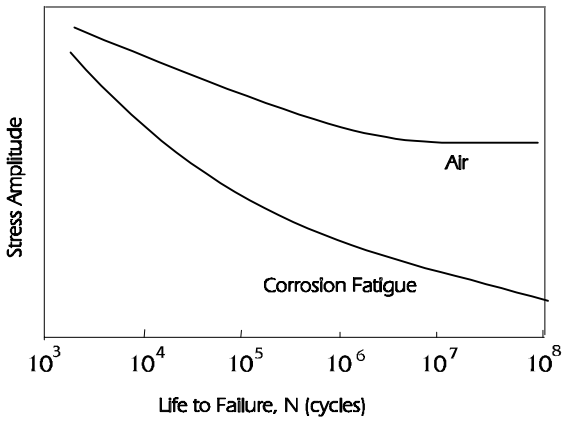


Figure 1.18 Example of the effect of environment on fatigue strength

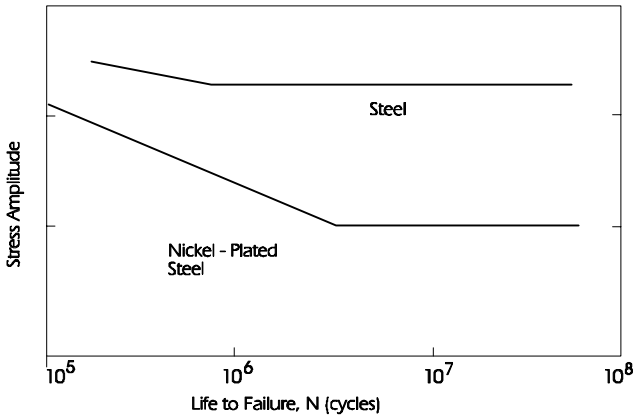


Figure 1.19 Effect of nickel plating on the fatigue strength of a steel

Fretting reduces fatigue strength at an endurance greater than 10^4 - 10^5 cycles. An example for an austenitic steel is shown in *Figure 1.20*. The extent of the reduction in fatigue strength is a function of the materials in contact, and the contact pressure. Ref 1.4 provides a comprehensive treatment of fretting fatigue and a bibliography.

The effects of residual stresses are usually limited to high endurance because the local yielding associated with low cycle fatigue will eliminate the residual stresses introduced by shot peening or rolling.

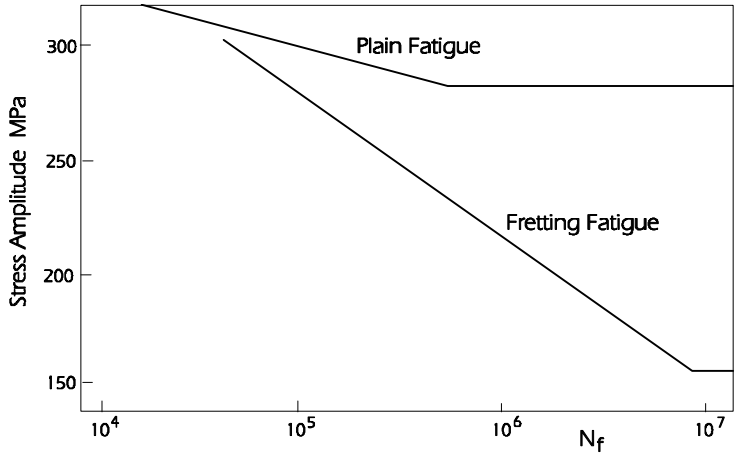


Figure 1.20 Effect of fretting on the fatigue strength of a steel

The cycle frequency of the applied loading, and the waveform, can be ignored in many fatigue calculations, and tests at frequencies of 20Hz and 20kHz have shown little difference in the total cycles to failure. Exceptions are corrosive environments, high temperature fatigue, and situations where the initiating crack is subjected to internal pressure (for example in hydraulic or pneumatic components).

Finally, and of great importance, smooth specimen test data is almost always obtained from tests on uniaxially stressed specimens. Biaxial stresses occur in many components, for example shafts and axles in combined bending and torsion, and can also occur at stress concentrations in components which are loaded uniaxially. The fatigue lives of biaxially stressed components can be orders of magnitude shorter than would be indicated by uniaxial endurance curves. Fatigue analysis for biaxial stresses is described in Chapter 7.

1.6 Chapter contents

The subjects are considered in the following sequence.

Chapter 2 considers fatigue crack initiation under local surface strains. This section includes analytical methods for the effects of stress concentrations and the effects of mean stress. It also shows which events in a load history constitute fatigue cycles.

Chapter 3 describes the materials data needed for local strain analysis.

Chapter 4 uses the methods developed in Section 2 to describe Rainflow cycle counting, and compare it with other methods of signal analysis.

Chapter 5 describes fatigue analysis methods using engineering stress - the S-N curve approach. It shows the empirical methods developed to describe the fatigue phenomena now treated more analytically by local stress-strain methods.

Chapter 6 describes stress concentrations and notch sensitivity in more detail.

Chapter 7 describes the theories for the analysis of fatigue under biaxial stresses.

Chapter 8 describes recent advances in the use of biaxial fatigue methods to provide designers with fatigue analysis techniques from finite element models.

Chapter 9 provides a brief description of fracture mechanics and crack growth calculations.

Chapter 10 describes the statistical analysis of fatigue data.

Chapter 11 describes the fatigue analysis of welded joints. This requires the information from chapters 9 and 10.

Chapter 12 describes recent work in the use of power spectra for fatigue analysis.

Chapter 13 looks more closely at the endurance limit and the effect of small cycles in real service loading, and describes the development of representative test signals.

Chapter 14 concludes with some guidance on practical fatigue investigation

1.7 References

1.1 Watson P, Hill S

Fatigue Life Assessments of Ground Vehicle Components

Design of Fatigue and Fracture Resistant Structures, ASTM STP 761, 1982, pp. 5-27 Abelkis P.R. and Hudson C.M., Eds.

1.2 Miner M A

Cumulative Damage in Fatigue

J Appl Mech, Vol 12, Trans ASME Vol 67, pp A159-A164, 1945

1.3 Miller K

Metal fatigue – past, current and future.

Proc. IMechE Vol 205, 1991

1.4 Waterhouse R B (editor)

Fretting Fatigue

Applied Science Publishers Ltd, Barking, Essex UK, 1981. ISBN 0-85334-932-0

2 Uniaxial strain-life fatigue analysis

2.1 Introduction

In Chapter 1 it was shown that fatigue cracks often initiate from stress concentrations caused by geometric shapes such as holes and fillet radii. Local stress-strain fatigue analysis presumes that the life to initiation of a small crack is determined by the sequence of stresses and strains developed in a small volume of material at the stress concentration. Consequently if the same stress-strain sequence is reproduced on a smooth specimen of the same material, the same fatigue life will be obtained.

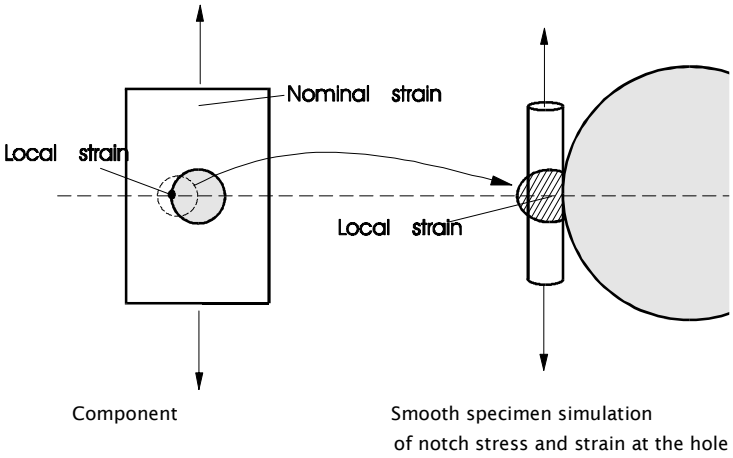


Figure 2.1 Fatigue life based on reproduction of true local stress - strain

Although many engineering components are designed so that stresses and strains are below the elastic limit under normal service loading, yielding can occur at a local stress concentration, and indeed must do so if fatigue cracks are to initiate. The application of strain-life analysis requires a description of the material response to cyclic elastic-plastic strains, and a relationship between these strains and fatigue endurance. This chapter describes the strain-life equations, shows how fatigue lives can be calculated from a sequence of local strains, and shows how stress concentration factors can be used to calculate fatigue lives for components which contain notches. This approach to fatigue analysis is referred to variously as local strain-life, local stress-strain, or critical location analysis.

Local strain-life methods are attractive for practical fatigue investigations where strains can be measured using strain gauges. Finite element models also give local stresses and strains at each location in the model, so local strain-life methods are very appropriate for fatigue design using finite element models.

The stress and strain at the critical location are termed the local stress (σ) and the local strain (ε). Stress and strain remote from the notch and unaffected by it are the nominal stress (S) and nominal strain (e).

2.2 True stress and strain

When a cylindrical test specimen is loaded in tension, it increases in length and reduces in diameter.

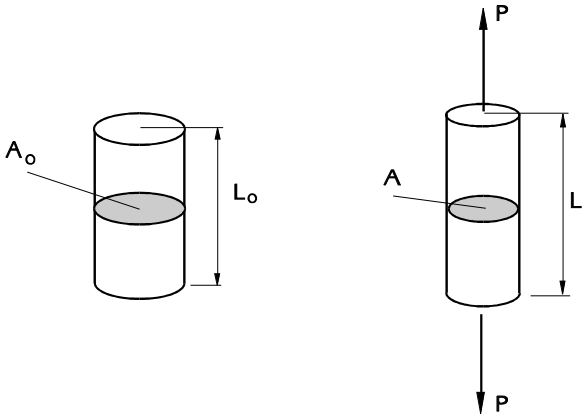


Figure 2.2 Engineering and true stress-strain

Engineering stress is the applied load divided by the original cross-section area.

$$S = \frac{P}{A_o} \qquad (2.1)$$

True stress is defined as the applied load divided by the actual cross-section area with the load applied.

$$\sigma = \frac{P}{A} \qquad (2.2)$$

Engineering strain is the total change in gauge length divided by the original gauge length.

$$e = \frac{(L - L_o)}{L_o} \qquad (2.3)$$

True strain is calculated from the instantaneous gauge length

$$\varepsilon = \int_{L_o}^{L} \left(\frac{\delta L}{L}\right) = ln\left(\frac{L}{L_o}\right) \qquad (2.4)$$

True strain is the strain which would be measured with a strain gauge. Because true strains are calculated as increments from the instantaneous strain, they can be added and subtracted incrementally during the analysis of a service strain history, whereas engineering strains cannot.

The difference between engineering and true stress-strain can be illustrated by the conventional tensile test. This is usually plotted on the basis of engineering stress and strain, and shows a reduction in load and stress when necking of the specimen occurs. *Figure 2.3* shows the tensile test plotted as engineering stress-strain, and as true stress-strain.

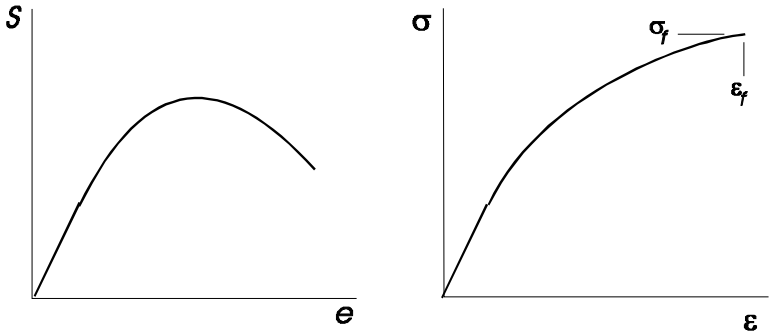


Figure 2.3 Monotonic tensile test plotted as engineering and true stress-strain

The true stress increases until failure occurs. (Note however that the simple relationships given by equations (2.1 to 2.4) are only valid up to necking. After this, the stress is not uniformly distributed across the section).

If A_f is the cross-sectional area at fracture, and σ_f is the true stress at fracture

$$\sigma_f = \frac{P}{A_f} \qquad (2.5)$$

and the true strain at fracture ε_f is

$$\varepsilon_f = ln\left(\frac{A}{A_f}\right) \qquad (2.6)$$

The true stress - true strain curve obtained from a single load application is called a monotonic curve. It consists of an elastic portion which is recovered when the load is removed

$$\varepsilon_e = \frac{\sigma}{E}$$

and a plastic portion which is not recovered, and which is given by the equation

$$\varepsilon_p = \left(\frac{\sigma}{K}\right)^{\frac{1}{n}}$$

The total strain is the sum of the elastic and plastic strains :

$$\varepsilon = \frac{\sigma}{E} + \left(\frac{\sigma}{K}\right)^{\frac{1}{n}} \qquad (2.7)$$

This relationship, proposed by Ramberg and Osgood, uses the following terms

E	the elastic modulus (Young's Modulus)
K	the strain hardening coefficient
n	the strain hardening exponent

Figure 2.4 shows two alternative methods of illustrating the elastic and plastic components of the total strain.

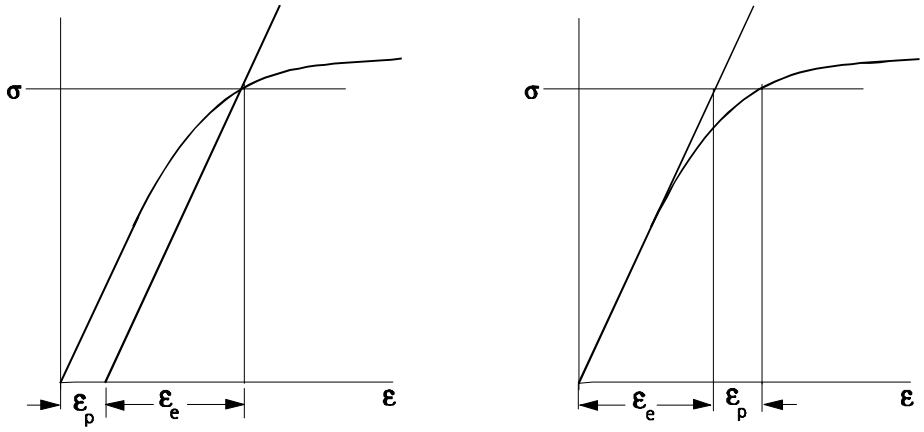


Figure 2.4 Total strain as the sum of elastic and plastic strains

2.3 Fatigue life relationships

If a cylindrical specimen is loaded in tension, then compression, and then tension again, and yielding occurs at each load application, the material response is a hysteresis loop of true stress and true strain *(Figure 2.5)*.

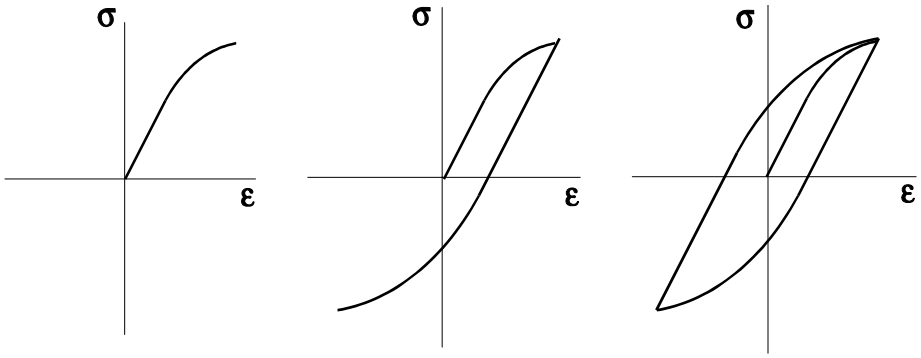


Figure 2.5 Cyclic behaviour as true stress-strain

(In practice many materials do not behave as simply as this, and exhibit some initial softening or hardening before the shape of the hysteresis loop stabilises. This is described in Chapter 3.)

A smooth test specimen can be cycled between fixed strain limits until it develops a fatigue crack. If several specimens are tested under different constant amplitude strain cycles, the test lives can be plotted on the basis of the true stress amplitude. If the axes are $\log_{10}$ (stress amplitude) and $\log_{10}$ (endurance), the test points will approximate to a straight line.

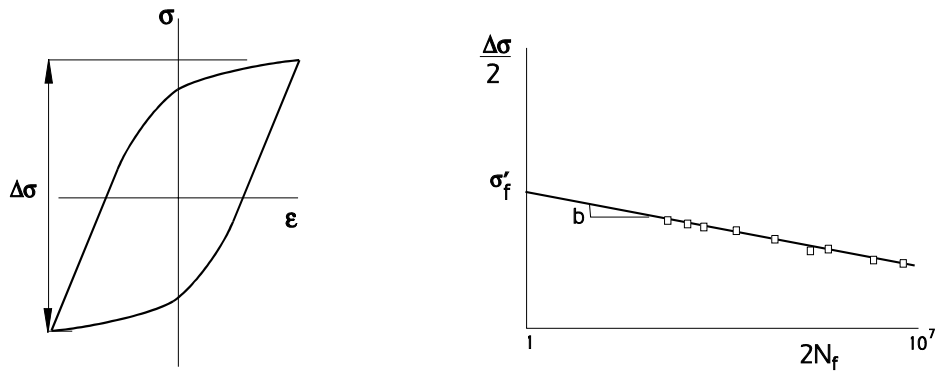


Figure 2.6 True stress amplitude vs endurance

Note that the endurance is not expressed as *N* cycles, but as $2N_f$ *half-cycles*, or reversals. **N_f is the number of cycles, $2N_f$ is the number of reversals.**

The equation for the relationship between stress amplitude and endurance is

$$log_{10}\frac{\Delta\sigma}{2} = b\, log_{10}(2N_f) + log_{10}(\sigma'_f) \qquad (2.8)$$

or

$$\boxed{\frac{\Delta\sigma}{2} = \sigma'_f(2N_f)^b} \qquad (2.9)$$

This linear relationship was first proposed by Basquin in 1910 (Ref. 2.1). The intercept σ'_f at $2N_f = 1$ reversal is the fatigue strength coefficient, and the slope b is the fatigue strength exponent (Basquin's exponent).

The test results can also be plotted as true strain amplitude versus endurance, again on log_{10} - log_{10} axes.

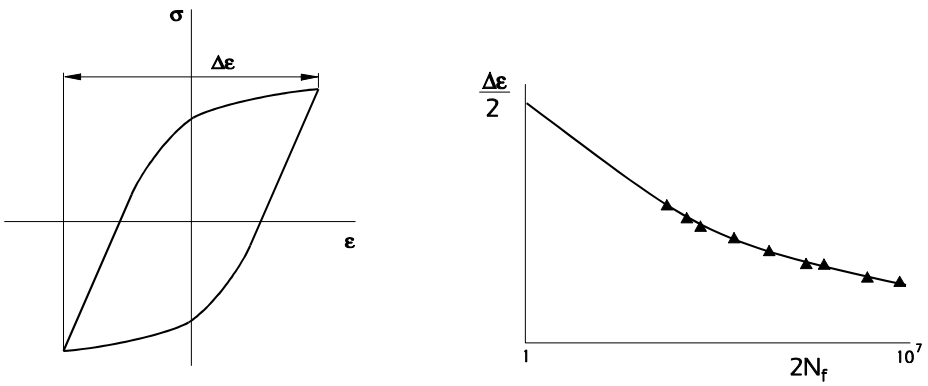


Figure 2.7 Total strain amplitude vs endurance

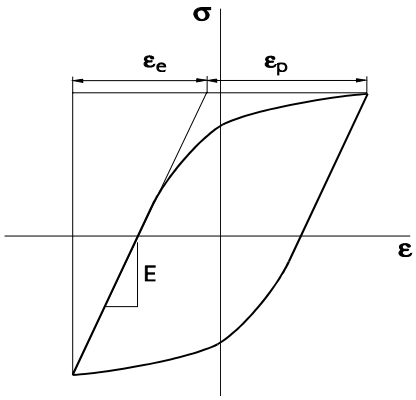


Figure 2.8 The cyclic stress-strain hysteresis loop
as the sum of elastic and plastic strains

The equation for the strain-life relationship can be obtained by considering the elastic and plastic components of the total strain *(Figure 2.8)*.

The stress and elastic component of strain are related by

$$\Delta\varepsilon_e = \frac{\Delta\sigma}{E} \qquad (2.10)$$

As there is a linear relationship between true stress and elastic strain, then by comparison with the stress-life relationship (eqn 2.9) the graph of elastic strain vs endurance is a straight line on axes of $\log_{10}$ (elastic strain amplitude) and $\log_{10}$ (endurance), with the same slope as the stress-life relationship shown in *Figure 2.6*.

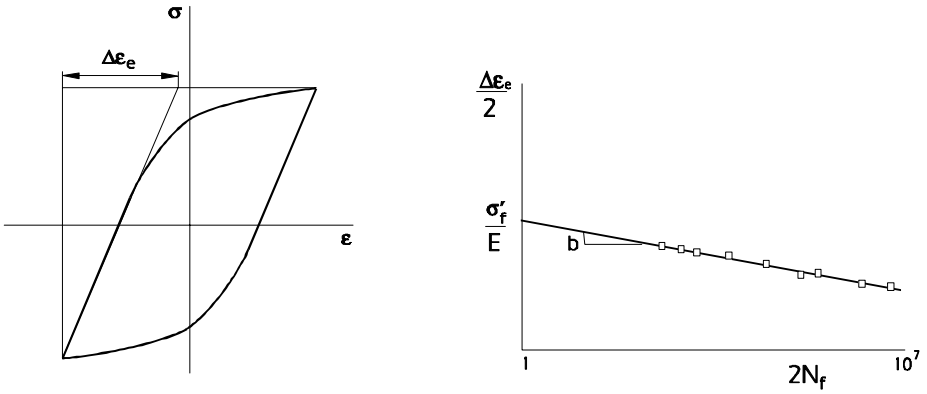


Figure 2.9 Relationship between elastic
strain amplitude and endurance

Dividing both sides of equation (2.9) by the elastic modulus E gives

$$\frac{\Delta\varepsilon_e}{2} = \frac{\sigma'_f}{E}(2N_f)^b \qquad (2.11)$$

Manson and Coffin (Ref. 2.2) showed that the relationship between plastic strain amplitude and endurance could also be expressed as a straight line on $\log_{10}$-$\log_{10}$ axes *(Figure 2.10)*. The equation is

$$log_{10}\frac{\Delta\varepsilon_p}{2} = c\,log_{10}(2N_f) + log_{10}(\varepsilon'_f) \qquad (2.12)$$

so that

$$\frac{\Delta\varepsilon_p}{2} = \varepsilon'_f(2N_f)^c \qquad (2.13)$$

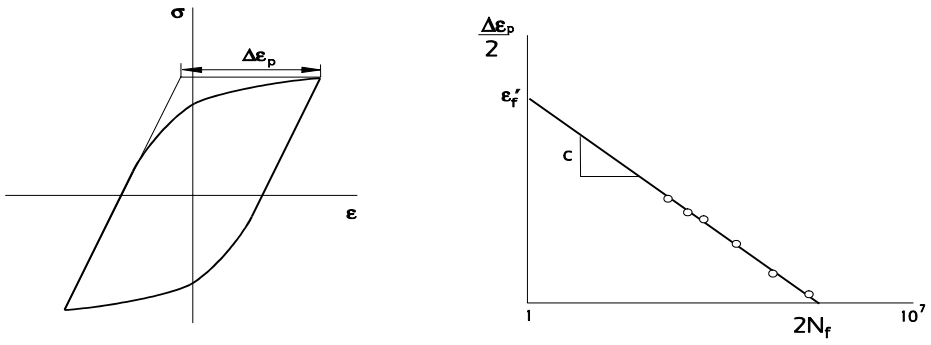


Figure 2.10 Relationship between plastic strain amplitude and endurance

The total strain is the sum of the elastic and plastic strains,

$$\Delta\varepsilon = \Delta\varepsilon_e + \Delta\varepsilon_p \qquad (2.14)$$

and in terms of strain amplitude

$$\frac{\Delta\varepsilon}{2} = \frac{\Delta\varepsilon_e}{2} + \frac{\Delta\varepsilon_p}{2} \qquad (2.15)$$

so that the strain-life relationship, from equations 2.11 and 2.13 is

$$\boxed{\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c} \qquad (2.16)$$

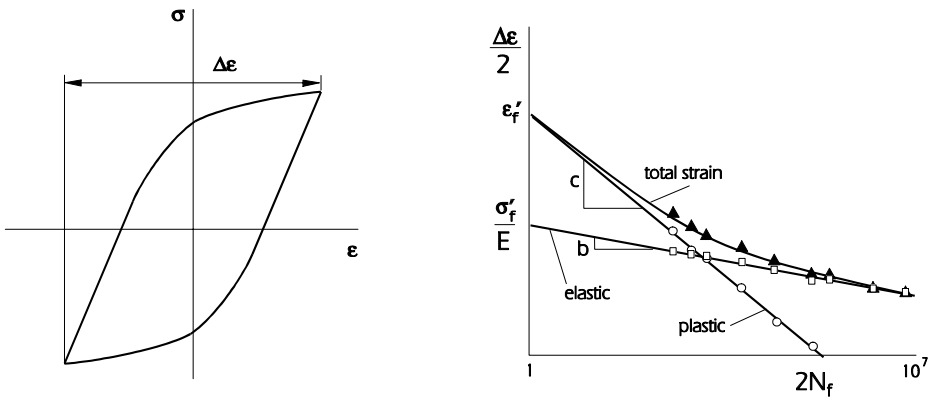


Figure 2.11 Relationship between total strain amplitude and endurance

The constants in this equation are

b	the fatigue strength exponent (Basquin's exponent)
σ'_f	the fatigue strength coefficient
c	the fatigue ductility exponent (the Coffin-Manson exponent)
ε'_f	the fatigue ductility coefficient, which is the plastic strain amplitude at $2N_f=1$
E	the elastic modulus

The test lives used to determine the stress-life and strain-life equations are the lives to a small crack, typically 1mm in length, in a smooth cylindrical test specimen. Local strain-life analysis is a method of calculating fatigue lives to the initiation of a small crack, and is a *crack initiation* criterion. The test procedures used to obtain the materials properties are described in Chapter 3.

2.4 Cyclic stress-strain relationships

The stable cyclic stress-strain hysteresis loops from a number of test specimens can be plotted, and a curve constructed through the tips of the hysteresis loops. *Figure 2.12* shows hysteresis loops from ten specimens, each tested at a different constant strain amplitude. The stress-strain curve constructed through the loop tips is called the stable cyclic stress-strain curve.

It is a rather artificial concept which is intended to represent the stable material response, after initial cyclic softening or hardening, to a strain application from a state of zero stress-strain.

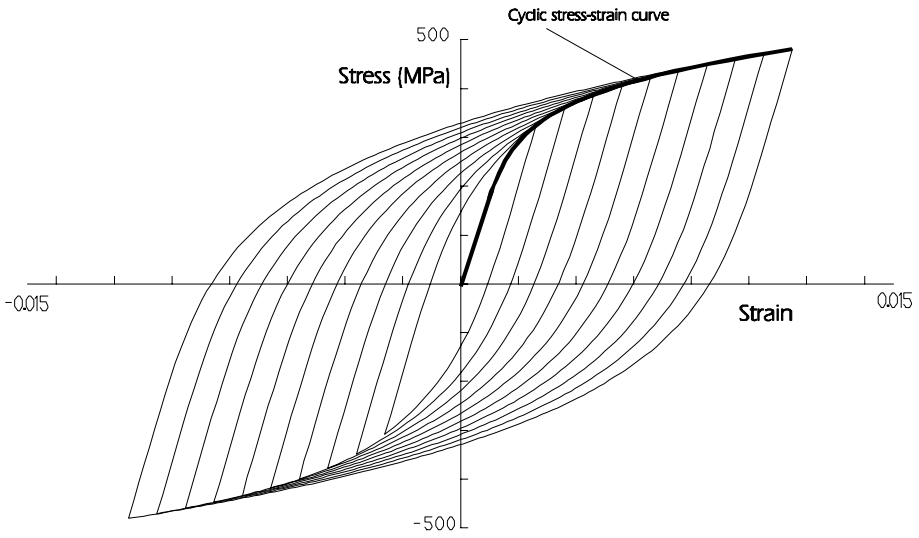


Figure 2.12 Stable cyclic stress-strain hysteresis loops

The cyclic stress-strain curve consists of an elastic component

$$\varepsilon_e = \frac{\sigma}{E^*} \qquad (2.17)$$

and a plastic component

$$\varepsilon_p = \left(\frac{\sigma}{K'}\right)^{\frac{1}{n'}} \qquad (2.18)$$

The primes are added to denote cyclic conditions, and the full equation for total strain is

$$\varepsilon = \frac{\sigma}{E^*} + \left(\frac{\sigma}{K'}\right)^{\frac{1}{n'}} \qquad (2.19)$$

where

E^*	is the cyclic elastic modulus
K'	is the cyclic strain hardening coefficient
n'	is the cyclic strain hardening exponent

The cyclic elastic modulus often has a value close the monotonic elastic modulus (Young's modulus), so that $E^* = E$.

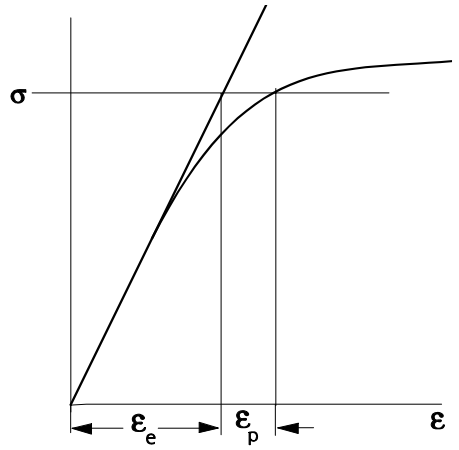


Figure 2.13 Cyclic stress-strain curve as the sum of elastic and plastic strains

For many metals, the shape of the hysteresis loop may be obtained by re-scaling the cyclic stress-strain curve by a factor of 2, in effect a photographic enlargement. This follows from the observation that a tensile strain followed by an equal and opposite compressive strain will produce equal values of tensile and compressive stress in a homogeneous material. *(Figure 2.14)*.

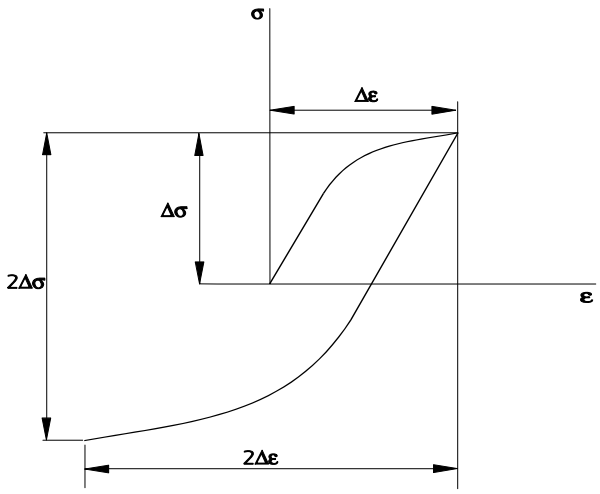


Figure 2.14 Symmetrical behaviour in tension and compression.

The hysteresis loop curve is sometimes referred to as 'twice the cyclic stress-strain curve', and the 'factor 2' relationship is Masing's hypothesis (Ref 2.3).

In *Figure 2.15*, the 'factor 2' enlargement means that a point at X on the cyclic stress strain curve becomes the point Y on the hysteresis loop curve, so that

$$\sigma_x = \frac{\sigma_y}{2} \quad \text{and} \quad \varepsilon_x = \frac{\varepsilon_y}{2}$$

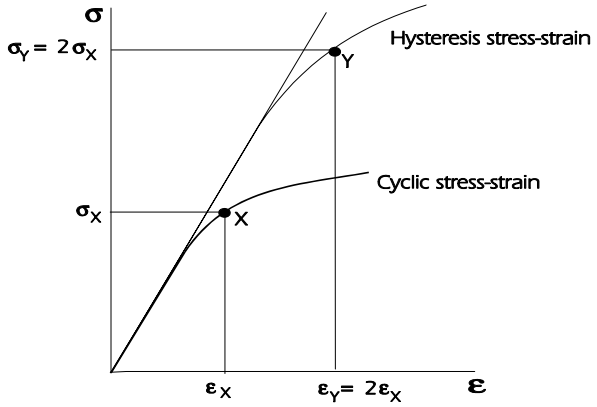


Figure 2.15 Cyclic and hysteresis stress-strain curves

Expressing the cyclic stress-strain curve as (from eqn 2.19)

$$\varepsilon_x = \frac{\sigma_x}{E} + \left(\frac{\sigma_x}{K'}\right)^{\frac{1}{n'}}$$

and so

$$\frac{\varepsilon_y}{2} = \frac{\sigma_y}{2E} + \left(\frac{\sigma_y}{2K'}\right)^{\frac{1}{n'}}$$

or

$$\varepsilon_y = \frac{\sigma_y}{E} + 2\left(\frac{\sigma_y}{2K'}\right)^{\frac{1}{n'}}$$

The hysteresis loop equation is used to calculate stress and strain ranges, so the equation in terms of ranges is

$$\boxed{\Delta\varepsilon = \frac{\Delta\sigma}{E} + 2\left(\frac{\Delta\sigma}{2K'}\right)^{\frac{1}{n'}}} \qquad (2.20)$$

where Δ is used to denote a stress or strain range, i.e. a change in stress or strain.

2.5 Material response to a sequence of strains

Figure 2.16 shows a short sequence of local strains at a notch.

The first strain excursion, from zero strain at O to the strain at A, follows the *cyclic stress strain curve* (Figure 2.17). All subsequent strain excursions follow the *hysteresis loop curve.*

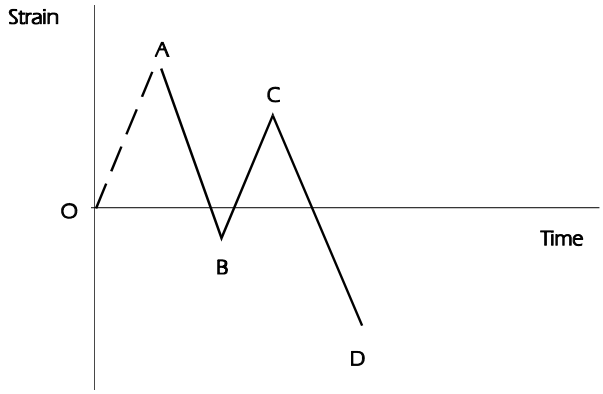


Figure 2.16

The material response to the strain range A - B is obtained by constructing the hysteresis loop curve with its origin at point A. The excursion from B to C is modelled by constructing the hysteresis loop curve with its origin at point B, with true stress and true strain axes as shown *(Figure 2.17)*.

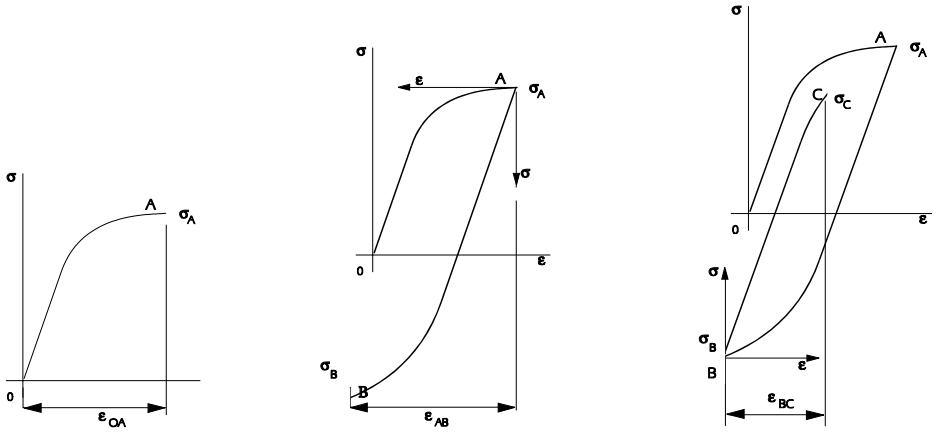


Figure 2.17 Stress response to strains oa, ab and bc

The strain excursion from C to D introduces another aspect of cyclic materials behaviour. If a hysteresis loop stress-strain curve is constructed with its origin at point C, it can be seen that this strain excursion closes a hysteresis loop. This is a fatigue cycle, and this method of modelling a sequence of strains will identify fatigue cycles reliably, however random the signal.

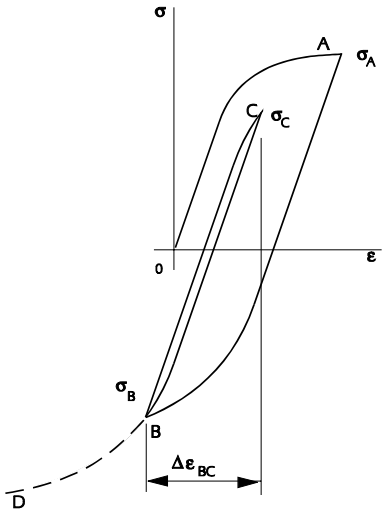


Figure 2.18 Cycle closure

The cycle B-C has a strain range $\Delta\varepsilon_{BC} = \varepsilon_C - \varepsilon_B$, a mean strain $(\varepsilon_C + \varepsilon_B)/2$ and a maximum stress σ_C The strain amplitude for the cycle, $\Delta\varepsilon_{BC}/2$, can be used to calculate the fatigue damage caused by the cycle. This is described in Section 2.7

The closed cycle is in effect an interruption to the larger strain excursion from A to D. After the cycle B-C has been closed, the strain excursion from A to D is modelled by returning to the hysteresis loop curve with its origin at point A and using the strain range from A to D. This behaviour is sometimes referred to as material memory. This is similar to material behaviour in a conventional tensile test, where after a reversal of loading the material stress-strain response continues almost as though the reversal had not occurred *(Figure 2.19)*.

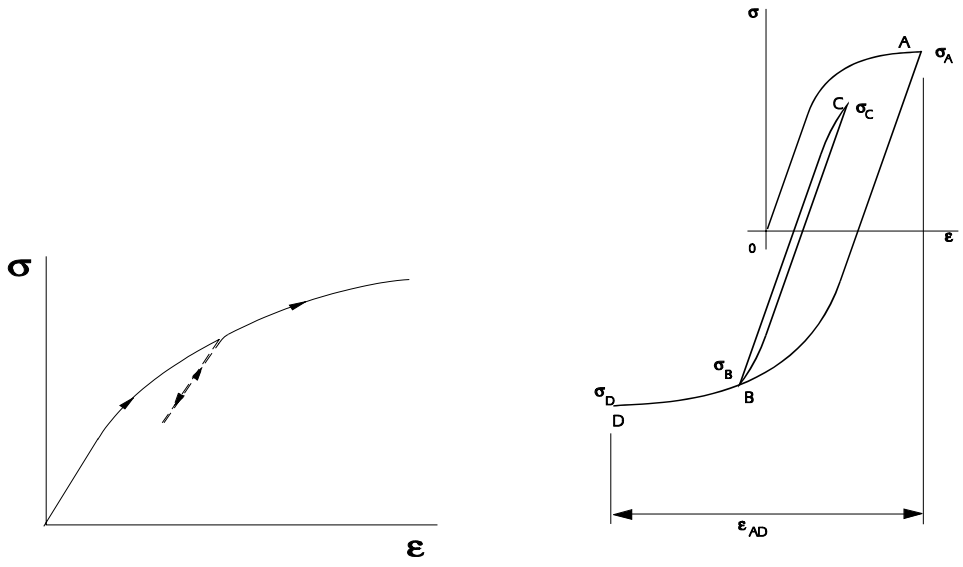


Figure 2.19 'material memory' in a tensile test (left)
and after cycle closure (right)

2.6 Effect of mean stress

It was shown in Chapter 1 that fatigue tests with a tensile mean stress produced shorter lives than tests with the same amplitude at zero mean stress. The different mean stresses can be displayed as stress-strain hysteresis loops *(Figure 2.20)*.

A number of methods of allowing for the effects of mean stress have been proposed. Smith, Watson and Topper (Ref 2.4) suggested that fatigue life was function not of strain amplitude alone but of the product of strain amplitude and the maximum stress in the cycle,

$$\text{i.e. } \frac{\Delta\varepsilon}{2}\,\sigma_{max}$$

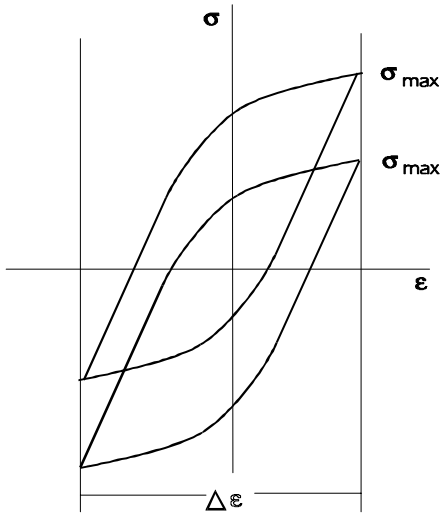
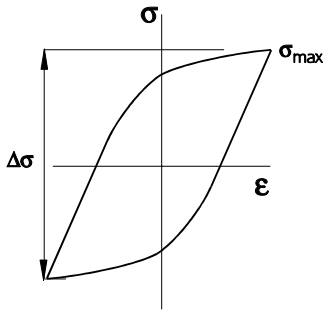


Figure 2.20 Hysteresis loops with different mean stresses

At zero mean stress

$$\sigma_{max} = \frac{\Delta\sigma}{2}$$



and from equation 2.9

$$\frac{\Delta\sigma}{2} = \sigma'_f\,(2N_f)^b$$

so that = $\sigma_{max} = \sigma'_f\,(2N_f)^b$

Taking the strain-life equation (2.16)

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c$$

and multiplying the left hand side by σ_{max} and the right-hand side by $\sigma'_f(2N_f)^b$

gives $$\frac{\Delta\varepsilon}{2}\sigma_{max} = \frac{(\sigma'_f)^2}{E}(2N_f)^{2b} + \sigma'_f\varepsilon'_f(2N_f)^{b+c} \qquad (2.21)$$

This is the Smith-Watson-Topper relationship.

The maximum stress is obtained for each closed hysteresis loop. The procedure is described in Section 2.7 for measured local strains, and in Section 2.10 for measured nominal strains. The Smith-Watson-Topper relationship is addressed futher in Section 2.8.

The Smith, Watson and Topper mean stress correction has been widely used since its publication in 1970, and has been shown to give acceptable results for a wide range of components and measured strain signals.

(In the original paper the equation was presented with the parameter

$$\sqrt{\frac{\sigma_{max}\,\Delta\varepsilon\,E}{2}}$$

on the left hand side. This gives units of stress, and so it is comparable with conventional stress-life curves. In many reproductions of this formula, and in some computer software, the square root and/or elastic modulus are omitted. When using published endurance curves care must be taken to determine which version of the Smith-Watson-Topper parameter is displayed.)

An earlier and more simple correction was proposed by Morrow (Ref 2.5) who corrected the elastic term of the strain-life relationship by subtracting the mean stress σ_m for the cycle.

$$\frac{\Delta\varepsilon}{2} = \underbrace{\frac{(\sigma'_f - \sigma_m)}{E}(2N_f)^b}_{\text{elastic}} + \underbrace{\varepsilon'_f(2N_f)^c}_{\text{plastic}} \qquad (2.22)$$

This equation is based on the observation that mean stress effects are more significant at long lives, where elastic conditions predominate. Morrow's mean stress correction is offered in fatigue analysis software, and gives acceptable results. However, it is theoretically incorrect, as by modifying only the elastic line it implies that the ratio of elastic to plastic strain changes with mean stress, and hence that the shape of the hysteresis loop changes with mean stress. Experimental evidence suggests that the shape of the hysteresis loop for a material is dependent only on strain range. (See pages 42 – 44 for more details of the Smith-Watson-Topper and Morrow mean stress corrections)

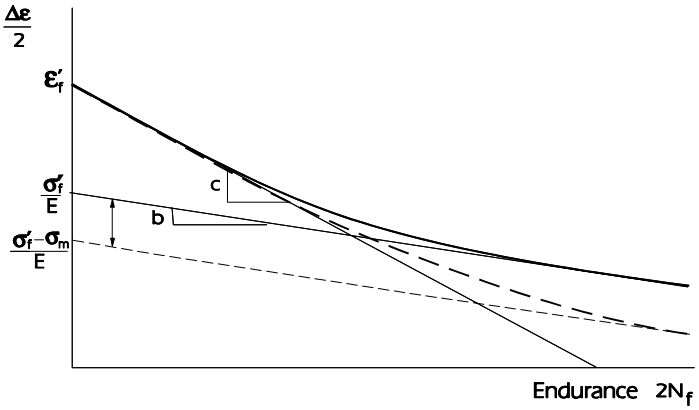


Figure 2.21 Morrow's mean stress correction

2.7 Analysing local strain histories

This section shows how the equations developed in Sections 2.1 - 2.6 can be used to calculate fatigue lives from a time history of local strain.

Figure 2.22 shows a short section of a local strain history - a strain history measured at a stress concentration. The fatigue life will be calculated for a component subject to this local strain history.

The initial strain excursion, from O to A, is modelled using the *cyclic stress-strain* curve. In Section 2.4 it was stated that this curve represents the stress-strain response for a stable material to an initial loading from a state of zero stress-strain.

The strain range O - A is plotted on the strain axis, and the stress at point A is calculated from the equation for the cyclic stress-strain curve

$$\varepsilon = \frac{\sigma}{E} + \left(\frac{\sigma}{K'}\right)^{\frac{1}{n'}} \qquad \text{from (2.19)}$$

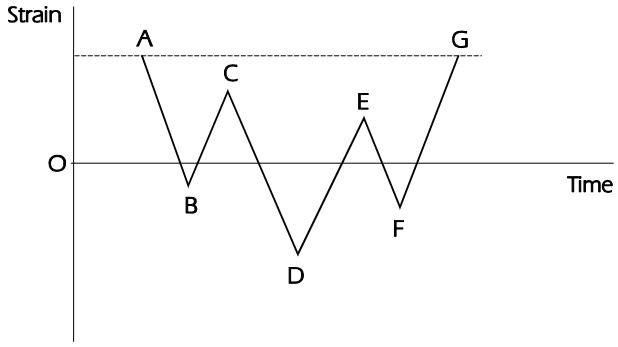


Figure 2.22 Measured time history of local strain

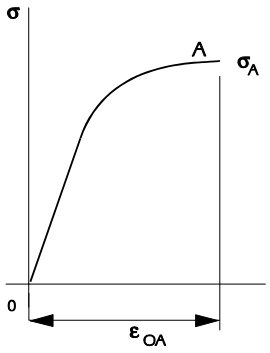


Figure 2.23 Strain excursion from *o* to *a* using the cyclic stress-strain curve

Point A is then taken as an origin, and the strain excursion A to B is plotted as shown in *Figure 2.24*. All strain excursions now follow the hysteresis loop curve, so the stress range from A to B is calculated from the equation for the hysteresis curve.

$$\Delta\varepsilon = \frac{\Delta\sigma}{E} + 2\left(\frac{\Delta\sigma}{2K'}\right)^{\frac{1}{n'}} \qquad \text{from (2.20)}$$

The stress at B is obtained by subtracting the stress range A-B from the stress at A.

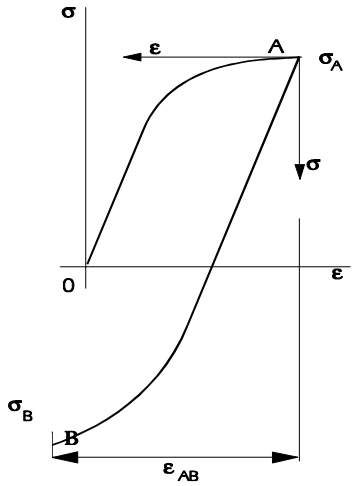


Figure 2.24 Strain excursion from A to B using the hysteresis loop curve

For the strain excursion from B to C, a new origin for the hysteresis loop curve is constructed at point B. The strain range from B to C is plotted on the strain axis, and the stress range B - C is calculated from the equation for the hysteresis loop curve. The stress at C is obtained by adding this stress range to σ_B

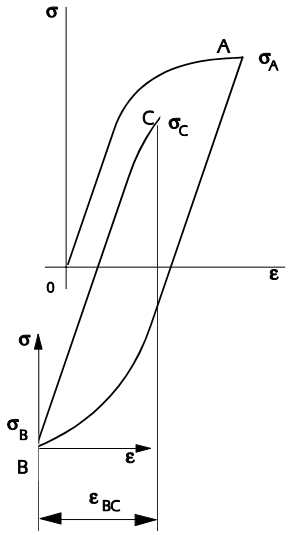


Figure 2.25 Strain excursion from B to C

If the analysis outlined above was repeated for strain C to D, a new hysteresis loop curve would be constructed from point C.

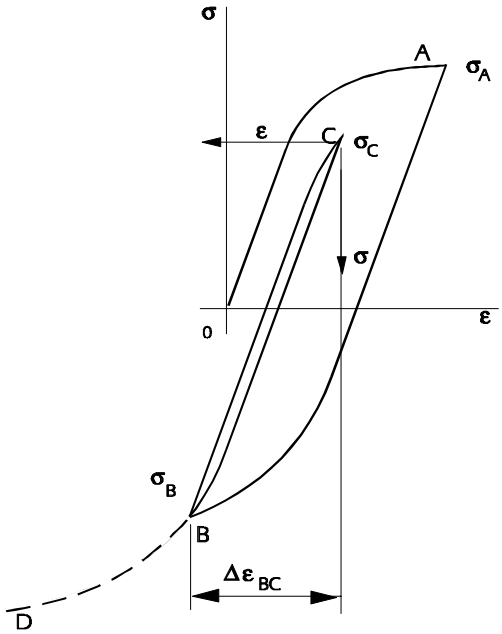


Figure 2.26 the strain excursion from C to D closes a cycle

However, *Figure 2.26* shows that this strain excursion forms a closed stress-strain hysteresis loop, with point B as its lower tip and point C at its upper tip. This is a fatigue cycle, and the fatigue damage it causes can be calculated.

For a strain-life calculation, the strain range $\Delta\varepsilon_{BC}$ from B to C is divided by 2 to obtain the strain amplitude $\frac{\Delta\varepsilon}{2}$, and the endurance $2N_f$ reversals is obtained from the strain-life equation

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c$$

If the Smith-Watson-Topper mean stress correction is being used, the strain amplitude is multiplied by the maximum stress in the cycle, σ_{max}, which is the stress σ_c at point C, to give the stress-strain product

$$\frac{\Delta\varepsilon}{2}\sigma_{max}$$

The endurance $2N_f$ reversals is obtained from the Smith-Watson-Topper equation

$$\frac{\Delta\varepsilon}{2}\sigma_{max} = \frac{(\sigma'_f)^2}{E}(2N_f)^{2b} + \sigma'_f\varepsilon'_f(2N_f)^{b+c} \qquad \text{from (2.21)}$$

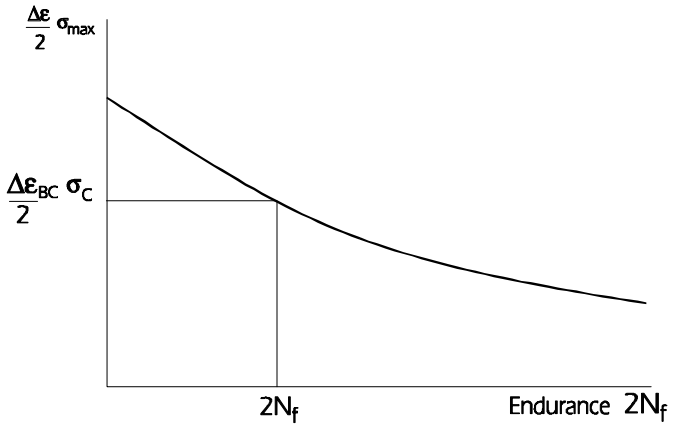


Figure 2.27 Endurance for the cycle B to C

The fatigue damage, using Miner's rule (from Chapter 1), is $\frac{n}{N_f}$ for this cycle, and as $n = 1$ cycle

$$damage = \frac{1}{N_f}$$

The cycle B-C is an interruption to the larger strain range A-D, so point D is calculated using the hysteresis loop curve constructed from point A with the strain excursion from A to D.

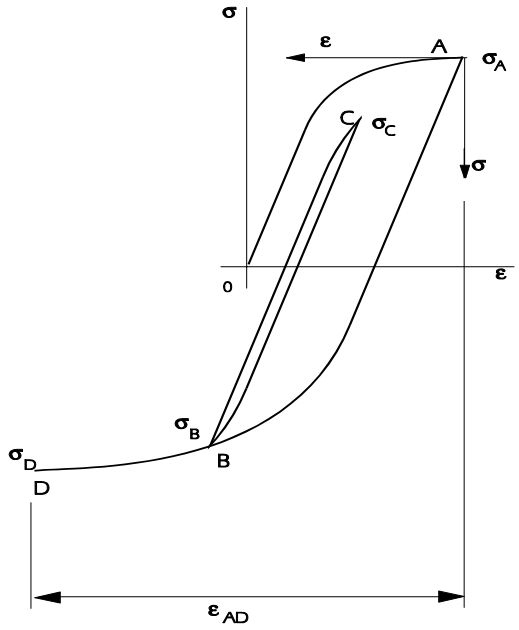


Figure 2.28 Effect of 'material memory'

Strain ranges D to E and E to F are used to calculate the stresses at points E and F.

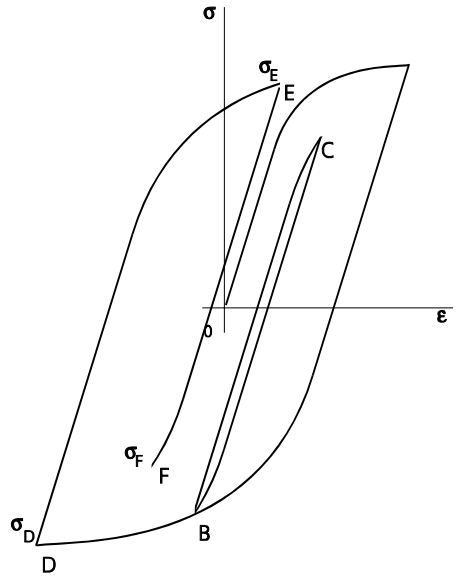


Figure 2.29 Strain excursions from D to E and E to F

The strain excursion F - G closes the cycle formed by E and F *(Figure 2.30)* and the damage is calculated from the strain range $\Delta\varepsilon_{EF}$ and the stress σ_E, .As the strain range F - G has closed the cycle, point G is calculated using a hysteresis loop curve with its origin at point D and strain range D-G. If the calculation has been performed with sufficient accuracy, point G will coincide with point A, and the largest cycle, given by points A, D and G, will close *(Figure 2.30)*.

The strain amplitude is $\Delta\varepsilon_{ADG}\ /2$, and this is multiplied by the stress σ_A to give the Smith-Watson-Topper parameter for this cycle. The damage for the cycle is calculated from the strain-life or Smith-Watson-Topper equations, as before.

Using Miner's rule, the total damage for this short signal is

total damage	=	damage for cycle B to C
	+	damage for cycle E to F
	+	damage for cycle A to D to G

i.e. $$\Sigma\frac{n}{N_f} = \left(\frac{1}{N_f}\right)_{B\text{-}C} + \left(\frac{1}{N_f}\right)_{E\text{-}F} + \left(\frac{1}{N_f}\right)_{A\text{-}D\text{-}G}$$

'Failure', in this case crack initiation, occurs when the fatigue damage $\Sigma\frac{n}{N_f} = 1$, so the fatigue life is the inverse of the damage

i.e. $$fatigue\ life = \frac{1}{\Sigma\frac{n}{N_f}}$$

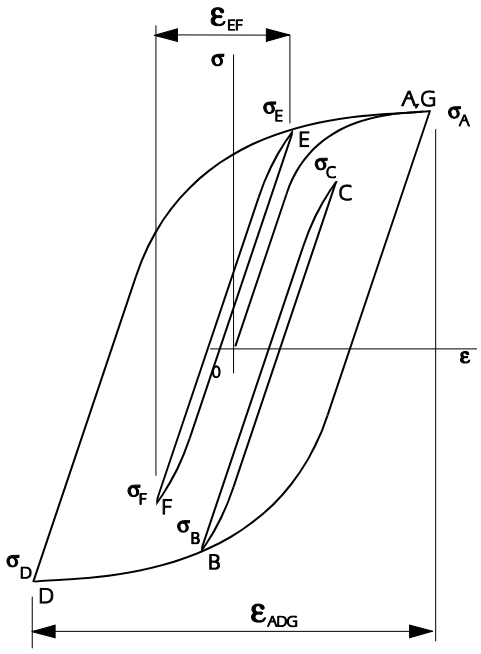


Figure 2.30 strain excursion from F to G

It is clear that a hysteresis loop is closed when a strain range is greater than the strain range immediately preceding it. The procedure of comparing successive strain ranges to determine the cycles in a signal is *fatigue cycle counting*, usually referred to as Rainflow cycle counting. It is described in Chapter 4.

Example 2.1

Figure 2.31 shows a short time history of local strain.

The strain values are :

POINT	STRAIN	POINT	STRAIN
A	0.003	D	-0.0025
B	-0.001	E	0.0014
C	0.0014	F	-0.001

A materials data set for SAE1045 steel will be used :

σ'_f = 948 MPa	b = -0.092	K' = 1258 MPa
E = 202000 MPa	ε'_f = 0.26	c = -0.445
n' = 0.208		

The stress-strain and endurance curves are shown in *Figure 2.32*.

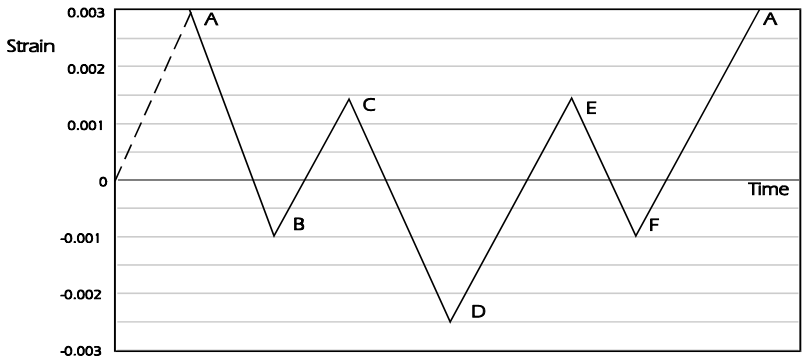


Figure 2.31 Local strain history for example 2.1

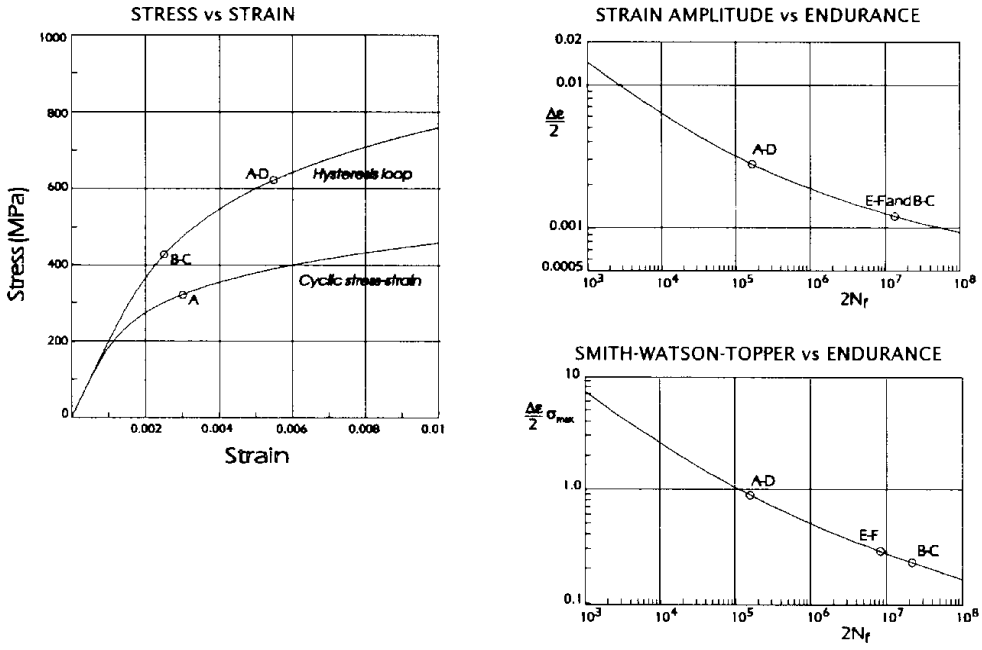


Figure 2.32 Materials data for SAE1045 steel

Referring to *Figure 2.31*.

The strain at point A lies on the cyclic stress-strain curve. A strain of 0.003 is found by iteration to correspond to a stress of 321.1 MPa.

The strain range from A to B follows the hysteresis loop curve, with its origin at A. The strain range is (0.003 - (-0.001)) = 0.004. By iteration, the stress range from A to B is 546.3 MPa. The stress at B is therefore (321.1 - 546.3) = -225.2 MPa.

The strain range from B to C is (-0.001 + 0.0014) = 0.0024. On the hysteresis loop curve, with its origin at point B, this represents a stress range from B to C of 415.1 MPa. The stress at C is (-225.2 + 415.1) = 189.9 MPa.

The strain range from C to D closes a hysteresis loop, because the strain range C-D is greater than the strain range B-C. The cycle B-C has a strain range of 0.0024, and a maximum stress at C of 189.9 MPa. Because of the material memory effect, the stress at D is calculated by using the strain range from A to D. The strain range is (0.003 - (-0.0025)) = 0.0055. On a hysteresis loop curve with origin at A, the stress range from A is 622.2 MPa. The stress at D is then (321.1 - 622.2) = -301.1 MPa.

The strain range from D to E is (-0.0025 - 0.0014) = 0.0039. On a hysteresis loop curve with origin at point D, this represents a stress range from D of 540.1 MPa, so the stress at E is (-301.1 + 540.1) = 239 MPa.

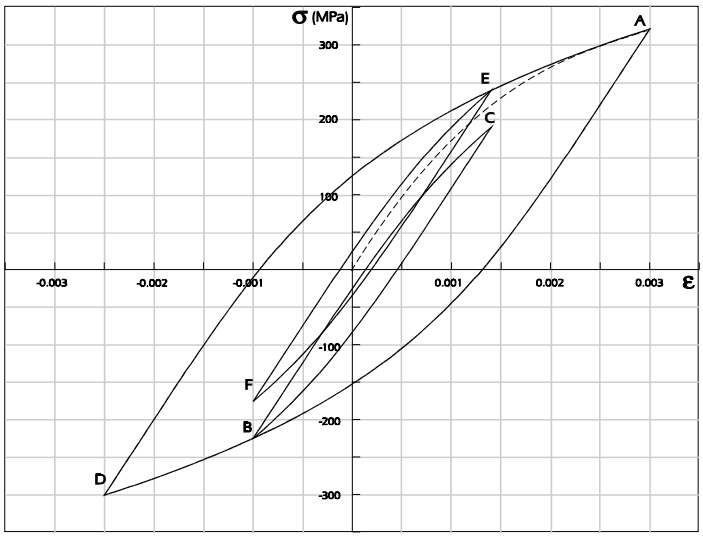


Figure 2.33 Slution to example 2.1

The strain range from E to F is (0.0014 - (-0.001)) = 0.0024. On the hysteresis loop curve with its origin at point E, the stress range is 415.1 MPa, and the stress at F is (239.1 - 415.1) = -176 MPa.

The strain range from F to A closes the cycle E-F. Its strain range is 0.0024 and the maximum stress at E is 239.1 MPa. Using material memory, the stress at A is calculated using a hysteresis loop curve with its origin at D. The strain range from D to A is 0.0055, and the stress at A is 321.1 MPa. This strain range has closed the largest cycle in the signal, that from A-D-A. Its strain range is 0.0055, and the maximum stress at A is 321.1 MPa.

A summary of the three cycles is shown in *Figure 2.33*. The endurance for a component subject to this local strain history could be calculated using a strain-life relationship, or the mean-stress corrected Smith-Watson-Topper relationship. The results, obtained by iteration, are shown in the table.

DAMAGE CALCULATION, STRAIN-LIFE RELATIONSHIP

CYCLE	$\Delta\varepsilon$	$2N_f$	DAMAGE $\frac{1}{N_f}$
B-C	0.0024	$1.437\text{x}10^{7}$	$1.392\text{x}10^{-7}$
E-F	0.0024	$1.437\text{x}10^{7}$	$1.392\text{x}10^{-7}$
A-D	0.0055	$1.751\text{x}10^{5}$	$1.142\text{x}10^{-5}$
			$1.170\text{x}10^{-5}$

The mean life to crack initiation is $\frac{1}{1.170\text{x}10^{-5}}$ = 85500 repeats of the signal.

DAMAGE CALCULATION, SMITH-WATSON-TOPPER RELATIONSHIP

CYCLE	$\Delta\varepsilon$	σ_{max}	$\frac{\Delta\varepsilon}{2}\sigma_{max}$	$2N_f$	DAMAGE $\frac{1}{N_f}$
B-C	0.0024	189.9	0.228	$2.135\text{x}10^{7}$	$9.366\text{x}10^{-8}$
E-F	0.0024	239.1	0.287	$8.012\text{x}10^{6}$	$2.496\text{x}10^{-7}$
A-D	0.0055	321.1	0.883	$1.614\text{x}10^{5}$	$1.239\text{x}10^{-5}$
				Total damage =	$1.273\text{x}10^{-5}$

The mean life to crack initiation is $\frac{1}{1.273 \times 10^{-5}}$ = 78500 repeats of the signal.

Comments on Example 2.1

The following points may be noted from the example.

The two strain points C and E have the same local strain, but different maximum stresses. This is not obvious from the strain-time history, but only becomes clear when the hysteresis loops have been constructed *(Figure 2.33)*.

As the cycles B-C and E-F have the same strain range, the strain-life analysis calculates the same damage for both cycles. The analysis with the mean stress correction calculates a higher damage for cycle E-F.

The cycle formed by points B-C has a positive mean strain, but a negative mean stress. This is shown by the hysteresis loop in *Figure 2.33*.

The short strain-time history started and ended at the numerically largest strain. If the numerically largest strain is not at the start and end of the history (and it rarely is) then the analysis still starts and ends at the numerically largest value. In computer software, this may be achieved by looping back to the beginning of the signal, and continuing the analysis until the start point is reached again. (If a compressive strain is the numerically largest value, the analysis is started by constructing the cyclic stress-strain curve in the compressive direction).

The justification for this procedure is that the measured signal represents a 'day in the life' of the component, which will be repeated until a crack initiates. This is implied by expressing the life as a number of repeats of the signal. As the start point is probably quite arbitrary - it depends on when recording was started - then any start point is as valid as any other, and starting and ending at the numerically largest value ensures that the outside hysteresis loop closes.

A further justification is that the mean stresses of the smaller cycles will be different before and after the numerically largest strain, and so the calculated damage will be different. This is shown in Figure 2.34. Placing the smaller cycles after the numerically largest value is more representative of the many repeats of the signal which will occur after the first one.

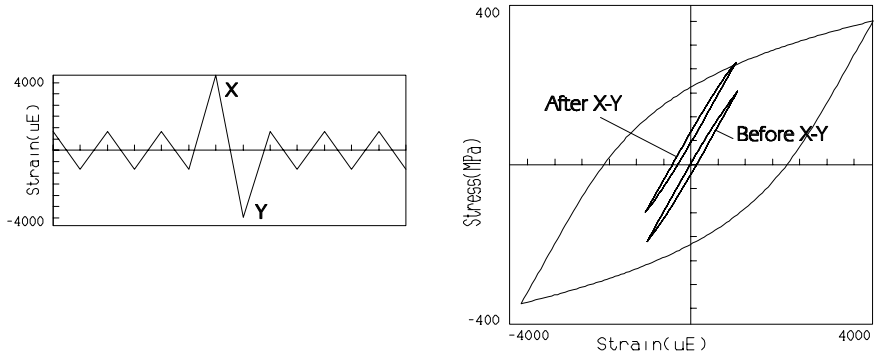


Figure 2.34 Risidual stresses after a large strain cycle

2.8 The Smith-Watson-Topper relationship

The Smith-Watson-Topper mean stress correction is given by

$$\frac{\Delta\varepsilon}{2}\sigma_{max} = \frac{(\sigma'_f)^2}{E}(2N_f)^{2b} + \sigma'_f\varepsilon'_f(2N_f)^{b+c} \qquad \text{from (2.21)}$$

This equation can be plotted to show the relationship between strain amplitude and mean stress for any chosen endurance. *Figure 2.35* was calculated for SAE 1045 steel at an endurance $N_f = 10^7$ cycles. The Morrow mean stress correction is shown for comparison.

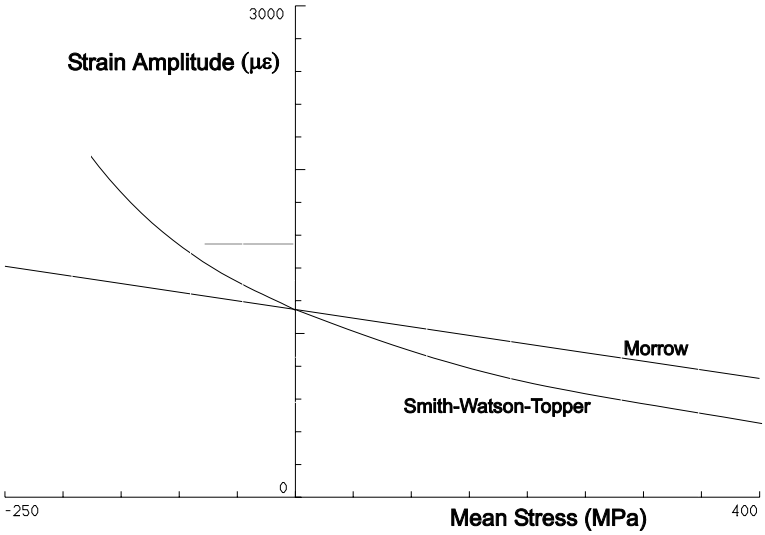


Figure 2.35 Mean stress correction plotted as true strain amplitude vs true mean stress for SAE1045 steel at $N_f = 10^7$ cycles

The Smith-Watson-Topper equation (2.21) becomes undefined if a cycle has a compressive *maximum* stress and it is usual to assume that these compressive cycles contribute no fatigue damage. This seems at first sight a quite sweeping assumption. However, for very many practical components the true strain at a critical notch will be within the range $\pm 5000\mu\varepsilon$ ($\varepsilon = \pm 0.005$). This limits the strain amplitude of compressive cycles.

Figure 2.36 shows the largest wholly compressive cycle that can be produced when the largest strain in the signal is $-5000\mu\varepsilon$. The cycle has a strain range of $2100\mu\varepsilon$ i.e. a strain amplitude of $1050\mu\varepsilon$. For this cycle the calculated endurance is greater than 10^7 cycles, and so the calculated damage for the cycle would be small whichever mean stress correction was being used.

The Smith-Watson-Topper relationship generally gives good agreement with test data for many engineering signals. The Morrow correction will give shorter calculated fatigue lives for signals where the local stresses at the notch have compressive mean stresses. For high compressive mean stresses, approaching the conventional yield stress, both methods may give non-conservative estimates of fatigue life.

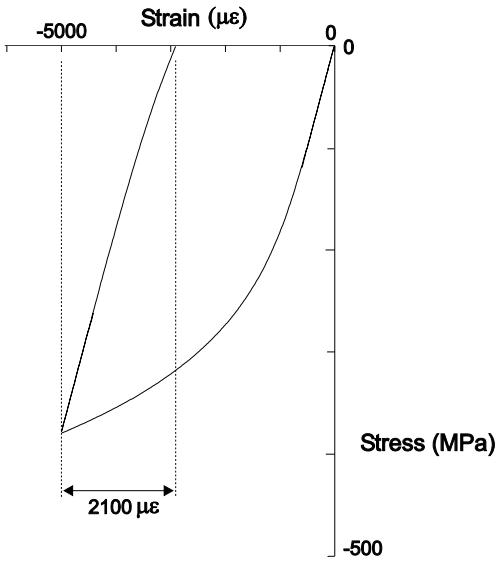


Figure 2.36 strain range of a wholly compressive cycle (SAE1045 cyclic stress-strain curves)

2.9 Application of stress concentrations

To apply local strain methods to real engineering components the local stresses and strains at the critical location must be known. For some components the local strains can be measured directly using strain gauges. In many other instances it may be possible only to measure strains some distance away from a stress concentration, because of physical constraints. Strains remote from and unaffected by the stress concentration are nominal strains.

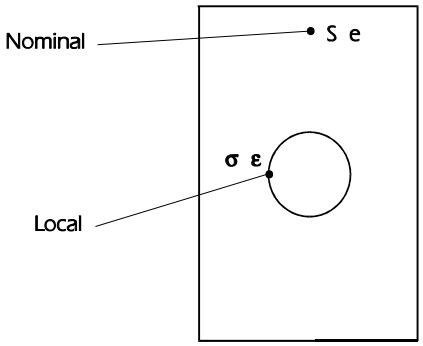


Figure 2.37 Nominal and local stress and strain

2.9.1 Neuber's Rule

Neuber (Ref. 2.6) proposed that the local stress and strain at the notch could be determined using the elastic stress concentration factor K_t.

If the nominal stress and strain are S and e,

and the local stress and strain are σ and ε,

the Neuber relationship is

$$\Delta\sigma\Delta\varepsilon = K_t \Delta S . K_t \Delta e \qquad (2.23)$$

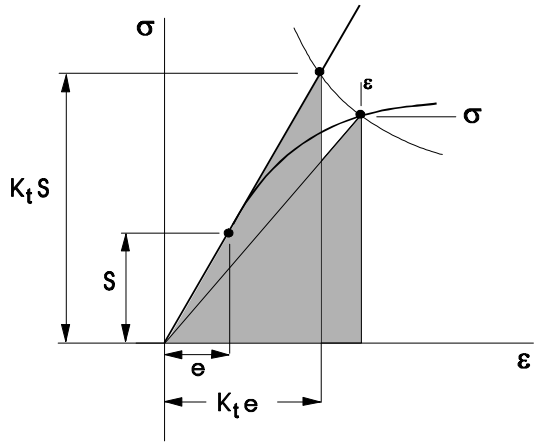


Figure 2.38 Neuber's rule

The left-hand side of the equation is the stress-strain product for the true local elastic-plastic stress and strain. The right hand side is the stress-strain product which would be calculated if the material was infinitely elastic.

As the stress-strain product is proportional to the strain energy at the notch, Neuber's rule essentially equates the strain energy for an elastic-plastic and an infinitely elastic material.

Neuber's Rule is applied as follows.

The first estimate of the local stress and strain σ, ε, can be obtained by multiplying the nominal stress ΔS and nominal strain Δe by K_t. In *Figure 2.39*, this produces the point X on the extension of the elastic portion of the stress-strain curve. This point will have a stress-strain product $\Delta\sigma_x \Delta\varepsilon_x$. Neuber proposed that the actual values of true stress and true strain, on the cyclic stress-strain curve, will have the same stress-strain product, i.e. $\Delta\sigma\Delta\varepsilon = \Delta\sigma_x\Delta\varepsilon_x$.

The equation $\Delta\sigma\Delta\varepsilon = constant$ is a hyperbola linking the point X on the elastic line and the true stress-strain point on the cyclic stress-strain curve. This is sometimes referred to as a Neuber hyperbola, and is defined by the equation

$$\sigma\varepsilon = constant, \text{ or } \Delta\sigma\Delta\varepsilon = constant$$

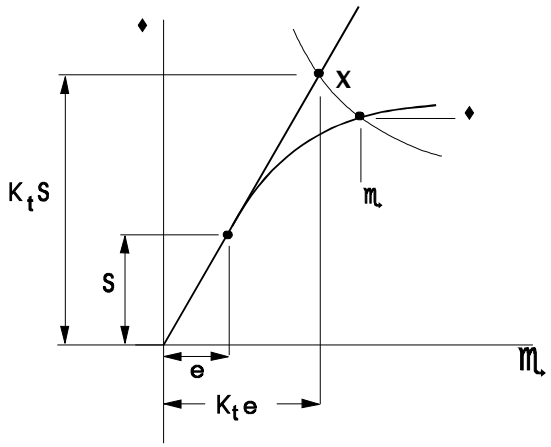


Figure 2.39 Calculation of local stress-strain using Neuber's rule

If the nominal stress/strain is elastic, or if stresses or strains have been calculated using an elastic analysis, for example in an elastic finite element model, Neuber's rule can be simplified to

$$\Delta\sigma\Delta\varepsilon = K_t^2 \frac{S^2}{E} \qquad (2.24)$$

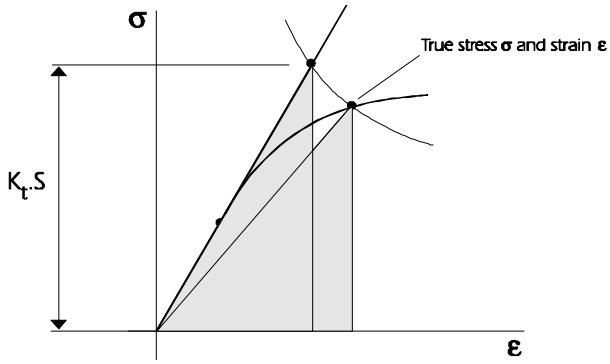
Use of this simplified equation is *not recommended* for analysing measured signals, as practical experience shows that in real signals even nominal strains exhibit some plasticity at the higher strains present in the signal.

Neuber's rule was formulated for shear-strained bodies, and was validated for one stress concentration. However, it is now used for locations which experience uniaxial direct stresses, and is used for any geometric notch. Experimental evidence shows that it overestimates the strains at stress concentrations, and this overestimation increases with high strain levels (Ref. 2.7).

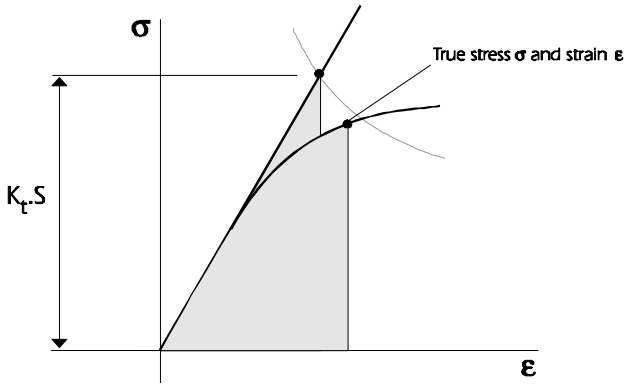
2.9.2 Glinka's Method

As shown in *Figure 2.38*, Neuber's rule makes approximately equal areas under the 'infinitely elastic' and cyclic stress-strain curve by selecting local stresses and strains which make the two triangles equal. Considering Neuber's stress-strain product as analogous to strain energy, Glinka proposed that it would be more consistent to integrate the cyclic stress-strain curve to obtain a better estimate of the area. This produces smaller values of true stress and true strain.

Experimental evidence indicates that Glinka's method underestimates local stresses and strains. A comparison between true stress/strain calculated from Neuber's rule and Glinka's method is shown in *Figure 2.40* (steel SAE1045, $K_t = 1$).



(a) Neuber's Rule



(b) Glinka's Method

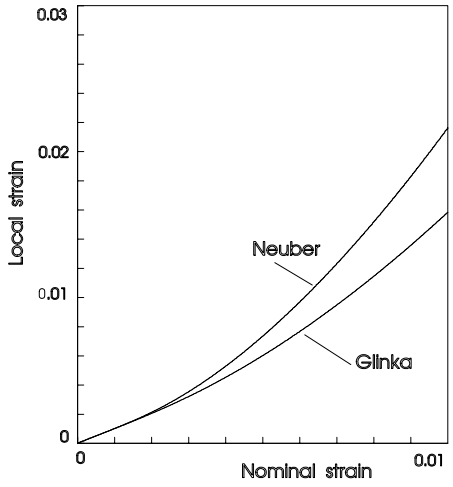


Figure 2.40 Comparison of Neuber's rule (a) and Glinka's method (b)

2.9.3 The Seeger-Heuler Method

The Neuber and Glinka equations apply when yielding is limited to a small volume of material at the stress concentration. Where the nominal stresses and strains are sufficiently high to cause general yielding across a component the Neuber and Glinka methods are not valid. Seeger and Heuler (Ref. 2.8) proposed that the elastic stress concentration factor K_t could be replaced by

$$K_p = K_t \frac{S_L}{\sigma_y}$$

where S_L is the nominal stress at which general yielding occurs, and σ_y is the conventional yield stress for the material. σ_y is therefore the local stress at which local yielding first occurs at the stress concentration. From the definition of the elastic stress concentration factor K_t, the nominal stress at which local yielding first occurs at the notch, S_y, is

$$S_y = \frac{\sigma_y}{K_t} \quad \text{and so} \quad K_p = \frac{S_L}{S_y}$$

Instances were the nominal stresses in a notched component approach the yield stress are unusual, and the Seeger-Heuler equations are required only in these special cases.

2.9.4 True Stress and True Strain Concentration Factors

The true stresses and strains calculated from Neuber's Rule (or any other rule) can be used to calculate true stress and strain concentration factors, which include the effects of plasticity.

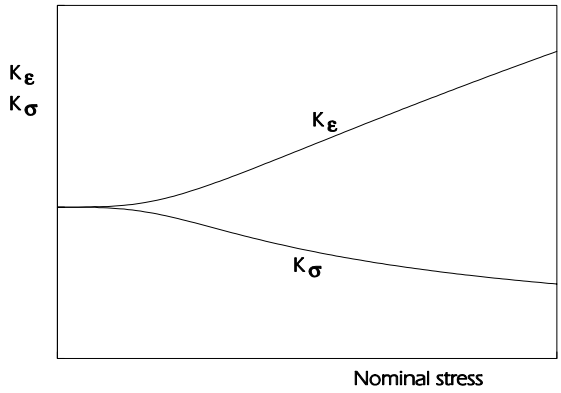


Figure 2.41 True stress and true strain concentration factors

The true stress concentration $K_\sigma = \frac{\sigma}{S}$

and the true strain concentration $K_\varepsilon = \frac{\varepsilon}{e}$

K_σ and K_ε can be plotted against elastic nominal stress (*Figure 2.41*). K_σ and K_ε both equal the elastic stress concentration factor K_t for elastic local stresses. Yielding at the notch causes the true stress concentration K_σ to become progressively smaller, and the local strain concentration K_ε to become progressively larger. The K_σ K_ε plots for two steels are shown in *Figure 2.42*.

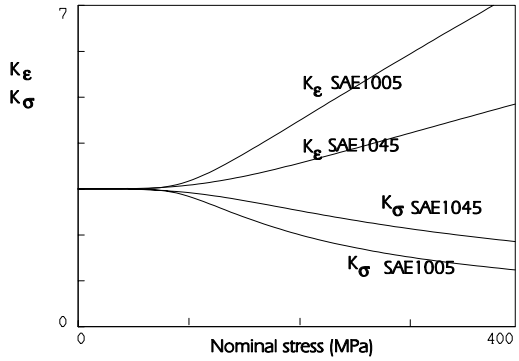


Figure 2.42 Comparison of true stress and true strain concentration factors for two steels

For the same nominal stress, the true stress and strain concentrations for the two materials are quite different. Identical components made from the two materials would produce the same nominal stresses and strains (providing they were below the yield stress) but quite different local stresses and strains. This is illustrated in *Figure 2.43*, where the Neuber hyperbola shows the local strains in the two materials for the same nominal strain.

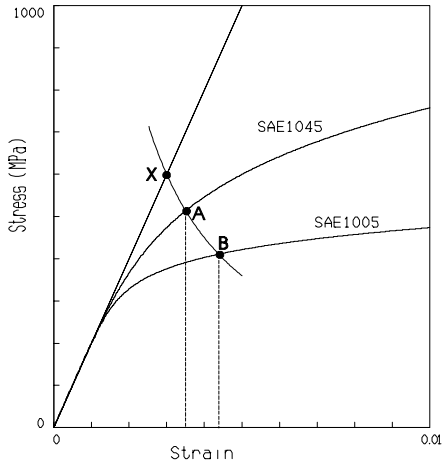


Figure 2.43 Local strains calculated from the same nominal strain for two steels using neuber's rule

This means that *measured local strains cannot be used to evaluate alternative materials*, because the same local strains would not be produced at a notch in different materials. Instead, the Neuber or Glinka rules can be used to convert the measured local strains into 'infinitely elastic' strains (point X in *Figure 2.43*), and these strains used to calculate local strains in another material.

Figure 2.43 also suggests that the fatigue strength of materials cannot be compared simply by comparing the strain-life or Smith-Watson-Topper endurance curves. These curves relate fatigue life to local strain. However, components made from two different materials may develop different local strains at a notch in response to the same loading, and so a comparison must consider both strain-life and cyclic stress-strain properties.

The true stress concentration factor K_σ in *Figure 2.41* can be used to calculate the stress-life and strain-life curves in terms of the nominal strain and stress amplitude remote from the notch. *Figure 2.44* shows 'nominal stress' and 'nominal strain' endurance curves for $K_t = 1$ and $K_t = 3$.

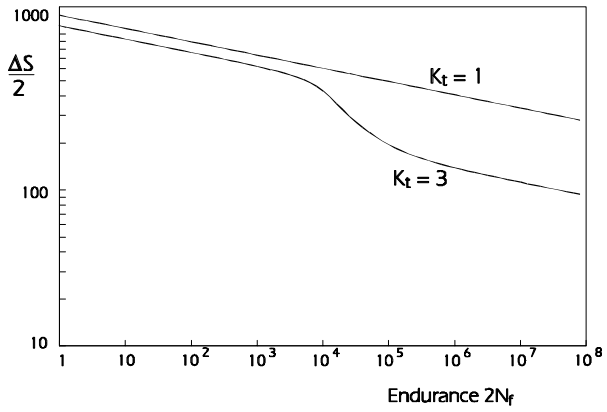


Figure 2.44 (a) Calculated nominal stress-life curves for smooth and notched specimens ($K_t = 1$ and $K_t = 3$)

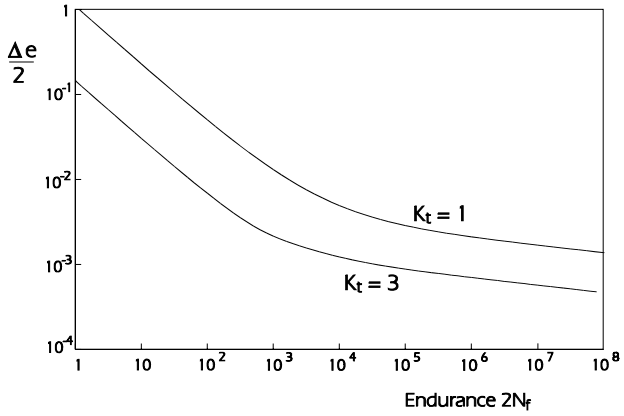


Figure 2.44 (b) Calculated nominal strain-life curves for smooth and notched specimens (K_t = 1 and K_t = 3)

2.10 Analysing nominal strain histories

Section 2.7 showed how fatigue lives can be calculated from a local strain history measured at a notch. A nominal strain history measured at a location unaffected by the stress concentration can be analysed in the same way, except that the effect of the stress concentration factor must be included to convert nominal strains into local strains at the notch, allowing for the effects of local plasticity. As Neuber's rule is the most widely used method, it will be used as an example, although the procedure is very similar if the Glinka method is used.

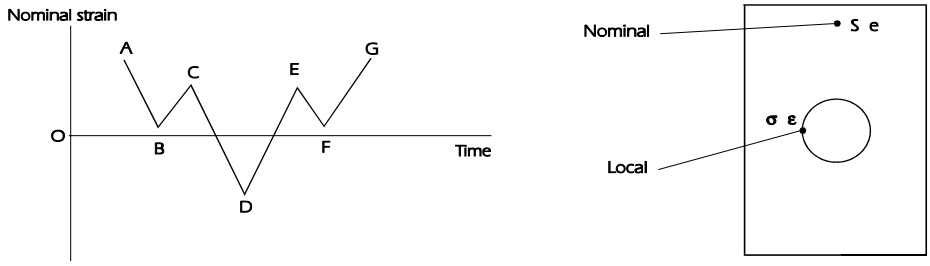


Figure 2.45 Nominal strain history

In *Figure 2.45*; For the strain excursion O to A, the nominal strain e_{OA} is used with the cyclic stress-strain curve to calculate the nominal stress S *(Figure 2.46)*. Multiplying both stress and strain by the stress concentration factor K_t gives a point on the extension of the elastic part of the cyclic stress-strain curve. A Neuber hyperbola through this point intersects the cyclic stress-strain curve to give the true strain and true stress at point A. This is the calculated value of the local strain range ε_{OA} which corresponds to the 'measured' value in Section 2.7.

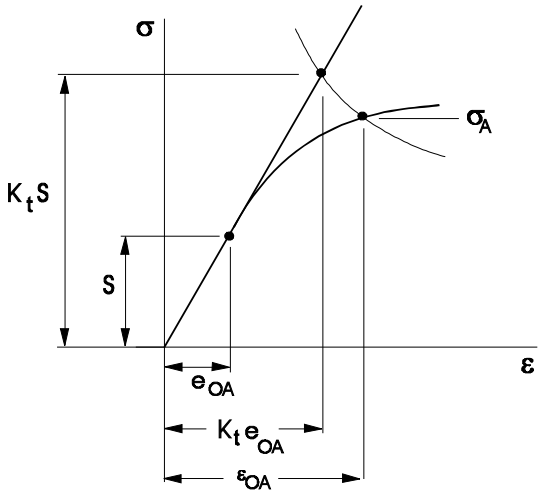


Figure 2.46 True stress and strain at point A using Neuber's rule

For the strain excursion from A to B, the hysteresis loop curve is used, with its origin positioned at the *local stress and strain values* calculated for point A. The nominal stress range ΔS_{AB} and nominal strain range Δe_{AB} are both multiplied by K_t to give the point on the extension of the elastic part of the hysteresis loop curve, and a Neuber hyperbola constructed to intersect the hysteresis loop curve (*Figure 2.47*).

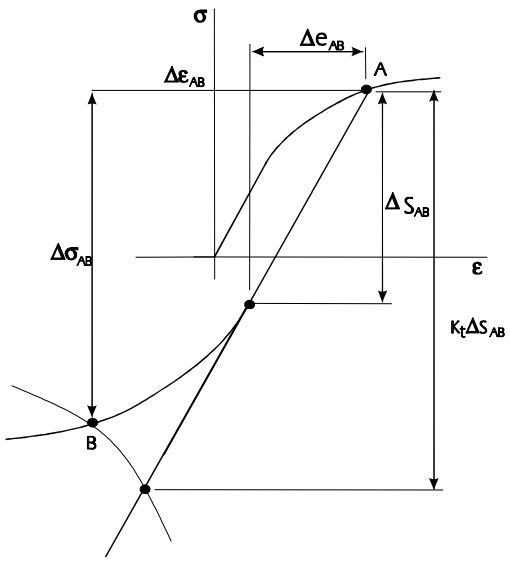


Figure 2.47 Neuber's rule used to calculate the true stress and true strain at point B for the strain excursion A to B

It was mentioned in Section 2.4 that some nominal strains in a real signal may exhibit a small amount of plasticity, on either the cyclic stress-strain or hysteresis loop curves. Neuber's rule was formulated for elastic nominal strains. However, the method may be extended by using the equation for the stress-strain curve to calculate nominal stress from nominal strain. Both elastic and elastic-plastic nominal strains will then be covered by the same calculation. *Figure 2.48* shows this method in graphical form (only stresses are shown for simplicity).

The product of nominal stress and strain will have the same stress-strain product as equivalent nominal values on the elastic extension of the stress-strain curve (ΔS^*). These new values of stress and strain can then be multiplied by the stress concentration, and Neuber's rule applied as before.

This method seems to be justified because equations (2.19 - 2.20) show that there is really no such thing as purely elastic behaviour (only very small plastic strains).

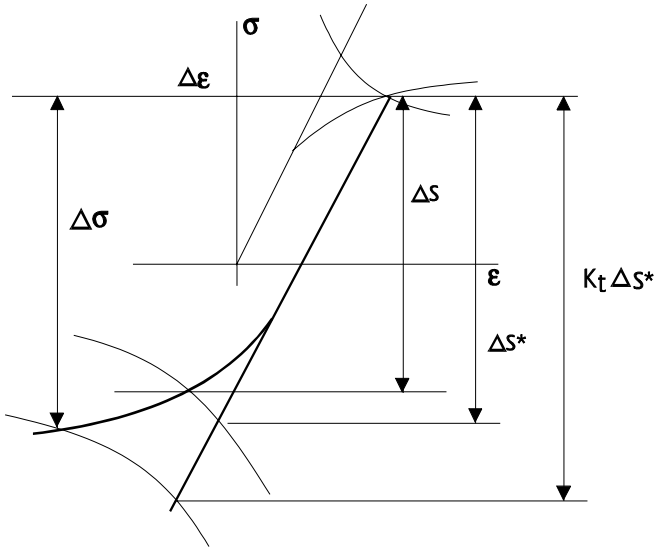


Figure 2.48 Application of neuber's rule when the nominal strain exceeds the elastic limit

2.11 Local strain analysis from a cycle histogram

If a cycle histogram has been generated from fatigue analysis software, or obtained from a cycle-counting signal analyser, this data may be post-processed to calculate a fatigue life. Some approximations are necessary, because the essential information on the sequence of events in the original signal has been lost.

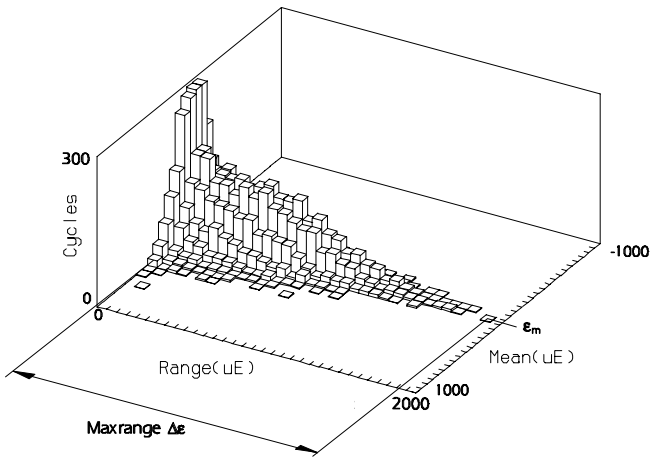


Figure 2.49 Cycle histogram

In *Figure 2.49* the largest cycle in the histogram will form the largest stress-strain hysteresis loop. From the strain range and mean strain for this cycle (or cycles if there are more than one in the histogram bin) the maximum and minimum strains can be calculated. By definition,

$$\varepsilon_{max} = \frac{\Delta\varepsilon}{2} + \varepsilon_m \quad \text{and} \quad \varepsilon_{min} = \frac{\Delta\varepsilon}{2} - \varepsilon_m$$

The largest absolute value of these two strains will lie on the cyclic stress-strain curve. This is shown in *Figure 2.50*, for the case where the most positive strain is the largest absolute strain. In this case the minimum strain will be obtained by constructing the hysteresis loop curve from the loop tip.

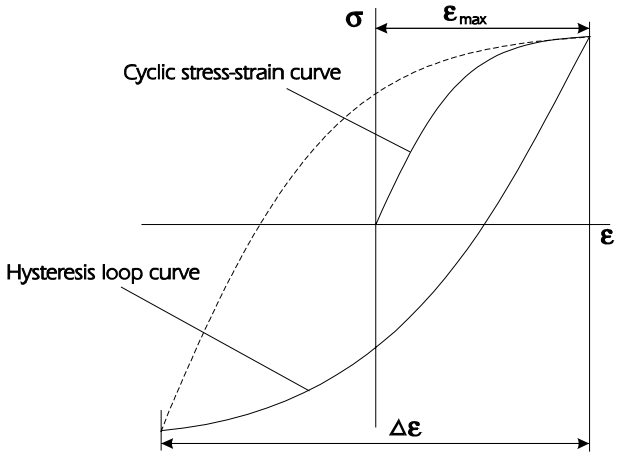


Figure 2.50 Fatigue analysis of the largest strain range

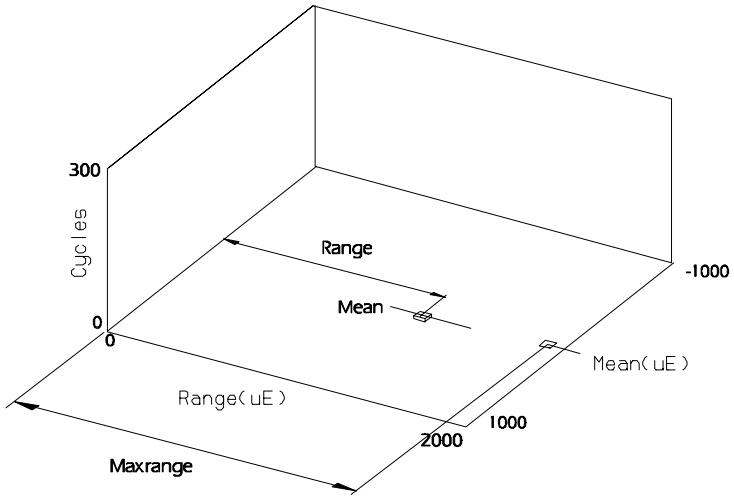


Figure 2.51 Strain range and mean strain for a cycle bin

The fatigue damage for this strain range can be calculated, using the stress at the loop tip if a mean stress correction is required.

For the remaining cycles in the histogram, the maximum and minimum strains for each 'bin' can be calculated. As these cycles will not lie on the cyclic stress-strain curve, the associated stresses cannot be calculated. However, these cycles must lie within the outside hysteresis loop. The most severe assumption would be that the loop tips for these cycles touched the outside loop, as this gives the maximum possible loop tip stress. This stress can be calculated by constructing a hysteresis loop curve from the minimum strain in the outside loop to the maximum strain in the present cycle. This strain range is ε^*_{max} in *Figure 2.52*. The fatigue damage for this cycle can be calculated, and the process repeated for each bin in the histogram. The total calculated damage will be conservative because the highest possible stresses have been used for each strain cycle.

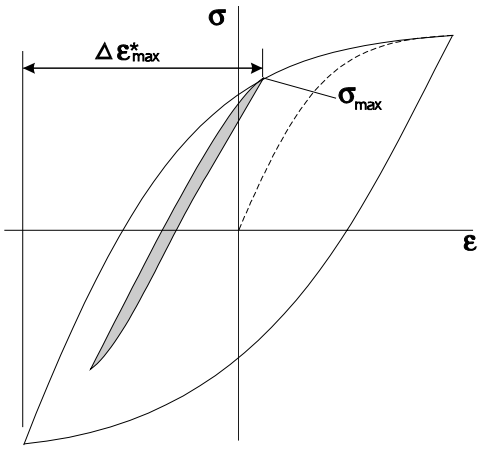


Figure 2.52 'worst case' assumption for σ_{max}

To produce a non-conservative damage estimate, the calculation can be repeated assuming each cycle has its lower tip on the outside loop. This will produce the lowest possible estimate of the maximum stress associated with each cycle. The calculation is shown in *Figure 2.53*. The minimum and maximum strains for the cycle are obviously unchanged, and the minimum stress is calculated by taking a strain range from the tip of the outside loop to the minimum strain in the cycle. This strain range is ε^*_{min} in the diagram 2.53. The maximum stress for this cycle is calculated using the strain range for the cycle and the hysteresis loop curve.

This method has been compared with cycle-by-cycle analysis of real signals, and generally introduces an additional uncertainty factor of approximately ±2 to the calculated lives. It shows the effect on fatigue life of the sequence in the signal. The conservative and non-conservative life estimates show the shortest and longest lives which could be obtained if the sequence of events was changed. It is now being used as an additional calculation from measured signals, and is referred to as *damage bounding*.

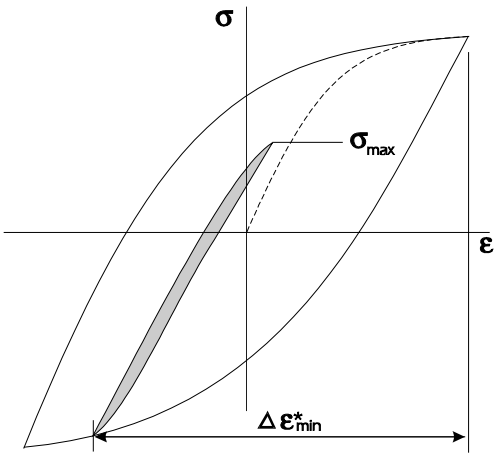


Figure 2.53 'best case' assumption for σ_{max}

2.12 Analysis of cast iron

Cast irons are stiffer in compression than in tension. Stress-strain hystersis loops are non-symetrical. The effect of most pronounced for grey iron. This section describes the analysis of grey cast iron, but the methods are equally applicable to compacted (CG) and nodular (SG) irons. The method is taken from Downing (Ref 2.9).

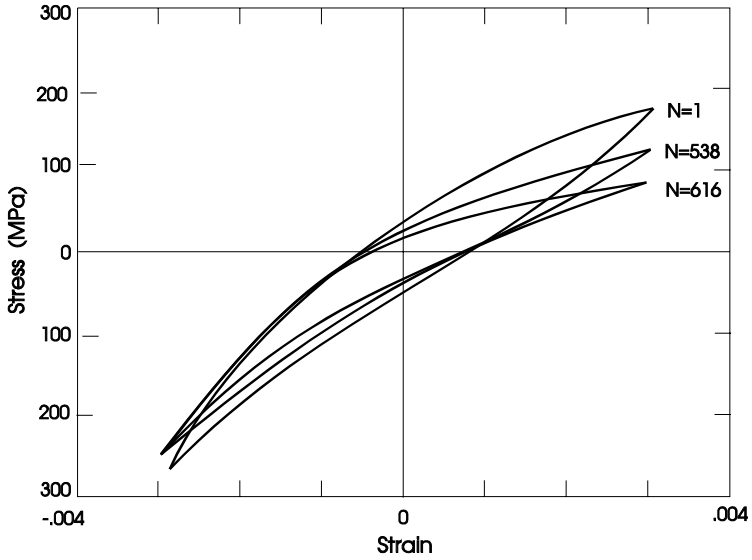


Figure 2.54 Hysteresis loops for a grey iron (from Ref. 2.9)

The stress-strain response of grey iron is considered to consist of the following components

- a bulk steel matrix modified by the presence of graphite
- the effects of the internal graphite
- the effects of surface graphite
- the effects of crack closure under compressive stresses

In addition, it is proposed that fatigue damage accumulates non-linearly.

Each of these effects is described below.

2.12.1 Bulk effects

The relationship between bulk stress σ_B and strain is similar to that of steel, except that the presence of graphite means that there is no initial 'elastic' response.

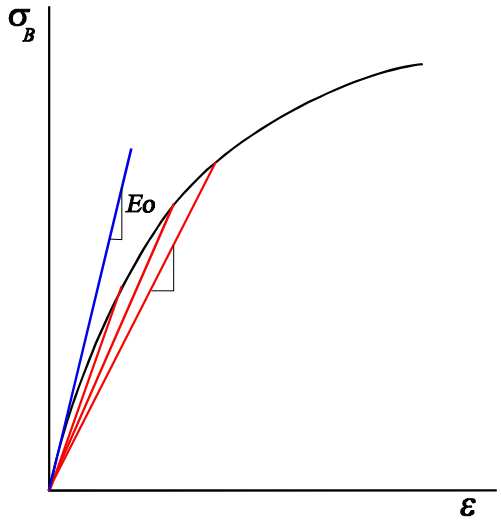


Figure 2.55. Initial stress-strain curve

This initial non-linear response can be characterised by a tangent modulus E_0 for the slope of the stress strain curve at zero stress (shown blue in the schematic diagram *Figure 2.55*), and by a secant modulus. The secant modulus is the slope of a line joining the origin of the stress-strain axis and a point on the stress-strain curve. Examples are shown (red) in *Figure 2.55*.

A plot of the secant modulus against stress σ_B, approximates to a straight line of slope m. An example is shown in *Figure 2.56*.

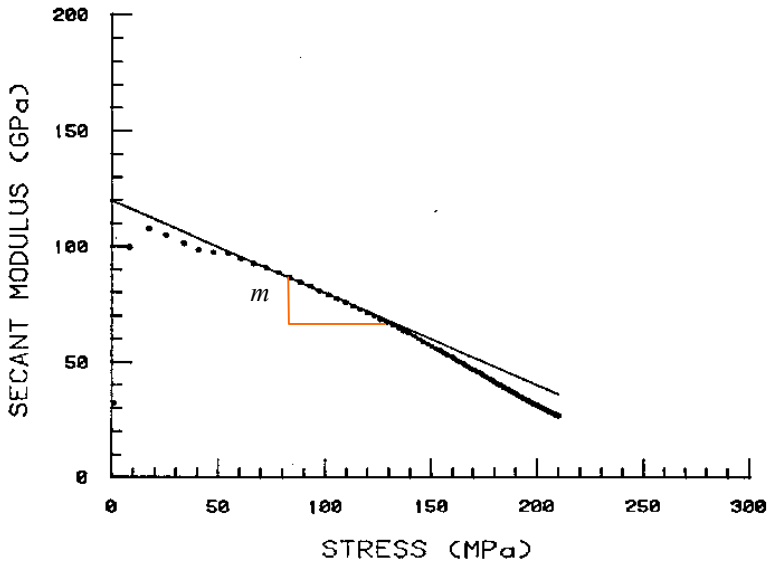


Figure 2.56. Plot of secant modulus versus stress

The effective 'elastic' modulus is given by $E = E_0 + m\sigma_B$

and the stress-strain relationship over this initial 'elastic' part of the curve is then

$$\varepsilon_e = \frac{\sigma_B}{E_0 + m\sigma_B}$$

The plastic component of strain is characterised by the standard power law, so that the cyclic stress strain curve is

$$\varepsilon = \frac{\sigma_B}{E_0 + m\sigma_B} + \left(\frac{\sigma_B}{K'}\right)^{1/n'}$$

and, using Massing's hypothesis, the hysteresis stress-strain curve for bulk stress is given by

$$\Delta\varepsilon = \frac{\Delta\sigma_B}{E_0 + 0.5m\Delta\sigma_B} + 2\left(\frac{\sigma_B}{2K'}\right)^{1/n'}$$

The slope m of the plot of secant modulus versus stress is a required material parameter, and can be obtained by standard regression analysis. It may be necessary to define two secant modulus slopes – one for tensile stresses and one for compressive stresses.

Because steels have a linear elastic response, the slope m is zero, and E_0 is equal to the elastic modulus E. These equations therefore become the standard equations used to model the stress-strain response of steels.

2.12.2 Internal graphite effects

In compression the graphite is essentially incompressible, and additional compressive stresses are transferred to the steel matrix. This additional stress is termed σ_G, and is the difference between the observed stress-strain response and the bulk response σ_B. The value of σ_G is zero if the total stress is tensile, and non-zero if the total stress is compressive. The effect of σ_G is therefore to produce a kink in the stress-strain response. This is shown in *Figure 2.60*.

2.12.3 Surface graphite effects

Surface graphite produces crack-like defects on the surface of the material. This changes the stiffness or compliance of the material at higher tensile stresses. The effect is apparent when the material is unloaded from a tensile stress, in that the unloading modulus $E_u = \sigma/\varepsilon$ reduces as the peak tensile stress is increased. This is shown schematically in *Figure 2.57*, with an example in *Figure 2.58*.

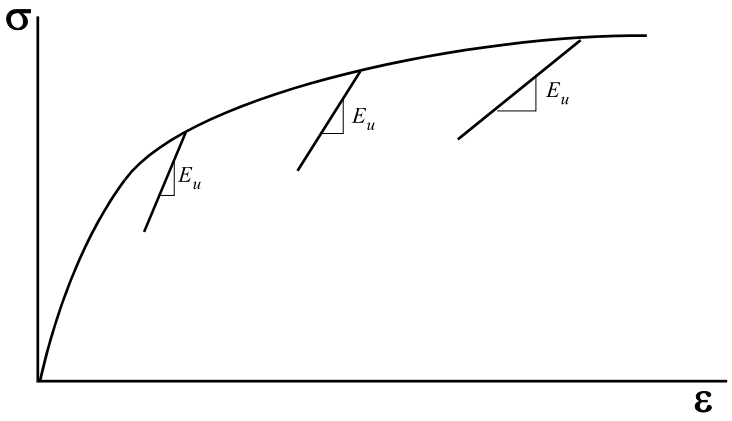


Figure 2.57. Schematic diagram of the unloading modulus.

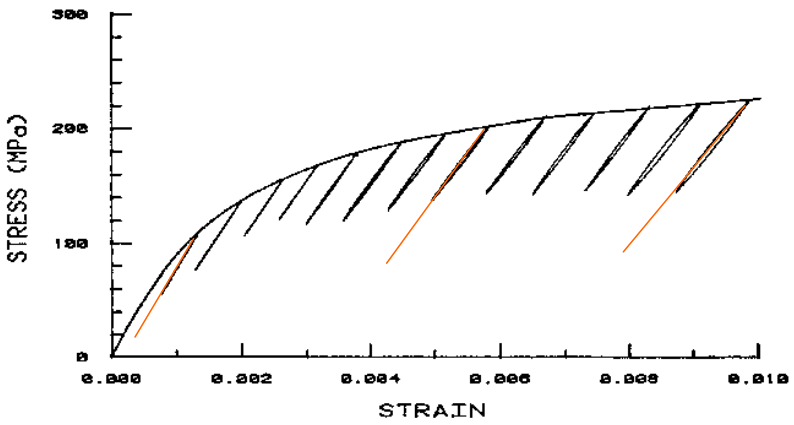


Figure 2.58. Example of the change in unloading modulus with peak tensile stress.

This effect has been attributed to the reduction in bulk cross-sectional area caused by the matrix tearing away from the graphite, an effect which becomes more pronounced with increasing peak stress. The effective area A_{eff} can be estimated by $A_{eff} = E_u / E_0$ where E_0 is the conventional elastic modulus.

A plot of E_u versus σ_{max} is an approximate straight line relationship, with a slope of m_u *(Figure 2.59)*.

m_u is therefore a second material parameter.

E_u is then given by $m_u \sigma_{max}$ and effective area is $A_{eff} = 1 + \dfrac{m_u \sigma_{\max}}{E_0}$

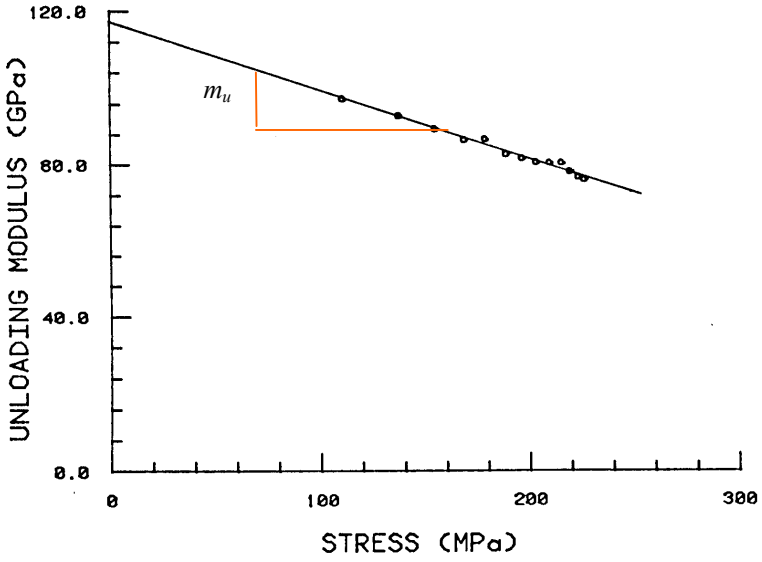


Figure 2.59 Unloading modulus versus peak tensile stress.

2.12.4 Compressive crack closure

The surface cracks are at their widest when the maximum stress in the loading history occurs. As the stress is reduced the cracks are closed. This crack closure stress is given by

$$\sigma_{cc} = Q(\varepsilon_{\max} - \varepsilon)^q$$

where Q and q are the coefficient and exponent of crack closure response.

Q and q are evaluated from boundary conditions of the cycle and from standard material properties. They do not represent additional materials properties.

The total stress to be used in the stress-strain response is then defined by the equation

$$\sigma = A_{eff}(\sigma_B + \sigma_G) + (1 - A_{eff})\sigma_{cc}$$

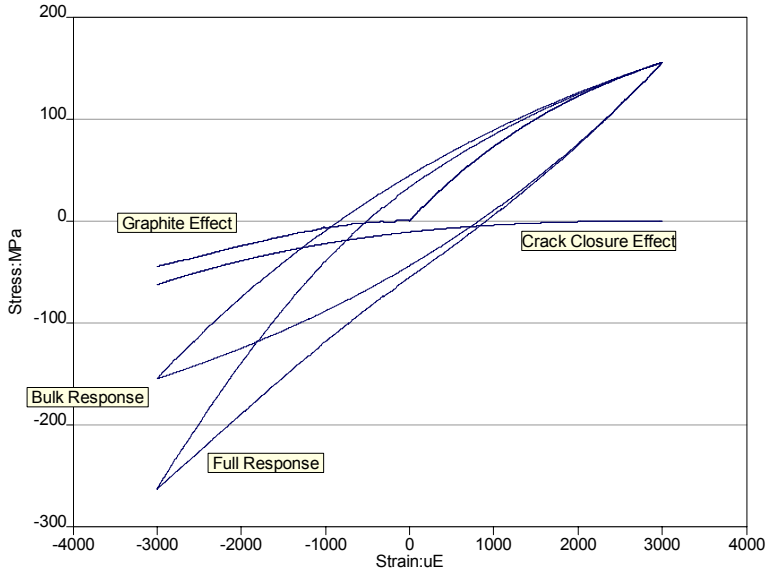


Figure 2.60. The components of the stress-strain response.

2.12.5 Fatigue damage calculation

The life is summed using a Continuum Damage Model. This considers that the rate at which damage occurs is not linear, but is related to the damage already accumulated. An incremental damage procedure calculates the number of repeats of the loading which gives a Miner's summation of 0.1. The damage parameters are modified as described below, and the procedure calculates the number of repeats of the loading to give another increment of 0.1 in Miner's summation. This procedure is repeated until Miner's summation equals 1.0

The incremental fatigue damage is calculated using:

$$\Delta D = \frac{(1 - D_i)^{Pi}}{(P_i + 1)N_{fi}}$$

where ΔD is the damage for the cycle, in the current damage increment

D_i is the damage so far accumulated

P_i is the damage rate parameter so far

N_{fi} is the endurance of the cycle.

P_i for a cycle is defined by the relationship :

$$Pi = 2.55(\sigma_{\max}\varepsilon_a)^{-0.8}$$

These equations become the conventional linear Miner's summation if Pi is set equal to zero.

2.12.6 Mean stress correction.

A modified Smith Watson Topper mean stress correction may apply. Unlike the conventional version of the SWT mean stress correction for steels, for grey irons the SWT relationship seems to be linear on a log-log plot of the SWT parameter versus endurance in cycles *(Figure 2.61)*.

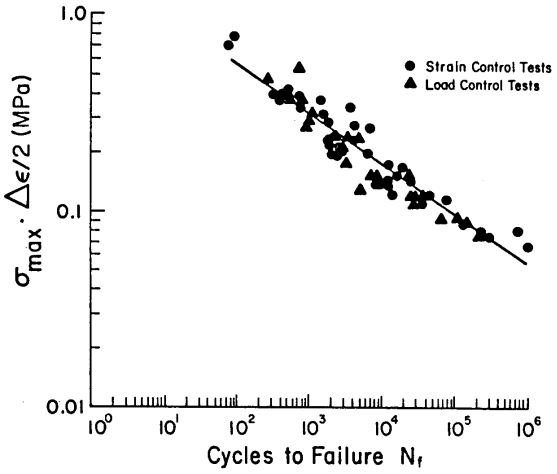


Figure 2.61. The Smith-Watson-Topper relationship for grey iron.

The SWT life curve for grey irons is defined by two parameters - a slope and an intercept at 1 cycle. These parameters should be determined experimentally.

For grey irons there is some experimental evidence for a 'knee' in the curve, with a second slope *b2* at higher values of endurance. This must be determined experimentally

NOTE that the SWT mean stress correction assumes that cycles which are totally compressive (in terms of local stress) are non damaging.

2.13 Applicability of local strain analysis

The main assumptions of local strain fatigue analysis are :

1. The fatigue life is determined by the local stress-strain response at the notch
2. The relationships between elastic strain and endurance, and between plastic strain and endurance, are linear on log-log axes
3. The hysteresis loop shape is determined by Masing's hypothesis.
4. The effect of mean stresses can be allowed for by using a one of the proposed mean stress corrections - Smith-Watson-Topper or Morrow for example
5. The effect of stress concentrations can be incorporated using the Neuber or Glinka relationships, or, if gross yielding occurs, by the Seeger-Heuler relationship.
6. Miner's rule can be used to calculate the total fatigue damage from a strain history, and a fatigue crack will initiate when Miner's summation equals 1.0

Although it is possible to solve the local strain equations with great accuracy (of the order of $1\mu\varepsilon$ over a signal containing many millions of data points), the relationships themselves are empirical, and there will be some error in calculated lives. It is probable that for many components made from conventional steels, achieved fatigue lives will be between 50% and 200% of the calculated life. (But see Chapter 13 on the effect of small cycles).

Local strain analysis does not take into account the size of the component, and experimental evidence indicates that a large component will have a shorter life than a small component of the same shape, for the same local strain history. Nor does the analysis allow for the effect of strain gradient. The effect of stress gradient is discussed in Chapter 6.

One of the most powerful features of local strain analysis is the ability to separate nominal and local stress-strain behaviour. *Figure 2.62* shows a test specimen loaded in constant amplitude 4-point bending. The nominal stresses and strains in the top of the specimen will cycle between zero and compression. If the loads are sufficiently large, yielding will occur in a small volume of material at the notch. When the load returns to zero, the local strain at the notch will return to a value close to zero. This is because the bulk of the unyielded material near the notch will 'pull' back to zero strain the

material which has yielded. Strains at a notch are usually considered to be under ‘strain control’, even if the overall loading on a component is load-controlled.

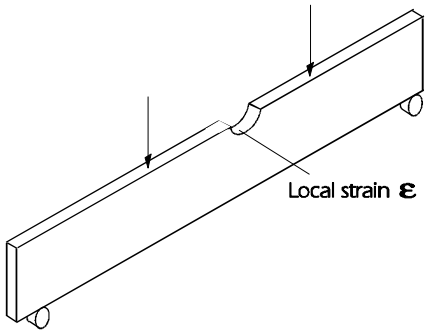


Figure 2.62 Test specimen in 4-point bending

The local stress-strain response is shown in *Figure 2.63*. The first strain application, from O to A, produces compressive yielding. When the load is removed, (point B), the local strain returns to zero (or something very close to it), leaving a residual tensile stress at load point B. All the subsequent cycles B-C-D etc. will produce tensile stresses at the notch each time the load is removed, and fatigue crack initiation is therefore possible.

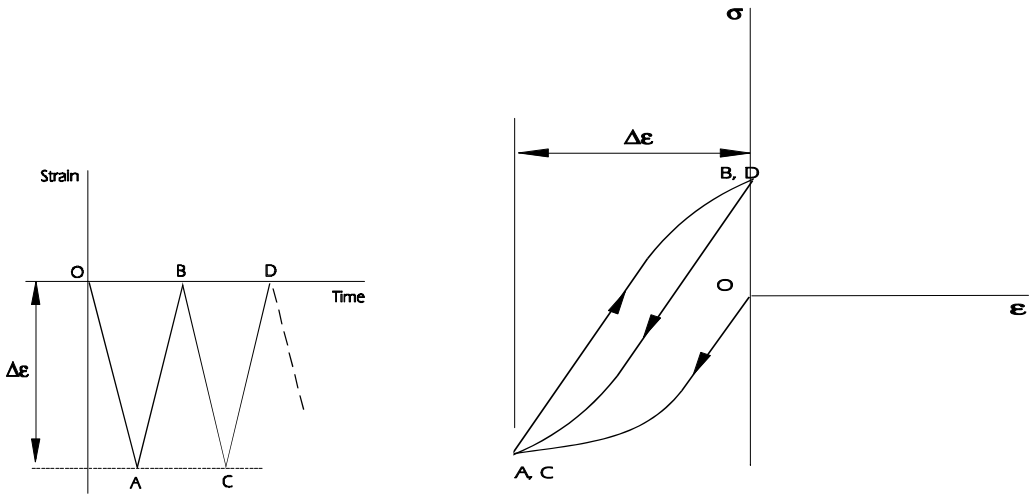


Figure 2.63 Local stress-strain response to compressive load cycles

This example is taken from a test performed in the 1970’s to investigate so-called ‘compressive fatigue failure’ in a machined component, and shows that although the nominal stress history is purely compression, the local stress cycles between compression and tension.

2.14 Summary of strain-life equations

The relationship between true local stress amplitude and endurance is

$$\frac{\Delta\sigma}{2} = \sigma_f'(2N_f)^b$$

The relationship between true local strain amplitude and endurance is

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma_f'}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c$$

The Smith-Watson-Topper mean stress correction is

$$\frac{\Delta\varepsilon}{2}\sigma_{max} = \frac{\sigma_f'^2}{E}(2N_f)^{2b} + \varepsilon_f'\sigma_f'(2N_f)^{b+c}$$

Morrow's mean stress correction is

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma_f' - \sigma_m}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c$$

The cyclic stress strain curve is defined by

$$\varepsilon = \frac{\sigma}{E} + \left(\frac{\sigma}{K'}\right)^{\frac{1}{n'}}$$

The hysteresis loop shape is defined by

$$\Delta\varepsilon = \frac{\Delta\sigma}{E} + 2\left(\frac{\Delta\sigma}{2K'}\right)^{\frac{1}{n'}}$$

Neuber's rule is

$$\Delta\sigma\Delta\varepsilon = K_t\Delta S.\,K_t\Delta e$$

2.15 References

2.1 Basquin O H
The Exponential Law of Endurance Tests
Am Soc Test Mater Proc. Vol 10, 1910, pp625-630

2.2 Coffin Jr L F
A Study of the Effects of Cyclic Thermal Stresses on a Ductile Metal
Trans ASME Vol 76, 1954 pp931-950

2.3 Masing G
Eigensannungen und Verfestigun im beim messing
Proc Second Int Cong Appl Mech, Zurich 1926

2.4 Smith K N, Watson P, Topper T H
A Stress-Strain Function for the Fatigue of metals
J Mater, Vol 5, No 4, 1970, pp767-778

2.5 Morrow J
Fatigue Design Handbook
Society of Automotive Engineers, 1968, Sec 3.2, pp21-29

2.6 Neuber H

Theory of Stress Concentration for Shear Strained Prismatic Bodies with Arbitrary Non-Linear Stress-Strain Law.

J Appl Mech, Trans ASME, Vol E28, 1961

2.7 Glinka G

Calculation of Inelastic Notch-Tip Stress-Strain Histories Under Cyclic Loading

Engnr. Fract Mech Vol 22, No 5, 1986, pp839-854

2.8 Seeger T, Heuler P

Generalised Application of Neuber's Rule

J Test Eval, Vol 8, No 4, 1980, pp199-204

2.9 Downing S D

Modelling Cyclic Deformation and Fatigue Behaviour of Cast Iron Under Uniaxial Loading

Report No 101, Ref UILU 84-3601, University of Illinois, 1984.

3 Local strain materials data

3.1 Introduction

Chapter 2 described the cyclic behaviour of materials, and the relationships between endurance and stress or strain amplitude.

The equation for the cyclic stress strain curve is

$$\varepsilon = \frac{\sigma}{E} + \left(\frac{\sigma}{K'}\right)^{\frac{1}{n'}} \qquad (3.1)$$

The equation for the hysteresis loop stress-strain curve is

$$\Delta\varepsilon = \frac{\Delta\sigma}{E} + 2\left(\frac{\Delta\sigma}{2K'}\right)^{\frac{1}{n'}} \qquad (3.2)$$

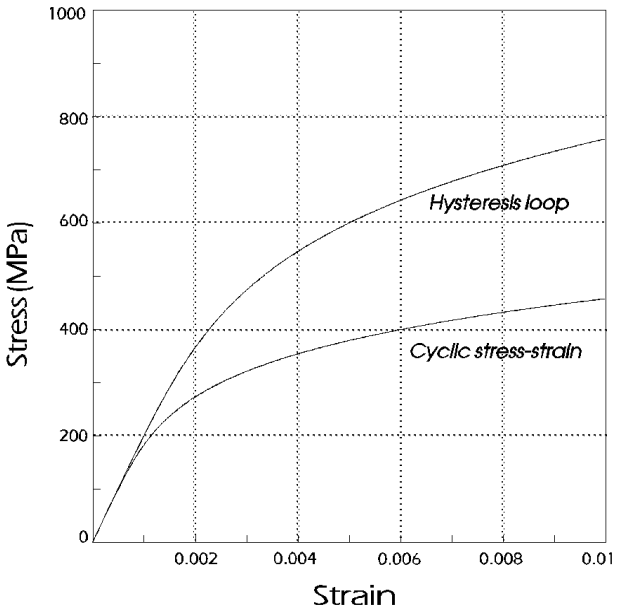


Figure 3.1 Stress-strain relationships

The relationship between true local stress amplitude and endurance is

$$\frac{\Delta\sigma}{2} = \sigma_f'(2N_f)^b \qquad (3.3)$$

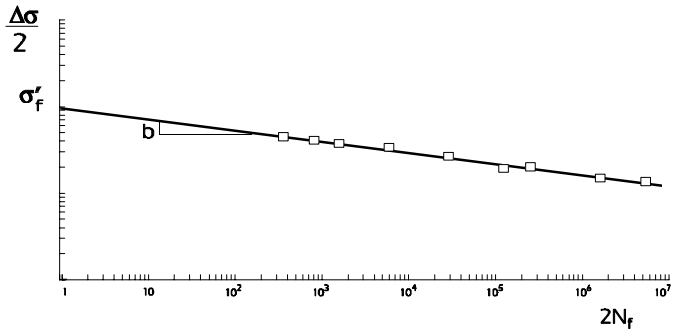


Figure 3.2 Stress-life relationship

The relationship between true local strain amplitude and endurance is

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma_f'}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c \qquad (3.4)$$

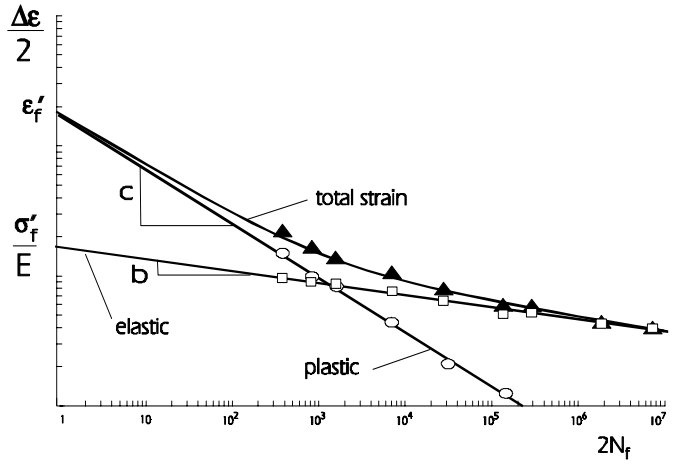


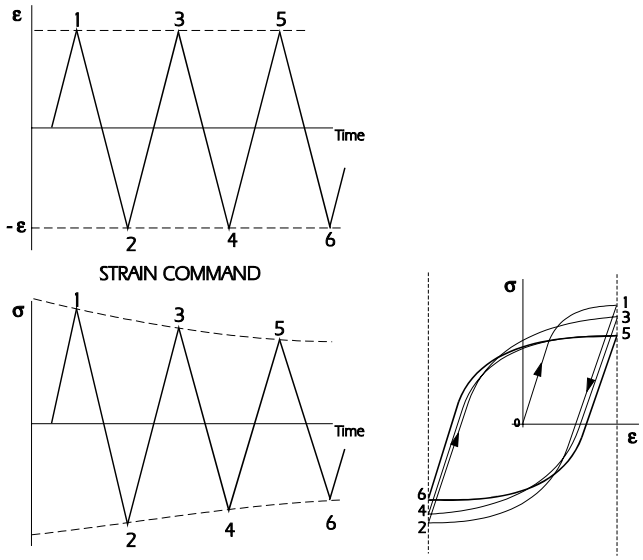
Figure 3.3 Strain-life relationship

Seven materials properties are used in these equations

E	the elastic modulus (Young's Modulus)
K'	the strain hardening coefficient
n'	the strain hardening exponent
b	the fatigue strength exponent (Basquin's exponent)
σ_f'	the fatigue strength coefficient
c	the fatigue ductility exponent (the Coffin-Manson exponent)
ε_f'	the fatigue ductility coefficient

3.2 The cyclic stress- strain curve

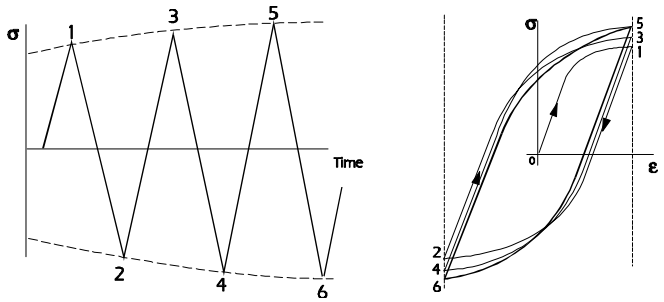
When a smooth test specimen is cycled between fixed strain limits, the stress response may show that the material is softening, producing lower stresses for each strain application. After a number of cycles the cyclic stress-strain curve stabilises, producing a stable hysteresis loop. *(Figure 3.4)*



stress response cyclic stress-strain

Figure 3.4 Cyclic softening

For other materials the stress response may show that the material is hardening.



stress response cyclic stress-strain

Figure 3.5 Cyclic hardening

Cyclic hardening and softening were first reported by Bauschinger in 1886, and quoted by Unwin later in the same year.

In Chapter 2 it was shown that the first strain excursion (of a cyclically stabilised material) follows the cyclic stress strain curve, and that subsequent strain excursions follow the hysteresis stress-strain curve.

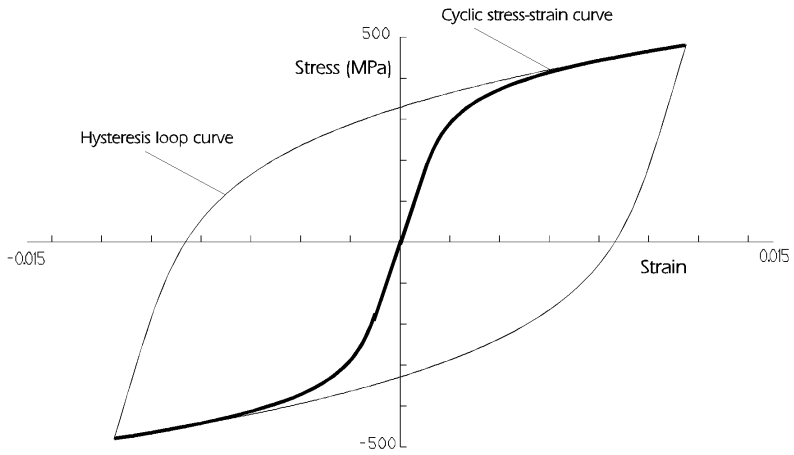


Figure 3.6 Cyclic and hysteresis loop curves

Other strain amplitudes will produce stable hysteresis loops of different sizes, but all the loops will have their tips on the cyclic stress strain curve.

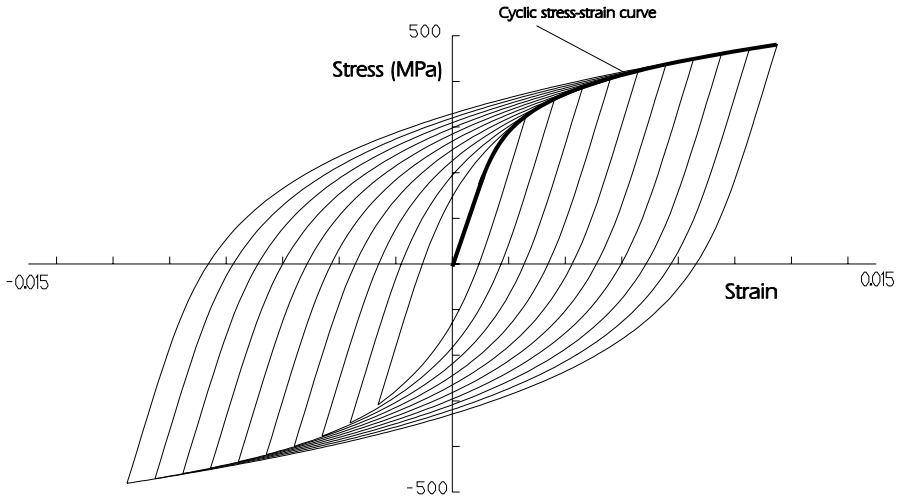


Figure 3.7 Cyclic stress-strain curvethrough the hysteresis loop tips

Figure 3.8 compares the monotonic tensile stress-strain curves produced in a tensile test, and stable cyclic stress strain curves for three typical materials.

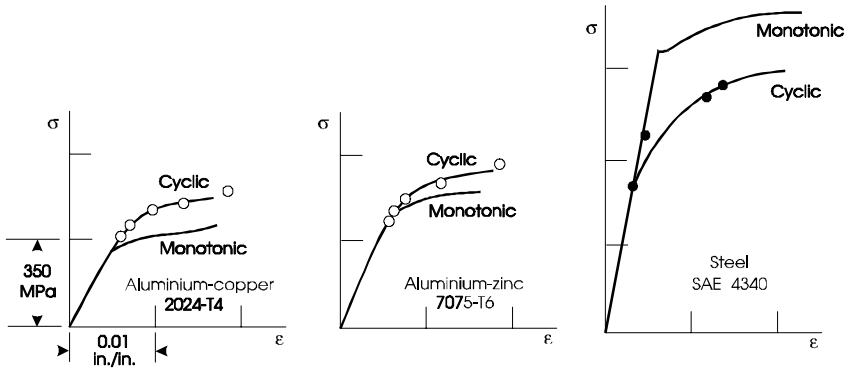


Figure 3.8 Monotonic and cyclic stress-strain curves

The material UTS and 0.2% proof stress σ_y can be used to predict whether a steel will cyclically harden or soften.

If $\frac{UTS}{\sigma_y} > 1.4$ the material will harden

and if $\frac{UTS}{\sigma_y} < 1.2$ the material will soften

In smooth specimen testing a specimen is cycled between fixed strain limits until the material stabilises. Cyclic stability is indicated by the force required to achieve the specified strain. The stress at the loop tips can be calculated from the applied load and the specimen diameter under load. Specimen diameter may be measured, or may be calculated using the axial strain and Poisson's ratio. Other specimens are tested at different strain amplitudes. The set of stress-strain co-ordinates for the loop tips may be plotted. They will lie on the cyclic stress-strain curve.

Similar but not identical results can be obtained from an incremental step test, which uses one specimen.

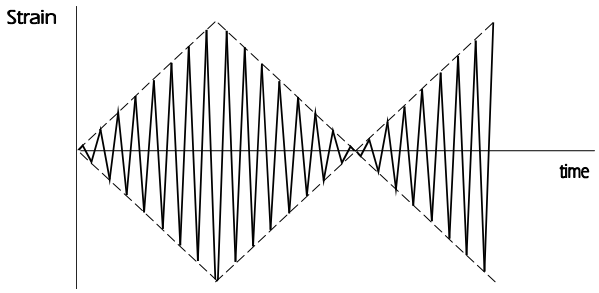


Figure 3.9 Applied strains for an incremental step test

The specimen is subjected to a series of increasing and decreasing strain amplitudes. Once the material has stabilised, the stresses for each strain amplitude can be plotted to show the cyclic stress-strain curve.

The data from both types of test can be analysed to determine the material properties n', K' and E.

The equation for the cyclic stress-strain curve is

$$\varepsilon = \frac{\sigma}{E} + \left(\frac{\sigma}{K'}\right)^{\frac{1}{n'}}$$

Taking the plastic component of strain,

$$\varepsilon_p = \left(\frac{\sigma}{K'}\right)^{\frac{1}{n'}}$$

Re-arranging this equation gives

$$\sigma = K' (\varepsilon_p)^{n'}$$

Taking logs of both sides of this equation

$$log_{10} \sigma = log_{10} K' + n' log_{10} (\varepsilon_p)$$

which is the equation for a straight line on log-log axes.

The data points are plotted on axes of $log_{10}(\sigma)$ vs $log_{10}(\varepsilon_p)$. *(Figure 3.10)*

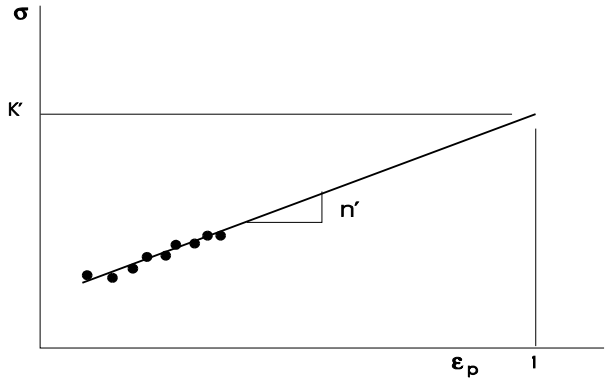


Figure 3.10 Calculation of n' and k'

A least-squares regression analysis may be used to calculate the best-fit straight line through the data. The slope gives the value of n'.

For an extrapolated strain of $\varepsilon_p = 1$, in the equation

$$log_{10}\,\sigma \;=\; log_{10}\,K' \;+\; n'\,log_{10}\,(\varepsilon_p)$$

if $\varepsilon_p = 1$, then $log_{10}(\varepsilon_p) = 0$ and so $\sigma = K'$

3.3 The strain-life curve

Constant amplitude tests on a number of specimens, each at a different strain amplitude, are used to produce a strain-life curve. The tests are performed between fixed strain limits. Initially the material stabilises, indicated by a change in stress for the applied strain, followed by a long period of stable behaviour. During this period, the stress and strain co-ordinates of the stable hysteresis loop can be measured. Eventually the stress reduces, indicating that a fatigue crack is propagating, and the endurance for the specimen is usually based a 5% or 10% reduction in stress for the applied strain, which indicates that a small crack is present.

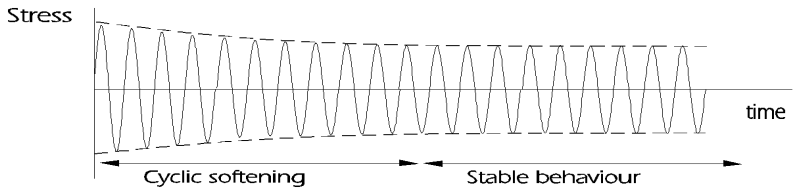


Figure 3.11 Stress response to constant amplitude strain cycling

The strain-life equation may be separated into the elastic and plastic components, each of which is a straight line on log-log axes.

$$\frac{\Delta\varepsilon_e}{2} \;=\; \frac{\sigma_f'}{E}(2N_f)^b$$

$$\frac{\Delta\varepsilon_p}{2} \;=\; \varepsilon_f'(2N_f)^c$$

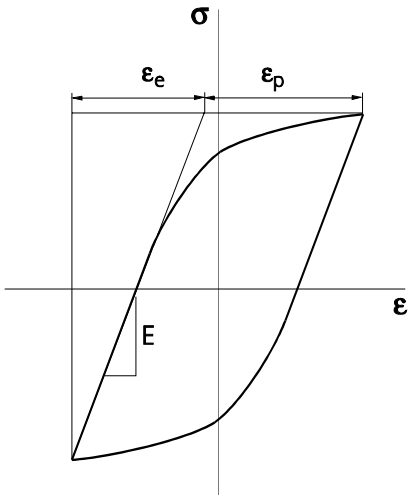


Figure 3.12 The cyclic stress-strain hysteresis loop as the sum of elastic and plastic strains

The elastic and plastic components of strain can be determined for each strain amplitude (see *Figure 3.12*) using the equation for the hysteresis stress-strain curve (3.2). Regression analysis is used to give a best fit straight line for the elastic and plastic curves, and provide the four parameters (*Figures 3.13* and 3.14). The two curves are added to give the strain-life curve, as shown in *Figure 3.15*.

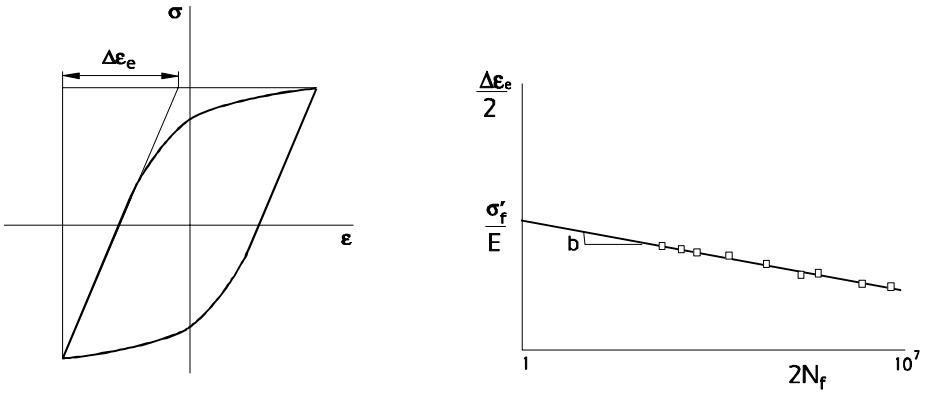


Figure 3.13 Relationship between elastic strain and endurance

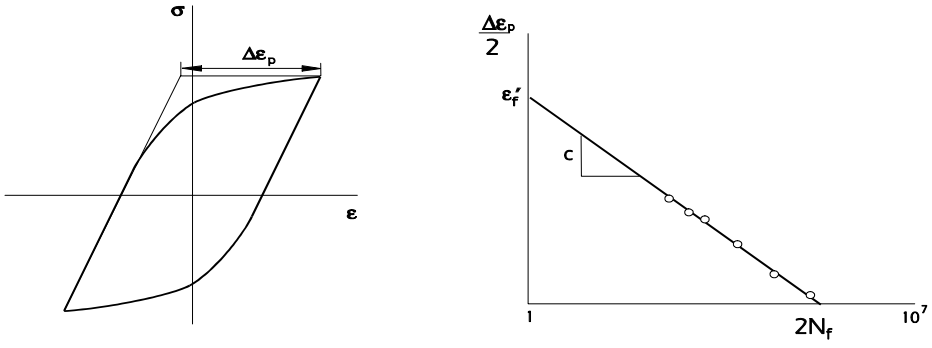


Figure 3.14 Relationship between plastic strain and endurance

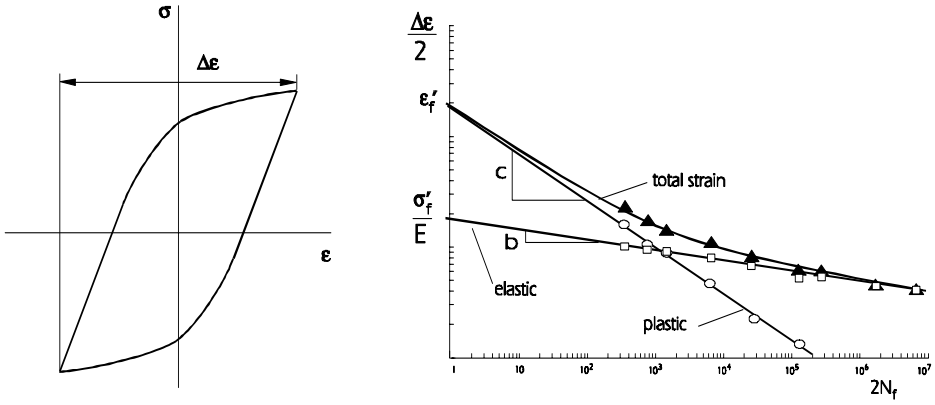


Figure 3.15 Example of test data to determine the strain-life curve

3.4 Practical testing

Local strain fatigue analysis can provide accurate fatigue lives if the materials properties are determined carefully. The following points can be noted.

The tests are at constant strain amplitudes between fixed strain limits. These may be computer-controlled for increased precision.

Axial strains are measured directly using a clip gauge, and stresses are determined from a load cell.

The true stress is calculated from the load and instantaneous cross-sectional area, which is calculated from the instantaneous axial strain and Poisson's ratio.

The results, as pairs of data points of true strain and true stress, are digitised and stored on computer to allow the hysteresis loops to be plotted.

A precise definition of specimen failure is used - usually a 5% or 10% reduction in stress response to the applied strain amplitude. This is generally taken to be equivalent to a short crack of 1 - 2 mm length.

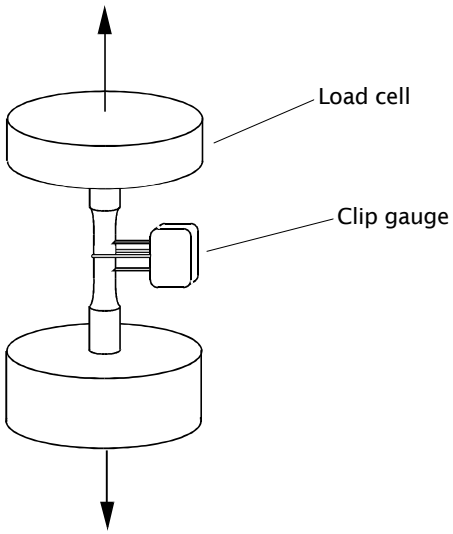


Figure 3.16 Smooth specimen material's test

Great care must be taken to ensure that the specimens are loaded axially, with no extraneous bending forces.

3.5 Approximations for other materials

3.5.1 Simple Approximations

A number of approximate relationships have been proposed for the seven materials properties E, b, c, σ_f', ε_f', n' and K'.

The fatigue strength coefficient σ'_f is approximately equal to the true fracture strength, σ_f. In Section 2.2 the true fracture strength (or stress) was shown to be

$$\sigma_f = \frac{P_f}{A_f}$$

where P_f is the load at fracture, and A_f is the cross-sectional area at fracture.

The fatigue ductility coefficient, ε'_f is approximately equal to the true fracture ductility ε_f. In Section 2.2 the true fracture ductility was shown to be

$$\varepsilon_f = ln\left(\frac{A_o}{A_f}\right)$$

where A_o is the original cross-sectional area.

The fatigue strength exponent, b, will usually be between -0.05 and -0.12.

The fatigue ductility exponent, c, will usually be between -0.5 and -0.7.

The transition life can also be approximated. This is the endurance at which the elastic and plastic strain amplitudes are equal

$$2N_f = \left(\frac{\varepsilon'_f E}{\sigma'_f}\right)^{\frac{1}{b-c}}$$

From equations (3.1) and (3.3) the following relationships can be derived, which may be valid for some materials.

$$K' = \frac{\sigma'_f}{(\varepsilon'_f)^{n'}} \qquad n' = \frac{b}{c}$$

3.5.2 Four-Point Correlation

This method (Ref 3.2) is shown in *Figure 3.17*. Note axes of $\Delta\varepsilon$ and N_f

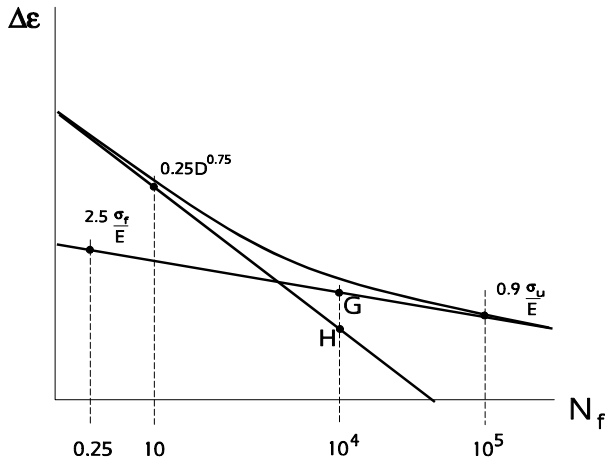


Figure 3.17 4-point correlation method

σ_f is the true fracture stress.

σ_u is the conventional monotonic ultimate tensile stress.

D is the logarithmic ductility of the material, $\ln(1\text{-RA})^{-1}$, where RA is the reduction in area.

H is given by $\frac{0.0132 - G}{1.91}$ where G is $\Delta\varepsilon_e$ at $N_f = 10^4$ cycles.

3.5.3 Universal Slopes

This method (Ref. 3.2) assumes that the elastic and plastic lines for all materials have the same slopes, b = -0.12 and c = -0.6.

The intercept at $2N_f = 1$ for the elastic line : $\frac{\sigma'_f}{E} = 3.5\,\frac{\sigma_u}{E}$

for the plastic line : $\varepsilon'_f = [\ln(1\text{-RA})^{-1}]^{0.6}$

3.5.4 Method of Socie et al

This method (Ref. 3.3) proposes that the elastic line has an intercept $\sigma'_f = \sigma_f$ and the slope is obtained using a point with an ordinate $0.5\sigma_u$ at $2N_f = 10^4$ reversals. The plastic line is obtained by assuming $\varepsilon'_f = \ln(1\text{-RA})^{-1}$ and a value of c = -0.6 for a ductile material or c = -0.5 for a high-strength material.

For these methods, if the true fracture stress is not known, it may be estimated using $\sigma_f = \sigma_u$ [1 + logarithmic ductility]

3.5.5 Comparison of the Methods

In (Ref. 3.1) fatigue lives are calculated for 49 steels using published values of σ'_f, ε'_f, b and c. These lives (termed 'experimental' lives) were compared with lives calculated using each of the approximation methods described above (termed 'estimated' lives). The lives covered a range from 10 to 10^7 reversals. The steels covered values of UTS from 345 MPa to 2585MPa, and Brinell hardness values from 80 to 660. They included the steels SAE1005, SAE1015, and SAE1045.

In all cases correlation between the 'experimental' and the 'estimated' lives was poor. The table shows the proportion of life estimates (using estimated materials properties) which fell within the specified factors of the lives calculated using measured properties. The 4-point correlation method was the best of the three. This produced estimated lives, for which 54% were within a factor of ± 2 of the lives calculated from real materials test data for the 4 parameters, and 86% were within a factor of ± 5.

In general, for all methods, the life predictions from estimated properties tended to be optimistic, and the scatter was greater than shown in the table at very short values of endurance, and at endurances greater than 10^6 reversals. For each method more than 5% of all the life estimates were outside a factor of ± 10, and all produced some estimates which were outside a factor of ± 20 of those produced from measured material properties.

METHOD OF LIFE ESTIMATES	LIFE FACTOR	%
4-POINT CORRELATION	±2	54
4-POINT CORRELATION	±5	86
UNIVERSAL SLOPES	±3	50
UNIVERSAL SLOPES	±5	80
SOCIE et al	±2	32
SOCIE et al	±5	59

If the results are expressed as allowable strain amplitude at a given life, the 4-point correlation method generally predicted allowable strains to within a factor of ± 2 of the true values, and was generally optimistic. The other methods were less successful.

When the estimated values of the four parameters were compared with test values, only the estimates of σ'_f showed any correlation with test values. Estimates of ε'_f, b and c showed large errors.

Strain-life fatigue, using measured materials properties, is generally considered to give mean life estimates that are within a factor of ±2 or ±3 of real test lives. The scatter from using estimated materials parameters must be added to these factors. Many life estimates will then fall outside scatter bands of ±10 or greater.

3.5.6 Seeger's Method

Seeger has published the following method for plain carbon and low to medium alloy steels, and for aluminium and titanium alloys.

Plain carbon and low/medium alloy steels		**Aluminium and titanium alloys.**	
σ'_f	$1.5\sigma_u$	σ'_f	$1.67\sigma_u$
ε'_f	0.59a	ε'_f	0.35
b	-0.087	b	-0.095
c	-0.58	c	-0.69
n'	0.15	n'	0.11
K'	$1.65\sigma_u$	K'	$1.61\sigma_u$

In the table a = 1.0 for σ_u/E less than 0.003, otherwise a = $1.375\text{-}125\sigma_u/E$. Experience shows that this method gives satisfactory results.

3.6 Additional testing for cast irons

Cast irons (and particularly grey iron) require additional materials data to define the assymetrical shape of the hystersis loop, and the non-linear fatigue damage accumulation.

The conventional strain-life parameters, σ_f', b, ε_f', c are required.

The conventional cyclic stress-strain parameters, n' and K' are required, for the tensile cyclic stress-strain curve. The elastic modulus is the value of E_0 described below.

In addition, n' and K' are required, for the compressive cyclic stress-strain curve. These are termed *n'-Comp* and *K'-Comp*.

The slope m of the plot of secant modulus versus stress, for both tensile and compressive stresses, are required. These are termed *m'-Tens_Sec_Slope* and *m'-Comp_Sec_Slope* .

The slope m_u, of the plot of the unloading modulus versus stress is required.

The slope (exponent) and intercept at 1 reversal, of the log-log plot of Smith-Watson-Topper parameter versus endurance.

The possible 'knee' in the SWT-life curve, with a slope b2.

Materials properties are determined from a series of constant amplitude strain-controlled fully-reversed fatigue tests, with each specimen tested at a different strain amplitude.

From each strain amplitude test:

a. For each measured loop, determine the peak tensile and peak compressive stresses, and the unloading modulus from the peak tensile stress.

b. Determine the unloading modulus for a series of peak tensile stresses *(Figure 3.18)* Plot the unloading modulus versus peak tensile stress (see *Figure 3.19*). Use linear regression to determine the slope m_u and the intercept E_0. A check on this analysis is that E_0 should correspond to the tangent modulus of the stress-strain curve at zero stress/strain.

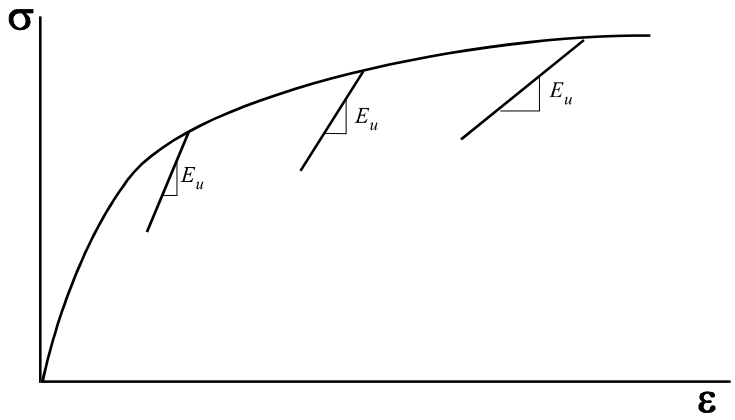


Figure 3.18. Schematic diagram of the unloading modulus

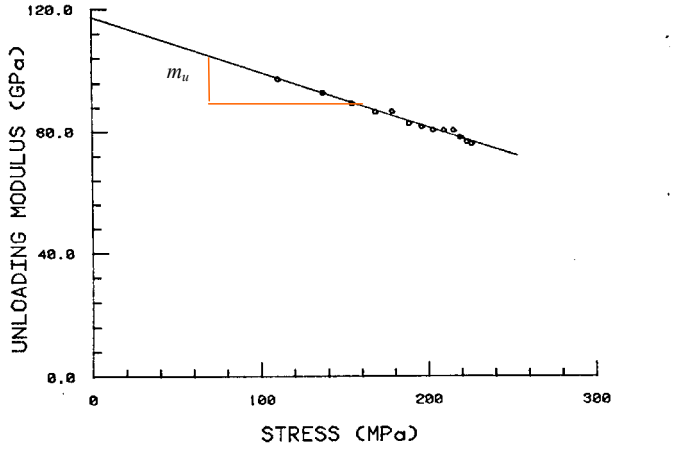


Figure 3.19 Plot of unloading modulus versus peak tensile stress

c. Initial peak tensile stresses are divided by their strain amplitude to give the secant modulus. As shown in *Figure 3.20*, there should be a linear section. Regression analysis will give the slope m and the intercept, which is another estimate of E_0.

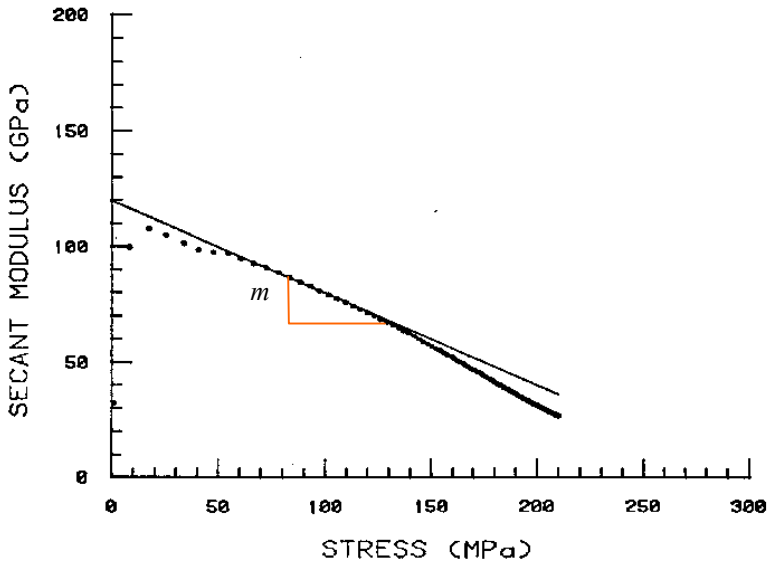


Figure 3.20. Plot of secant modulus versus stress

d. The value of m for compressive stresses is determined in a similar way.

To determine the peak tensile stresses to be used in steps b-d, Downing (Ref. 3.4) recommends that, for each test specimen, the peak tensile stress from each measured loop is plotted versus $(1 - N/N_f)$, on a log-log plot, where N is the cycle number at which the loop was measured, and N_f is the number of cycles to crack initiation *(Figure 3.21)*

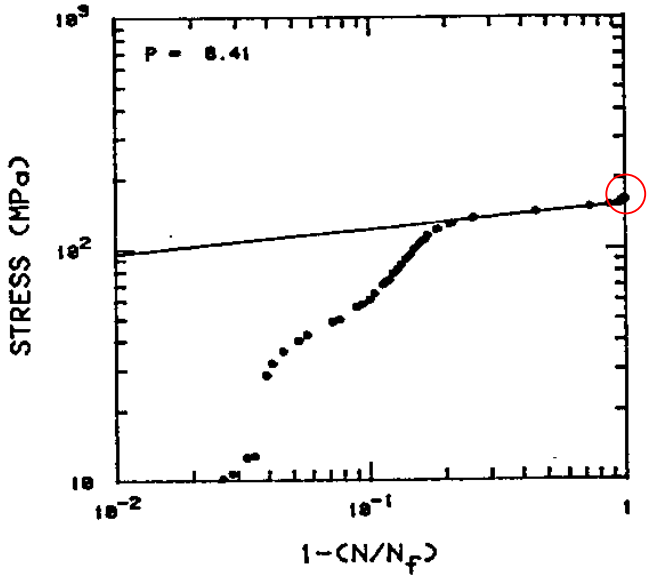


Figure 3.21 Realtionship between peak tensile stress and cycle number

The intercept of this plot at $(1 - N/N_f) = 1$ (effectively zero cycles), shown red, gives the required maximum stress.

The procedure is repeated with compressive stresses to determine the peak compressive stress.

3.7 Applicability of materials data

Local strain fatigue analysis can be used for many steels, and to some aluminium alloys. For cast irons the hysteresis loops are not symmetrical and the additional materials data (section 3.6) is required. This is particularly true for grey iron – nodular (SG) iron can often be modelled successfully by treating it as a steel.

Some aluminium alloys strain harden during smooth specimen testing, never attaining cyclic stability. An example is shown in *Figure 3.22*. The material continues to harden throughout the test life, with final stresses almost twice the initial stress for the same strain amplitude. For materials which produce symmetrical hysteresis loops but which are never really stable, it is usual to calculate *n'* and *K'* from hysteresis loops measured at the half-life point of each specimen. This would be difficult to justify for a material where the change in behaviour is as extreme as that shown in *Figure 3.22*, and the absence of stable behaviour would probably preclude the calculation of mean stresses and the application of Neuber's rule where significant plasticity was developed. When using published values of cyclic properties, additional test data on hysteresis loop shape, cyclic stability and the correlation coefficients obtained during the curve fitting should be examined if the data is available.

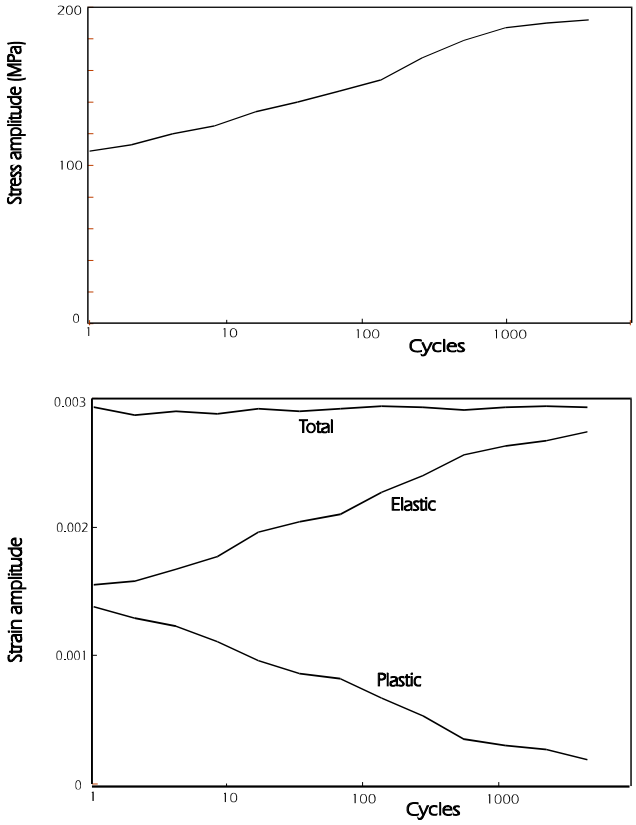


Figure 3.22 Cyclic hardening of an aluminium alloy sheet specimen

3.8 References

3.1 Ong J H
An evaluation of existing methods for the prediction of axial fatigue life from tensile data
Int J Fatigue vol 15 no 1 January 1993

3.2 Manson S
Fatigue: a complex subject - some simple approximations.
Exp Mech SESA (1965) pp193-226

3.3 Socie D F, Mitchell M R, Caulfield E M
Fundamentals of modern fatigue analysis
FCPR Report No 26 (University of Illinois, 1977)

3.4 Downing S D
Modelling Cyclic Deformation and Fatigue Behaviour of Cast Iron Under Uniaxial Loading
Report No 101, Ref. UILU 84-3601. University of Illinois, 1984

see also **Downing S D, Socie D F.**
Modelling Cyclic Deformation Behaviour of Grey Iron Under Uniaxial Loading
Proc SEECO 83, 'Digital Techniques in Fatigue', City University, London. pp152-175.

4 Signal processing for fatigue analysis

4.1 Introduction

Measurements of service histories - loads, strains, and accelerations - are required so that general information on service loading can be obtained, and so that the fatigue life of specific components can be determined.

Service histories must be analysed using methods which are relevant for fatigue analysis. In Chapter 2 it was shown that a cycle is a closed stress-strain hysteresis loop. Modern signal processing uses a cycle counting algorithm to extract these cycles quickly and accurately. However, cycle counting is a recent development, possible only with modern computer systems and solid state electronics, and many other less relevant methods developed in earlier days are still encountered.

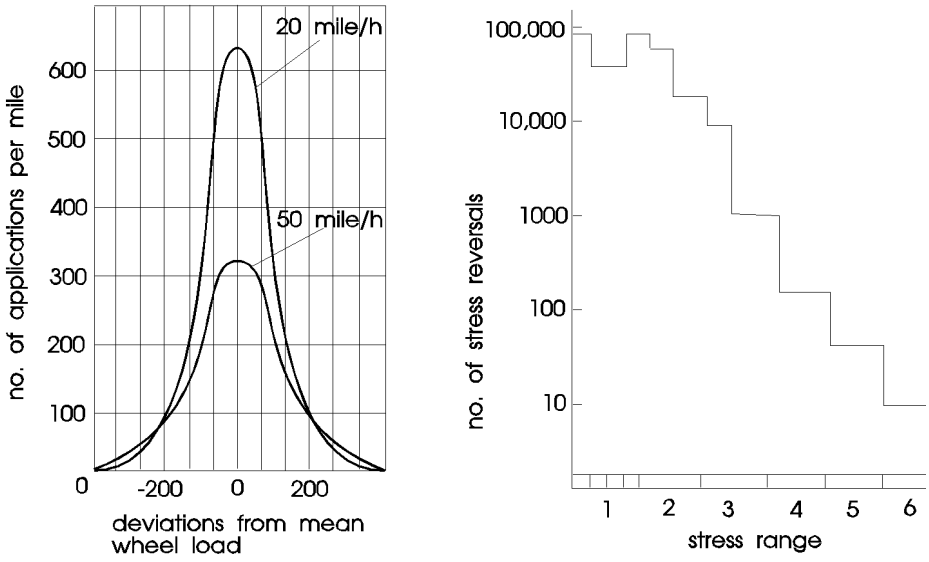


Figure 4.1 Some early methods of data presentation

Figure 4.1 shows two methods that have been used in the past in an attempt to present service loading in a way which was both concise and relevant for fatigue calculation. Some of these early methods will be described later in this chapter.

Modern signal processing for fatigue analysis uses Rainflow cycle counting, and this will be described first.

4.2 Rainflow cycle counting

Chapter 2 showed how a material responds to a series of applied loads or strains. It was shown that fatigue cycles are closed stress-strain hysteresis loops.

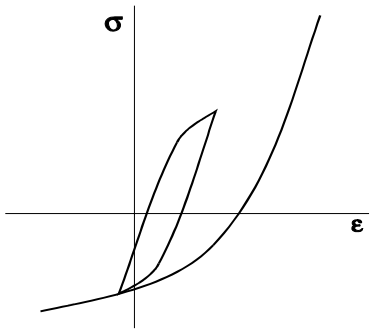


Figure 4.2 A cycle is a closed stress-strain hysteresis loop

The closure of the loops is quite complex, in that the loop tips can be formed from points in the signal which are separated by a large number of intermediate points. An algorithm is required which correctly determines the cycles present in a signal. As the tips of the cycle are formed by a peak and a valley in the signal, the intermediate data points between each peak and valley need not be considered.

Consider the following sequence of peaks/valleys *(Figure 4.3)*. The notation uses point A as the most recent data point, point B as the previous point and so on.

Because the range from A to B is greater than the range from B to C, a cycle is closed, and is represented by the range B to C. One criterion for cycle closure is then that the range between the two most recent peak/valleys must be greater than the preceding range, i.e.

range A to B > range B to C

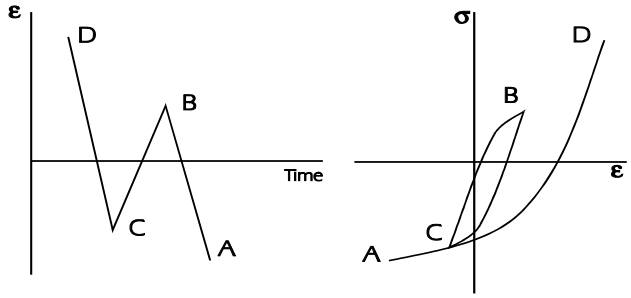


Figure 4.3

A more complex sequence is shown in *Figure 4.4*.

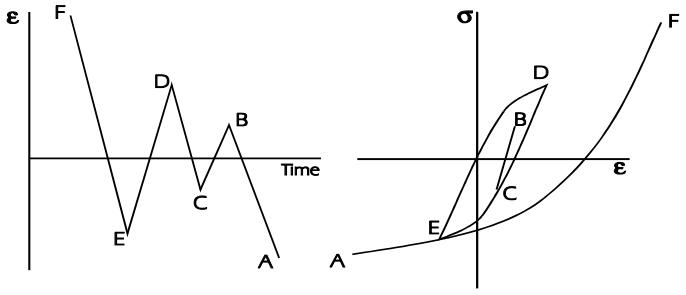


Figure 4.4

As before, the range from A to B is greater than that from B to C, so B to C is one cycle. However, the hysteresis loops show that another cycle has also been formed, with its tips at points D and E. This is closed because the range from A to D is greater than the range from D to E. This becomes clearer if, after closing the first cycle, from B to C, the signal is re-drawn as though this cycle had never existed *(Figure 4.5)*

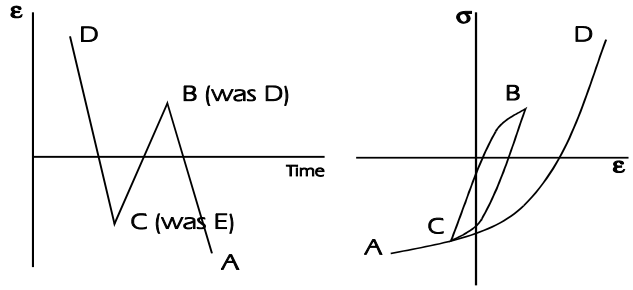


Figure 4.5

From this example it is clear than the most recent data point can cause more than one cycle to close. The simple criterion developed from the first example, that a cycle is formed if

range A to B > range B to C

can now be extended to

a cycle is formed if
range A to B > range B to C
extract the cycle, re-label the data points

a cycle is formed if
range A to B > range B to C

and carry on this procedure until no more cycles are closed by this point.

This is the basis of the cycle counting method published by Socie and Downing (Ref 4.1) and now used in many cycle counting programs. An example of the FORTRAN code for the algorithm is shown in *Figure 4.6*. It maintains a data array, or buffer, of peak/valleys which represents unclosed cycles.

```
INTEGER BUFFER (4096), INDEX, VALUE, RANGE, MEAN, X, Y

INDEX = 0
10  CONTINUE
    call 'get next peak/valley', VALUE
    INDEX = INDEX + 1
    BUFFER (INDEX) = VALUE
20  CONTINUE
    IF (INDEX.LT.3) THEN
            not enough points to form a cycle
            GOTO 10
    ELSE
            X = ABS (BUFFER(INDEX)      - BUFFER(INDEX  - 1))
            Y = ABS (BUFFER (INDEX - 1)  - BUFFER(INDEX -2 ))
            IF (X.GE.Y) THEN
```

```
c --  cycle has been closed
                RANGE = Y
                MEAN = (BUFFER(INDEX-1) - BUFFER(INDEX-2))/2
c--   remove the cycle
                INDEX = INDEX  - 2
                BUFFER(INDEX) = BUFFER(INDEX+2)
c--   see if this value closes any more cycles
                GOTO 20
            ELSE
                GOTO 10
            END IF
    END IF
```

Figure 4.6 Simple rainflow cycle algorithm

The analysis is known almost universally as *Rainflow cycle counting*, or simply as Rainflow. The term Rainflow derives from an earlier algorithm, proposed by Endo et al. (Ref. 4.2). In this algorithm the signal was turned through 90^{o}, and rain was envisaged as falling on the signal and dropping from surface to surface. Various rules were proposed for what happened to the rain, and the resulting algorithm correctly extracted each half-cycle, which eventually paired with another half cycle to make a complete cycle. The Rainflow method was a most important development at the time, because it provided a genuine method of extracting fatigue cycles from measured signals. It has been replaced by the more simple Socie-Downing algorithm.

It is interesting to note that the paper by Endo et al. (Ref. 4.2) includes not only the Rainflow algorithm but also another algorithm more closely related to hysteresis loop closure, and providing the same result in terms of cycles counted. Other methods - 'racetrack' and 'reservoir' - have been also proposed for analysing very short signals by hand, and give the same result as Rainflow counting.

The Rainflow algorithm will now be applied to a short length of signal, and the cycles extracted simply by comparing successive ranges. The signal could be any measured parameter, not necessarily a strain signal.

The signal starts at point A (the absolute maximum point in the signal). The first 3 points are A, B, and C.

Point D is the first point to consider. The range CD is greater than BC, so a cycle is closed. This is the cycle of range BC. Because the cycle was closed by a downward-going excursion, it is sometimes drawn as shown in *Figure 4.8*.

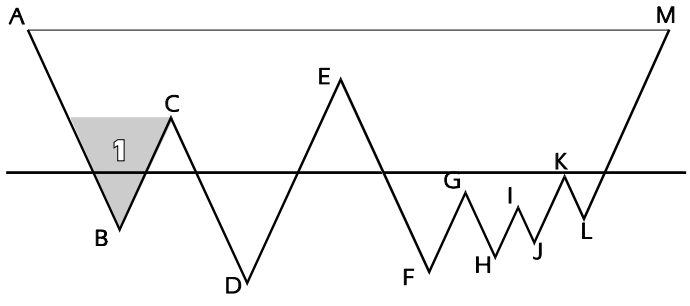


Figure 4.7

This shows more clearly that the cycle closed because CD was greater than BC.

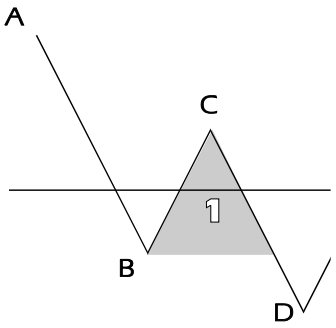


Figure 4.8

The signal can now be redrawn with this cycle removed:

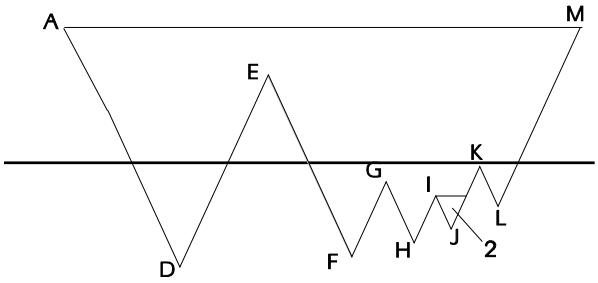


Figure 4.9

Taking each range in turn:

DE is less than AD, so no cycle is closed.
EF is less than DE, so no cycle is closed.
FG is less than EF, so no cycle is closed.
GH is less than EF, so no cycle is closed.
HI is less than GH, so no cycle is closed.
IJ is less than HI, so no cycle is closed.

JK is greater than IJ, so a cycle is closed. This is the cycle labelled 2, and has a range from I to J.

If cycle 2 is now removed, the range HK is also greater than GH, so point K closes another cycle, labelled 3. Cycle 3 can also be removed. Point K has therefore closed two cycles.

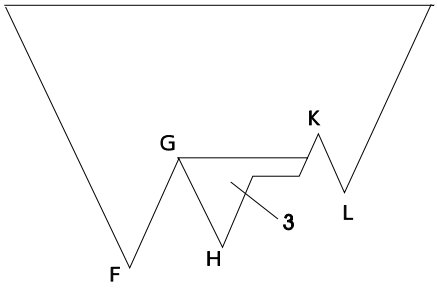


Figure 4.10

The range KL is less than FK, so no cycle is closed.

The range LM is greater than KL, so a cycle is closed *(Figure 4.11)*.

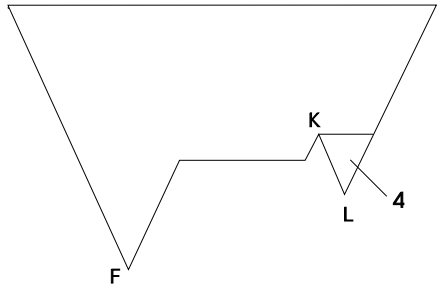


Figure 4.11

This has a range from K to L, and is labelled cycle 4. Removing this cycle leaves the following data points :

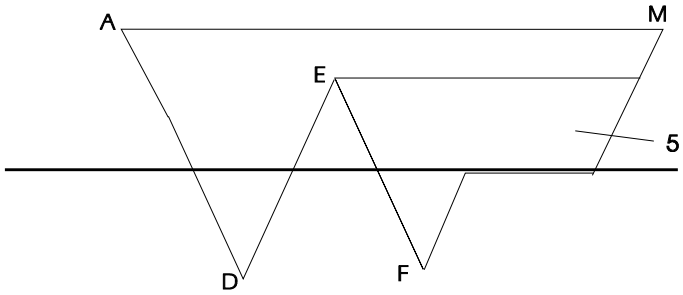


Figure 4.12

FM is greater than EF, so a cycle is closed with range EF, labelled 5. If this cycle is removed only three points remain, A, D and M, and these form the final cycle.

The six cycles extracted are summarised in *Figure 4.13*.

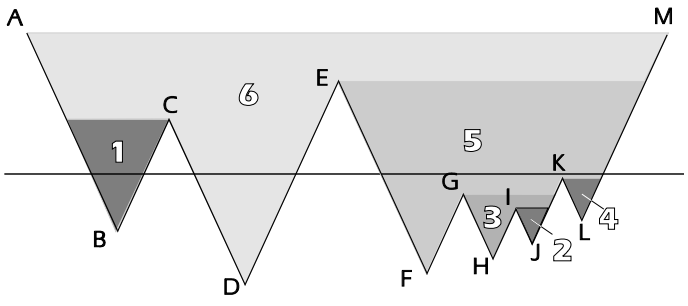


Figure 4.13

The cycles close in the order in which they are numbered, and the cycles have the following ranges:

Cycle number	Cycle range
1	B-C
2	I-J
3	G-H
4	K-L
5	E-F
6	A-D (= D-M)

Any parameter, such a load or deflection, can be cycle counted in this way.

The results of the Rainflow cycle analysis of a signal can be displayed as a histogram showing the distribution of cycle ranges.

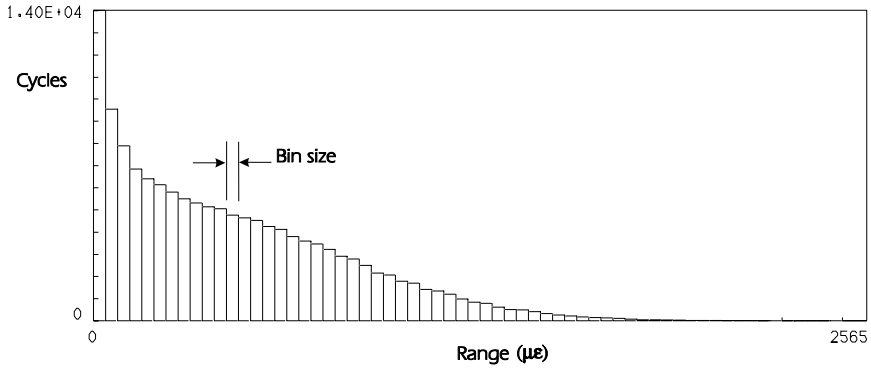


Figure 4.14 Histogram of cycle ranges

Histograms produced from different analyses may be difficult to compare, as the number of cycles in any bin depends on the bin size used.

If the number of cycles in each bin is divided by the bin size, a cycle density diagram is produced, where the area between any two range values represents the number of cycles with ranges between these values.

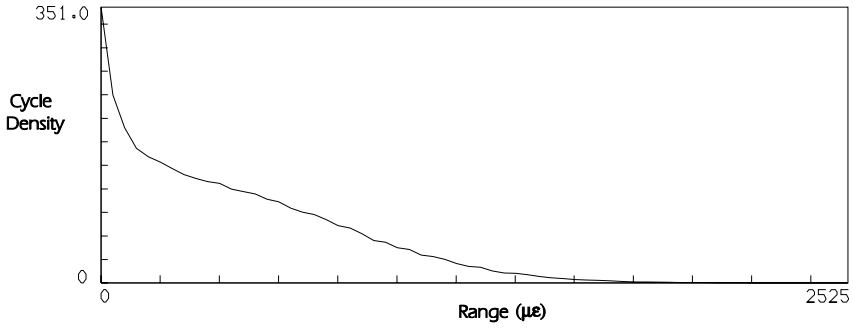


Figure 4.15 Cycle density diagram

As a cycle density diagram is independent of the bin width chosen for the analysis, it may be used to compare cycle distributions of signals obtained at different times and by different analysis programs.

The cycle density diagram may be integrated from the right hand side. This produces a cycle exceedence diagram, where the horizontal axis shows the number of cycles which exceed a given range.

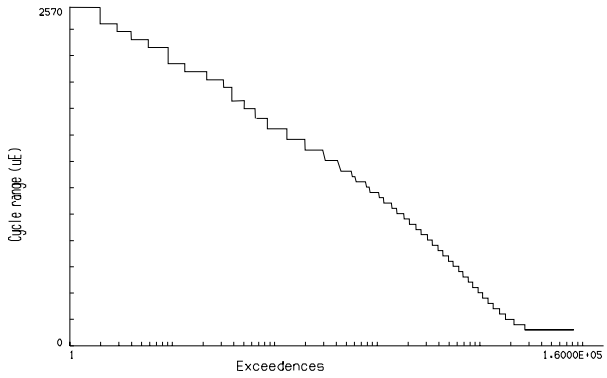


Figure 4.16 Cycle exceedence diagram

The cycle range and mean may also be used to form a range-mean histogram of cycle distribution.

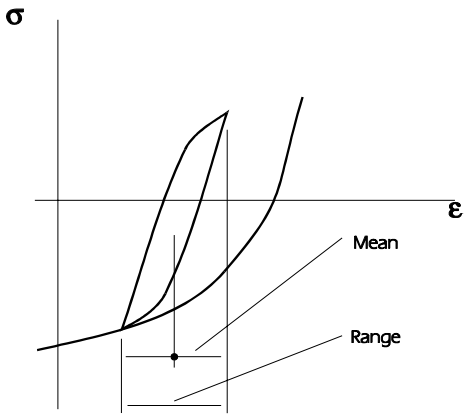


Figure 4.17 Definition of cycle range and mean

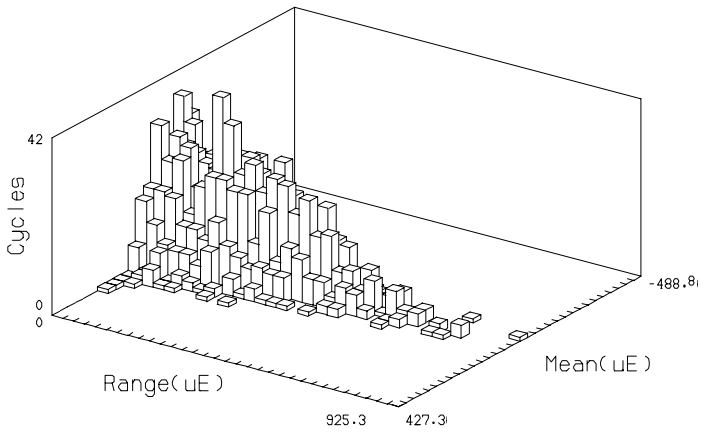


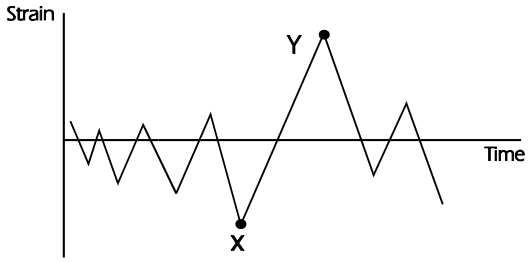
Figure 4.18 Cycle range-mean histogram

The simple cycle-counting algorithm given earlier works only if the analysis can start at the absolute maximum (numerically greatest) point in the signal. It therefore requires that the complete signal is available for searching to determine the position of the absolute maximum data point.

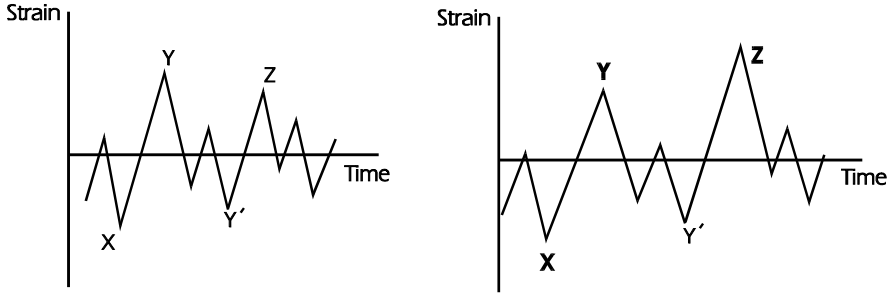
The analysis starts at this point, continues to the end of the signal, returns to the first point in the signal, and finishes at the absolute maximum again. (The largest cycle is formed by the signal maximum and minimum). This type of cycle counting is referred to as off-line analysis, and so the

algorithm given in *Figure 4.6* is an off-line Rainflow algorithm, because the complete signal must be captured before analysis can be attempted.

Socie and Downing (Ref. 4.1) published a modified algorithm which can be used in on-line signal processing. If the first few data points in a signal are as shown below, then a number of cycles can be closed.



The two points X and Y, however, cannot be analysed, because the rest of the signal is not yet known. Consider the following possibilities.



In the first case, X and Y are the maximum and minimum values in the complete signal, and form the largest cycle. Y' and Z form a smaller cycle. In the second case, X is the minimum value but a larger maximum, Z, occurs after Y, so X and Z form a cycle, and Y and Y' form a smaller cycle. The simple off-line algorithm must be modified so that decisions on some cycles are postponed until the end of the signal. At the end of the signal, the data buffer of unclosed cycles takes the form shown in *Figure 4.19*. It can be analysed using the simple algorithm to close the remaining cycles. The detailed rules for this method are given in Ref. 4.1.

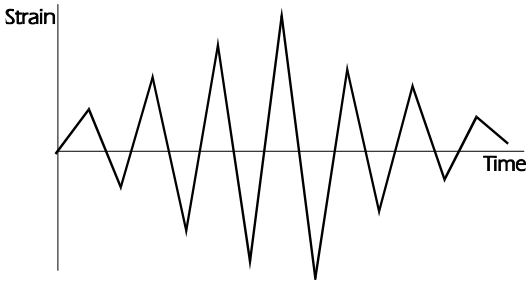


Figure 4.19 Buffer of unclosed cycles

Real signals contain large numbers of very small cycles, some of which may be a result of digitisation noise. These cycles will slow down the analysis, and distort the scaling of graphical displays. It is usual to set a hysteresis gate level which excludes very small cycles from the analysis. A gate of (say) 1% of the maximum signal range is often an appropriate value.

Before the adoption of Rainflow cycle counting, other methods of signal analysis were developed for life estimation or for comparing the severity of different load spectra. These are described in the following sections.

4.3 Level crossing analysis

Level crossing analysis provides an indication of the amplitudes present in a signal. Perhaps its greatest strength is that the results can be post-processed to provide information on the peaks and valleys in the signal. It was one of the earliest techniques to be provided in mechanical signal analysers. The original 'g' meter for measuring aircraft load spectra was a level crossing device.

The amplitude of the signal is divided into a number of levels, or bins. The boundary between adjacent bins is a threshold. A count is made of the number of times the signal crosses each threshold.

In DIN 45667, level crossing analysis is defined as

'a count of all positive slope threshold crossings when the signal is greater than zero, and a count of all negative slope threshold crossings when the signal is less then zero'.

Most analogue to digital converters convert a signal into positive integers, so the algorithm can be implemented simply as a count of all positive slope threshold crossings.

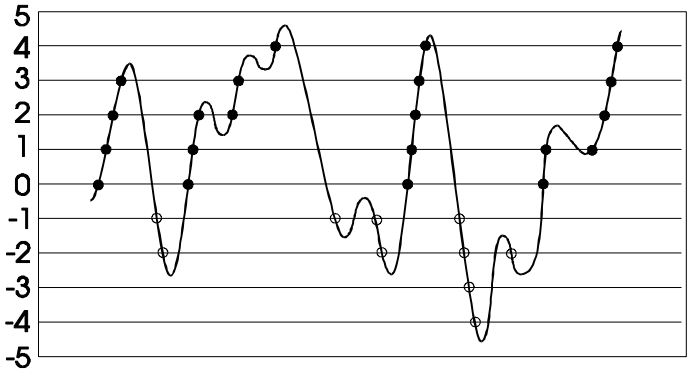


Figure 4.20 Level crossing counting

Small fluctuations in a signal could cause a large number of level crossings if they coincided with a threshold, but not otherwise. This anomaly can be avoided by specifying a reset level, or gate level. A positive slope threshold crossing is then counted only if the signal has also crossed at least one lower threshold with a negative slope *(Figure 4.21)*.

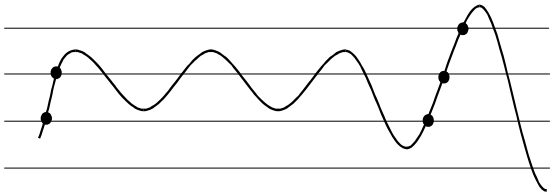


Figure 4.21 Reset level

Level crossing analysis produces a matrix of threshold counts, for example

LEVEL	NUMBER OF LEVEL CROSSINGS
10	0
9	1
8	5
7	35
6	100
5	300
4	120
3	35
2	5
1	0

The level crossing distribution can be analysed to provide a distribution of peaks and valleys. In the example above, if the signal crosses level 8 five times, and level 9 once, then there must be 4 peaks between levels 8 and 9.

4.4 Peak and valley counting

Peak-valley counting simply counts the number of peaks and valleys which fall within specified bands. A gate level may be set to exclude small fluctuations in the signal. In *Figure 4.22*, a peak and valley which formed a range less than the gate level would not be counted.

Figure 4.22 Peak-valley counting

The results of a peak/valley count may be displayed as an exceedence diagram which shows the number of peaks or valleys which exceed any specified value *(Figure 4.23)*.

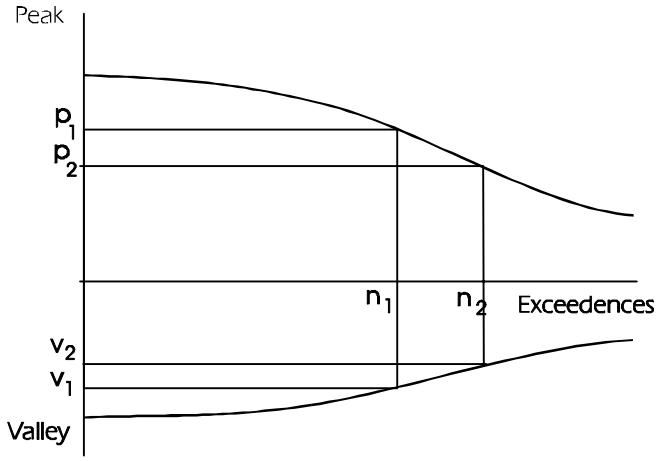


Figure 4.23 Peak-valley exceedence diagram

These exceedence diagrams were once used to compare the severity of different type of service duty, and can be of value in industries where signals are usually of a similar pattern but differ in scaling. A Rainflow cycle distribution will usually provide a much better comparison.

To obtain approximate cycle information from peak/valley exceedence diagrams, it is sometimes assumed that cycles are formed by peaks and valleys of equal probability. In *Figure 4.23*, the difference in numbers of exceedences n_1 and n_2 for the values of peaks p_1 and p_2 gives the number of peaks between these two values. The two values of the valleys v_1 and v_2 which give the same number of occurrences can then be obtained from the diagram. These peaks and valleys are assumed to form cycles, with ranges which lie between (p_1-v_1) and (p_2-v_2), and where (n_2-n_1) is the number of cycles. The process is repeated for the whole diagram, to give a cycle range spectrum. For the particular case of a narrow band signal the procedure is fairly accurate, but for broad band signals the analysis can give calculated fatigue lives which are conservative by orders of magnitude.

4.5 Range counting

Range counting preceded Rainflow counting as a fatigue analysis algorithm. A range is formed between each peak and valley. The analysis produces a histogram of ranges, which can be post-processed to produce range density diagrams and range exceedence diagrams, as described for Rainflow cycle counting.

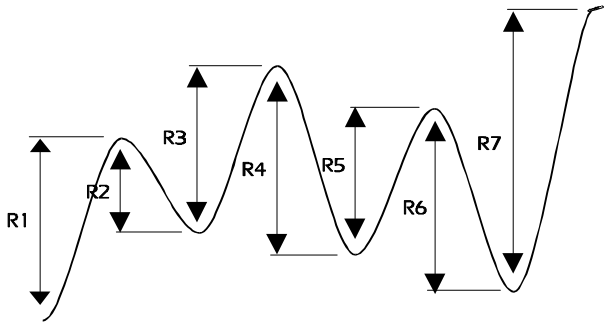


Figure 4.24 Range counting

For fatigue analysis, a range was assumed to be half a cycle, and so fatigue damage could be calculated. Consider however the following section of a signal *(Figure 4.25).*

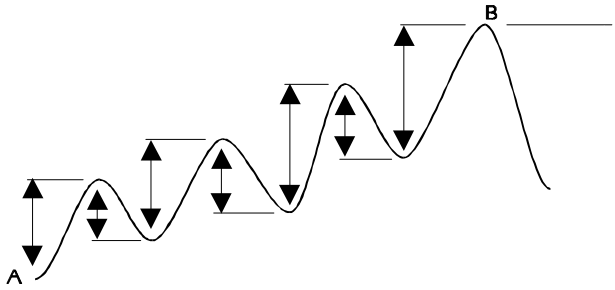


Figure 4.25 Simple range counting -effect on large ranges

The large range from A to B is broken by range counting into a larger number of smaller ranges. In terms of fatigue, this means that the analysis misses the larger cycles, and replaces them by a larger number of smaller cycles. A gate level may be set to exclude the smaller ranges, and so form a result closer to real Rainflow cycle counting.

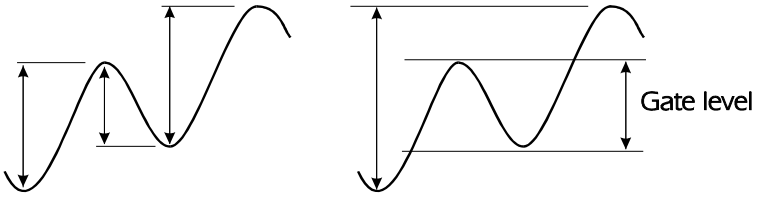


Figure 4.26 Effect of gate level on ranges

However, if the gate is made too small, it will not be effective, and the larger ranges will be broken into smaller ranges. If the gate is too large, larger ranges which may cause fatigue damage will be excluded. Watson and Dabel (Ref. 4.3) showed that the fatigue damage calculated from range counting varied with gate level, and that the optimum value of gate level varied from signal to signal. They also showed that range counting always produced a lower calculated damage than Rainflow counting.

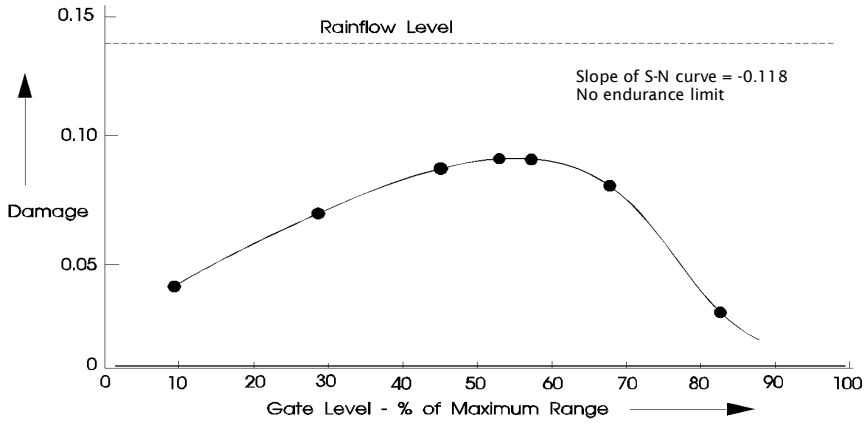


Figure 4.27 Fatigue damage calculated by range counting (from ref. 4.3)

4.6 Summary

It is now generally accepted that Rainflow cycle counting is the most appropriate method of signal analysis for interpreting fatigue lives. Cycle range histograms, cycle density diagrams and range-mean histograms provide valid methods of comparing signals. Chapters 2 and 5 show how Rainflow cycle histograms can be post-processed to provide fatigue life estimates.

The older methods such as level crossing and peak/valley matrices are much less valid than Rainflow counting for fatigue life calculation, and for 'eyeball' comparison of signals for fatigue purposes.

4.7 References

4.1 Downing S D, Socie D F

Simplified Rainflow Cycle Counting Algorithms

Int J Fatigue, Vol 4, No 1, 1982, pp31-40

4.2 Endo T, et al

Damage Evaluation of Metals for Random or Varying Loading

Proc 1974 Symposium on Mechanical Behaviour of Materials, Vol 1.

The Society of materials Science, Kyoto, Japan, 1974

4.3 Watson P, Dabel B
Cycle Counting Fatigue Damage
Journal of the Society of Environmental Engineers, Vol 15-3, Sept 1976

5 Fatigue analysis using stress-life (S-N) curves

5.1 Introduction

This chapter describes the theories for fatigue analysis using engineering stresses.

The average stress across the section XX of a component is the nominal stress, and does not allow for the local effects of the stress concentration. The nominal stress amplitude S_a is plotted against endurance in cycles N_f. This is an *S-N curve (Figure 5.2)*.

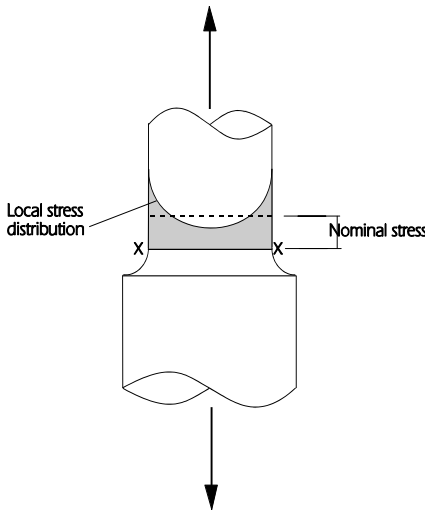


Figure 5.1 Nominal stress in a component

S-N curves are obtained by testing components or smooth specimens of material, under constant amplitude loading, often at zero mean stress (R=-1).

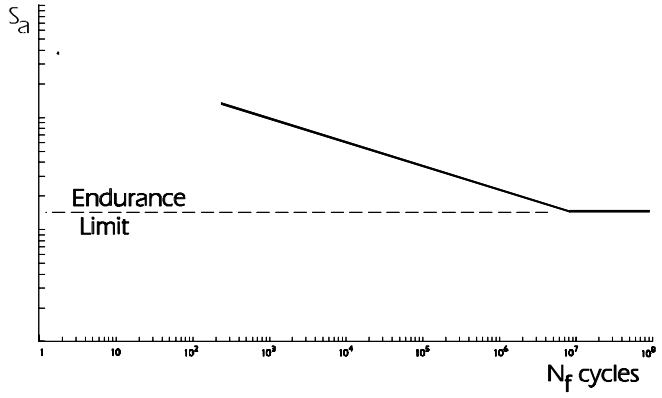
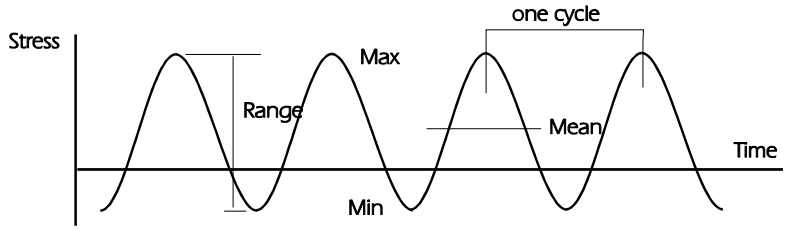


Figure 5.2 S-N curve plotted on log-log axes

The curves are plotted with a $\log_{10}$ (endurance) axis, showing the endurance in cycles N_f. A $\log_{10}$ (stress amplitude) axis is often used, as S-N curves for many materials and components approximate to a straight line on log-log axes at endurances greater than (say) $N_f > 10^3$ cycles.

For tests on smooth specimens, to obtain base-line materials data, N_f cycles is often the endurance to fracture of the specimen, not the life to crack initiation.

The following terms are used to describe constant amplitude loading.



$$\text{Amplitude} = \frac{\text{Range}}{2}$$

Figure 5.3 Terms used to describe constant amplitude loading

maximum stress = S_{max}

minimum stress = S_{min}

stress range = $\Delta S = S_{max} - S_{min}$

stress amplitude = $S_a = \frac{\Delta S}{2}$

mean stress = $S_m = \frac{S_{max} + S_{min}}{2}$

stress ratio = $\frac{S_{min}}{S_{max}} = R$

amplitude ratio = $\frac{S_a}{S_m} = A$

Materials data is often obtained by constant amplitude tests at zero mean stress on smooth specimens of material. In order to apply the data to real components, methods are required for application to other shapes, and to complex stress histories. In addition, it is often desirable to estimate fatigue properties of a material from data on other materials. Following the procedure in Chapter 2, this chapter describes the use of stress concentration factors, methods of mean stress correction, cumulative damage, and methods of approximating the fatigue properties of a material from other test data.

5.2 Application of stress concentrations

Most engineering components contain stress concentrations such as holes, grooves and fillet radii. In many instances fatigue failures occur in these areas. Taking as an example a flat plate containing a circular hole, a simple stress analysis would calculate the stress at the section A-A or, if the reduction in area at the hole was taken into account, the stress over the section B-B *(Figure 5.4)*. The stress distribution at B-B is actually much more complex than the average value calculated by simple stress analysis, and at the edge of the hole the value of the stress will be much higher.

The elastic stress concentration produced by the hole, K_t , is

$$K_t = \frac{\text{actual stress at the hole}}{\text{average stress at section B-B}} = \frac{\sigma}{S_{BB}}$$

with both stresses calculated for elastic conditions.

(An alternative value could be calculated by dividing by the stress at section A-A. Chapter 6 explains stress concentrations more fully).

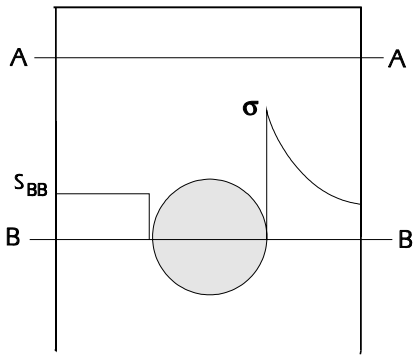


Figure 5.4 Stress concentration at a circular hole

An S-N curve for the plate with the hole could be obtained by testing, and the stress amplitude, S_a, used to plot the data, could be the stress calculated at section B-B in *Figure 5.4*. A comparison with an S-N curve for a smooth specimen is shown in *Figure 5.5*.

At high endurance, the reduction in fatigue strength can be calculated by dividing the smooth specimen fatigue strength by K_t, .

At shorter lives, the stress concentration causes a smaller reduction in fatigue strength. As shown in Chapter 2, this is because yielding at the notch reduces the true stress concentration factor K_σ.

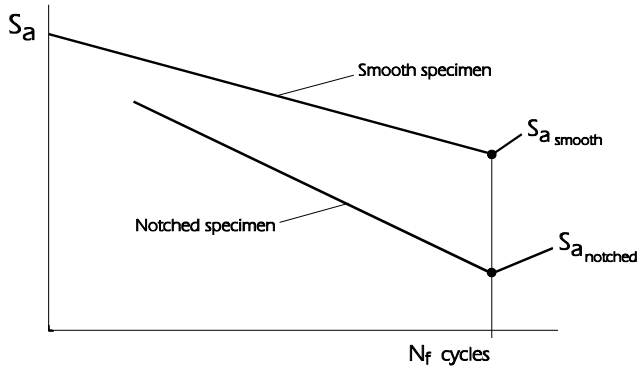


Figure 5.5 S-N curves for a smooth and notched specimen

Various empirical methods have been proposed to allow for the effect of plasticity at the notch. The most simple is to assume the stress concentration has no effect at $N_f = 1$ cycle, with a linear variation between $N_f = 1$ and $N_f = 10^7$ cycles.

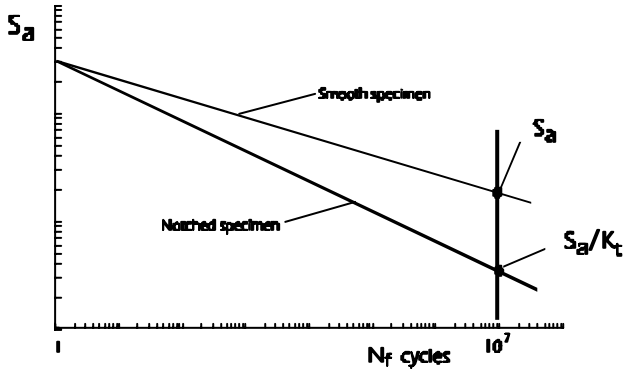


Figure 5.6 Approximate S-N curve for a notched component

An alternative method is to assume that the stress concentration has no effect at an endurance less than $N_f = 10^3$ cycles.

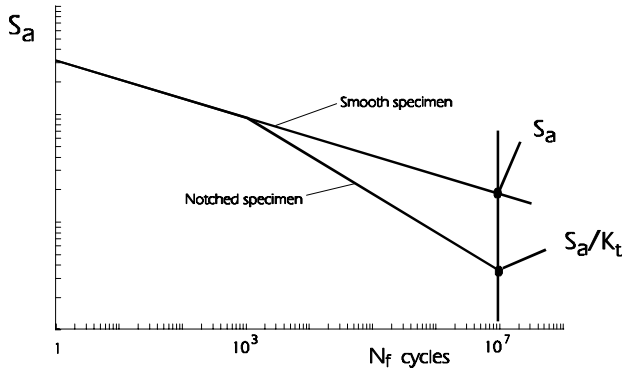


Figure 5.7 An alternative S-N curve for a notched component

Peterson (Ref. 5.1) proposed that the effective stress concentration $(K_t)_N$ at any endurance could be expressed by

$$(K_t)_N = 1 + \frac{K_t - 1}{0.915 + \frac{200}{(\log N)^4}} \qquad (5.1)$$

where K_t is the stress concentration factor at 10^7 cycles. This relationship is used widely in fatigue design, and is plotted in *Figure 5.8* for $K_t = 3$.

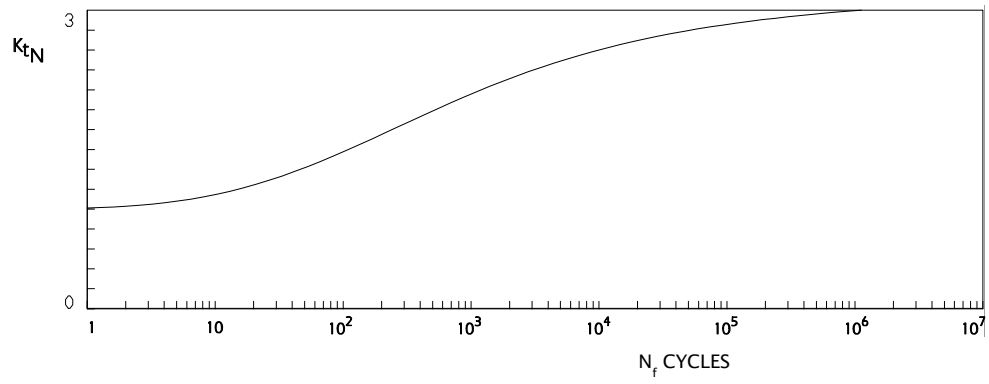


Figure 5.8 Peterson's relationship between endurance and effective stress concentration K_{tN} ($K_t = 3$)

Figure 5.9 shows a S-N curve for a smooth specimen, and the S-N curve calculated from Peterson's equation using K_t = 3.

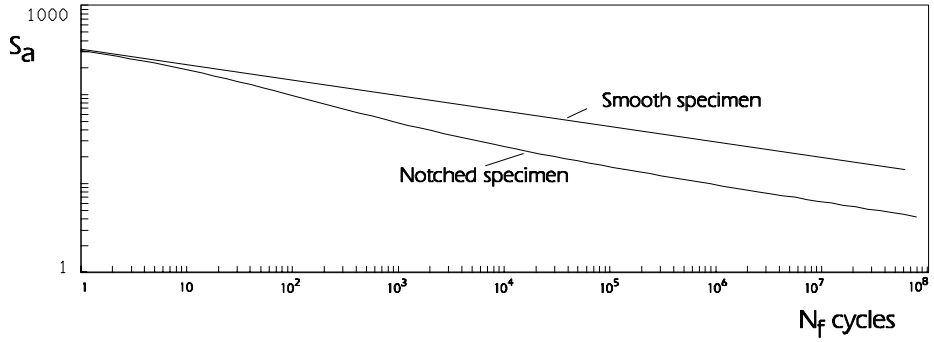


Figure 5.9 An S-N curve for a notched component calculated using Peterson's equation (K_t = 3)

It can be seen that the notched S-N curve has a similar form to that calculated using Neuber's rule, see *Figure 2.44(a).*

5.3 Complex load histories

Real service loading does not usually consist of a constant amplitude sine wave at zero mean stress. *Figure 5.8* shows an example of a strain history measured on the blade of wind turbine. The mean value of many cycles in this signal is clearly not zero, and it is not a constant amplitude signal. To apply an S-N curve to this type of signal, we need

- a method of applying fatigue data obtained at zero mean stress to cycles which are not at zero mean stress.
- a method of applying fatigue data obtained from constant amplitude tests to signals of varying amplitude

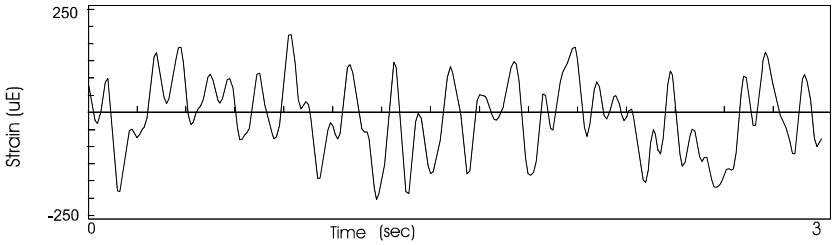


Figure 5.10 Measured strain signal

5.3.1 Effect of Mean Stress

Consider three constant amplitude signals, of the same stress amplitude but different mean stresses

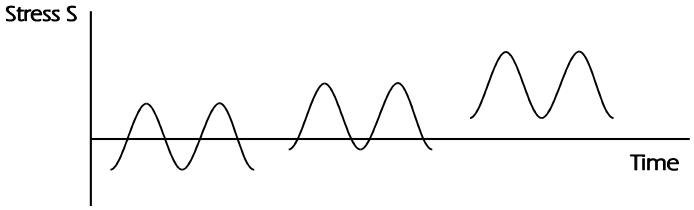


Figure 5.11 Constant amplitude test signals at different mean stresses

The signal with a small tensile mean stress would produce a shorter fatigue life than the signal with zero mean, and the signal with a larger tensile mean stress would produce an even shorter fatigue life *(Figure 5.12)*

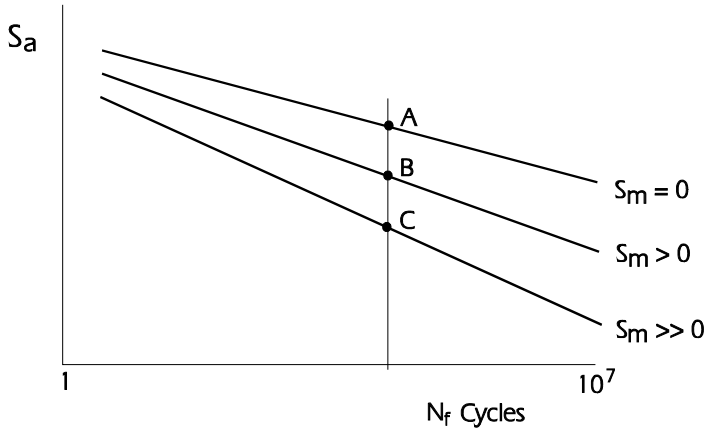


Figure 5.12 Effect of mean stress on endurance

At any endurance, this data can be cross-plotted to show the effect of mean stress on fatigue life at the chosen endurance. This type of plot is called a Haigh diagram *(Figure 5.13)*

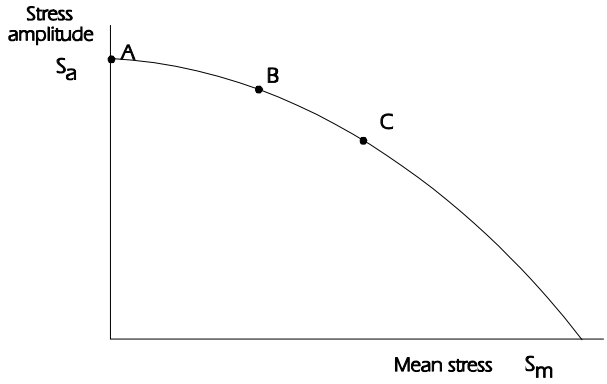


Figure 5.13 Effect of mean stress plotted as a Haigh diagram

To make the axes non-dimensional, the vertical axis can be expressed as a ratio

$$\frac{S_a}{S_{a0}}$$

where S_a is the stress amplitude

and S_{ao} is the stress amplitude at zero mean stress.

This ratio has a value of 1.0 at a mean stress of zero.

The mean stress axis can be made non-dimensional by dividing by the material ultimate tensile strength (UTS).

At a mean stress equal to the material UTS, the allowable stress amplitude is zero, as the material is on the point of fracture. The mean stress axis therefore has a value of 1.0 at an amplitude $S_a = 0$.

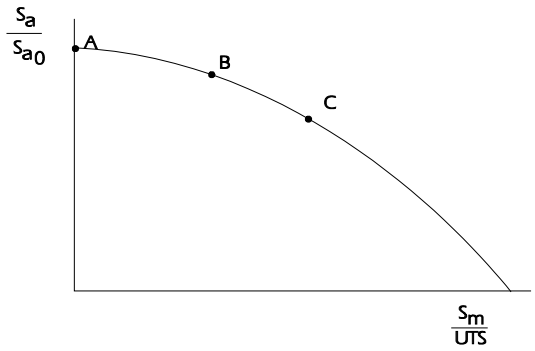


Figure 5.14 Normalised Haigh diagram

Two widely used equations have been used to describe the mean stress relationship shown in *Figure 5.14*.

For smooth polished specimens most test data lies between a linear relationship, proposed by **Goodman**.

$$\frac{S_a}{S_{a0}} + \frac{S_m}{S_{ult}} = 1 \qquad (5.2)$$

and the non-linear **Gerber** relationship

$$\frac{S_a}{S_{a0}} + \left(\frac{S_m}{S_{ult}}\right)^2 = 1 \qquad (5.3)$$

If the yield stress, instead of the UTS, is used to normalise the mean stress axis, the relationship, proposed by **Soderberg** in 1930, is

$$\frac{S_a}{S_{a0}} + \frac{S_m}{S_y} = 1 \qquad (5.4)$$

The Soderberg relationship is generally considered too conservative. The Goodman and Gerber curves are compared in *Figure 5.15*.

In most practical signals or design cases, mean stresses are small compared with the stress amplitude, and either the Goodman or Gerber relationship can be used.

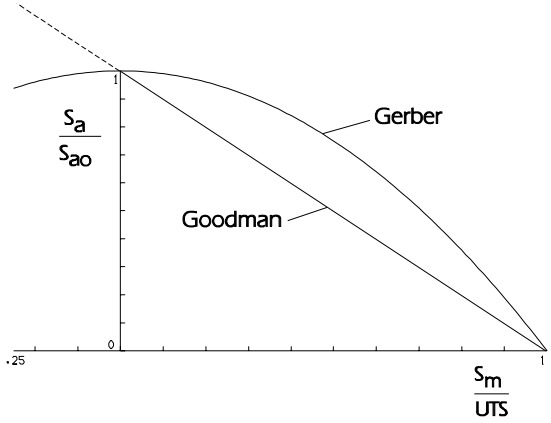


Figure 5.15 Comparison of Goodman and Gerber mean stress corrections

A compressive mean stress will produce an increased fatigue life. The Goodman line may be extended into the compressive mean stress region, and various geometrical methods have been proposed for determining the shape of the Goodman line for compressive mean stresses (See for example Ref. 5.2). Alternatively, the increase in endurance for compressive mean stresses may be ignored.

For infinite life design for very high mean stresses, the **Buch** mean stress correction has been proposed. The Goodman diagram is modified by the addition of two lines linking a point on the stress amplitude axis, at an amplitude equal to the yield stress, and points on the mean stress axis equal to the yield stress. For infinite life, the stress amplitude and mean stress must be within the shaded area. The Buch mean stress correction is more conservative than the Goodman diagram for high tensile or high compressive mean stresses.

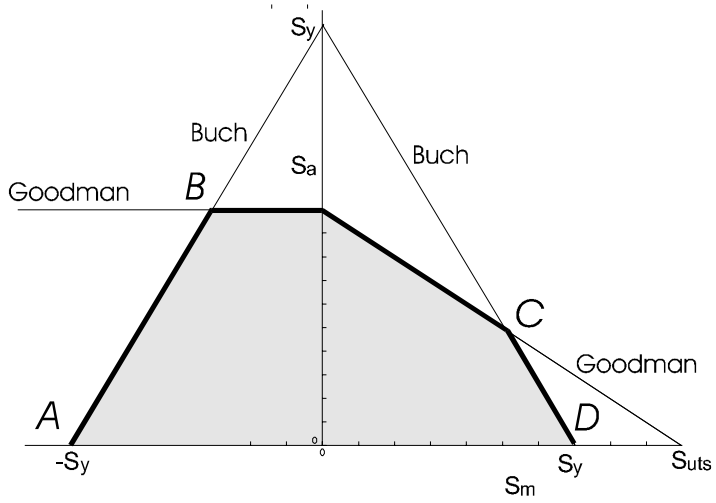


Figure 5.16. Buch mean stress correction

The Buch criterion is defined for the endurance limit. Although the conventional Goodman curve B-C can be defined for other lives, the lines A-B and C-D in *Figure 5.16* are specific to the endurance limit. The Buch criterion cannot be used finite life design.

It is important to remember that these relationships are really applicable only to smooth specimens. A comparison of Haigh diagrams for smooth and notched specimens is shown in *Figure 5.17*. In this case the Goodman or Gerber relationship is followed for small mean stresses, but the relationship becomes invalid when the implied peak stress ($S_m + S_a$) approaches the yield stress of the material

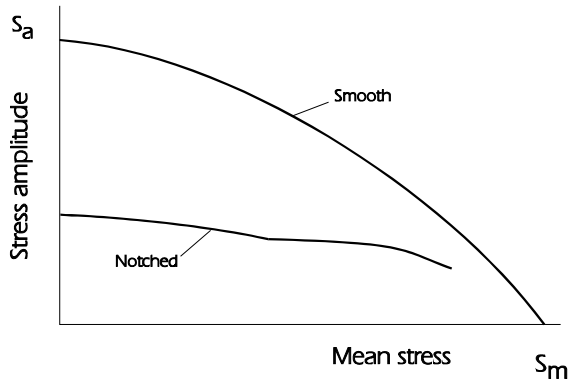


Figure 5.17 Effect of stress concentration on mean stress correction

The Goodman and Gerber mean stress corrections are widely used with S-N curves, even though S-N curves are usually plotted on the basis of nominal stress. It was shown in Chapter 2 that the local mean stresses at a notch could be quite different from the nominal mean stresses because of local plasticity at a notch. This is a fundamental objection to the use of Goodman or Gerber relationships, and limits their applicability to very long-life components where all cycles are essentially elastic.

5.3.2 Analysis Of Variable Amplitude Stress Histories

Constant amplitude sinusoidal stress histories rarely occur in real engineering components. Miner's Rule (see Chapter 1) is used to calculate fatigue damage for variable amplitude stress histories. Rainflow cycle counting (Chapter 4) can be used to obtain the cycles present in a time history of loads, stresses, etc.

Consider a simple stress history consisting of two constant amplitude blocks.

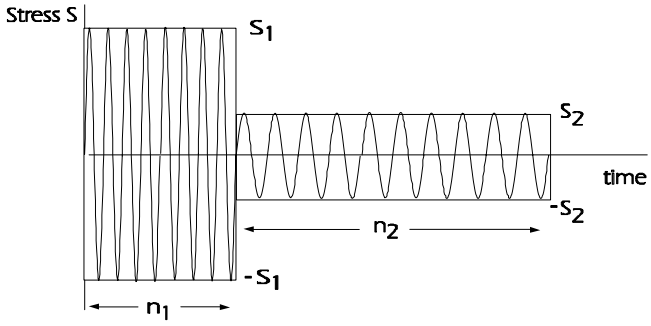


Figure 5.18 Block loading with two stress amplitudes

If the signal consists of only the large amplitude, failure of the specimen would occur when the number of cycles applied (n_1) equalled the number of cycles to failure obtained from the S-N curve for the component (N_1) - ignoring any scatter in test results.

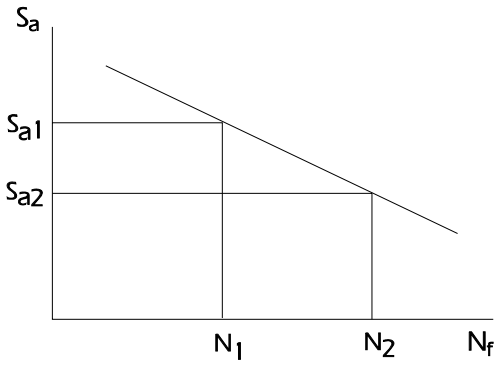


Figure 5.19 Endurance N for the two blocks of cycles

The ratio $\frac{n_1}{N_1}$ has a value of 1.0 at failure.

Similarly, if the signal consists of only the smaller amplitude, then failure would occur when the number of cycles applied (n_2) equalled the number of cycles to failure obtained from the S-N curve (N_2) for this amplitude.

Miner's Rule can be used to calculate the life for the combined signal: if, for either signal on its own, failure occurs when

$$\frac{n}{N} = 1$$

then the combined signal, failure will occur when

$$\frac{n_1}{N_1} + \frac{n_2}{N_2} = 1$$

and for more complex signals, with many blocks of different amplitudes, failure will occur when

$$\Sigma \frac{n}{N} = 1$$

Example: Damage calculation, no mean stress correction

An S-N curve has an intercept S_a = 800 MPa at N_f = 1 cycle, and a slope of –0.086

Its equation is therefore

$S_a = 800 \, (N_f)^{-0.086}$

A Rainflow cycle count of a test signal gave the following spectrum of stress amplitudes

S_a	S_m	n
340	170	5
310	155	31
280	140	49
250	125	74
220	110	101
190	95	258

Calculate the fatigue lives for this spectrum, using no mean stress correction, and using the Goodman and Gerber mean stress corrections.

S_a	S_m	n	N	$\frac{n}{N_f}$
340	170	5	20890	2.393×10^{-04}
310	155	31	61160	5.068×10^{-04}
280	140	49	199600	2.453×10^{-04}
250	125	74	746100	9.918×10^{-05}
220	110	101	3299000	3.062×10^{-05}
190	95	258	1.814×10^{7}	1.422×10^{-05}
				---
			damage =	1.135×10^{-03}

Calculated life $= \frac{1}{1.135 \times 10^{-03}} = 881$ repeats of this spectrum

Example: Damage calculation, Goodman mean stress correction

The Goodman mean stress correction is

$$\frac{S_a}{S_{ao}} + \frac{S_m}{UTS} = 1$$

so that the effective stress amplitude at zero mean, $S_{ao} = \frac{S_a}{1 - \frac{S_m}{UTS}}$

S_a	S_{ao}	S_m	n	N	$\frac{n}{N_f}$
340	432	170	5	1299	3.849×10^{-03}
310	384	155	31	4999	6.201×10^{-03}
280	339	140	49	21330	2.297×10^{-03}
250	296	125	74	103470	7.152×10^{-04}
220	255	110	101	590700	1.710×10^{-04}
190	215	95	258	4171000	6.185×10^{-05}
					---
				damage =	1.329×10^{-02}

Calculated life $= \frac{1}{1.329 \times 10^{-02}} = 75.2$ repeats of this spectrum

Example: Damage calculation, Gerber mean stress correction

The Gerber mean stress correction is

$$\frac{S_a}{S_{ao}} + \left(\frac{S_m}{UTS}\right)^2 = 1$$

so that the effective stress amplitude at zero mean, $S_{ao} = \dfrac{S_a}{1 - \left(\dfrac{S_m}{UTS}\right)^2}$

S_a	S_{ao}	S_m	n	N	$\frac{n}{N_f}$
340	352	170	5	12210	4.095×10^{-04}
310	322	155	31	39200	7.908×10^{-04}
280	289	140	49	139100	3.522×10^{-04}
250	256	125	74	559700	1.322×10^{-04}
220	224	110	101	2642000	3.823×10^{-05}
190	193	95	258	1.538×10^{7}	1.677×10^{-05}
				damage =	1.740×10^{-03}

Calculated life $= \dfrac{1}{1.740 \times 10^{-03}} = 574$ repeats of this spectrum

5.4 Calculation of fatigue safety factors

Mean stress diagrams – Goodman, Gerber, Soderberg, Buch and others, have been used to calculate stress-based factors of strength for a specified fatigue life. Using the Goodman diagram as an example, the method is as follows.

The most severe cycle, i.e. the one that comes closest to the Goodman line, is plotted on the Goodman diagram. A line is drawn through this point from the origin. This indicates how much the stress could be increased before it touches the Goodman line. The ratio A/B (shown below) indicates the factor of strength

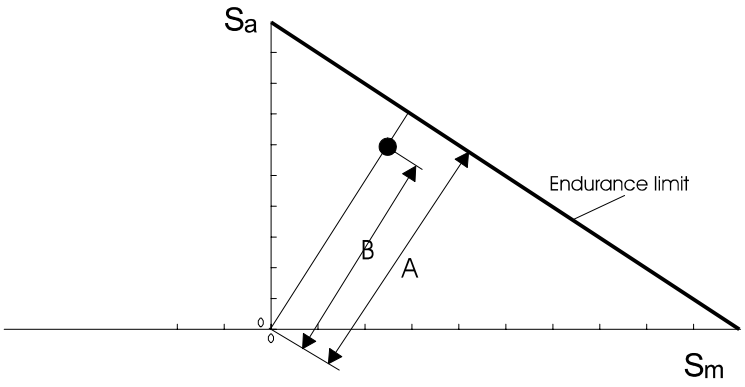


Figure 5.20 Factor of strength (FOS) - Goodman diagram.

To determine the allowable stress amplitude for a fixed mean stress, the ratio A/B may be determined from a vertical line *(Figure 5.21)*. A horizontal line may be used to determine the allowable mean stress for a fixed stress amplitude.

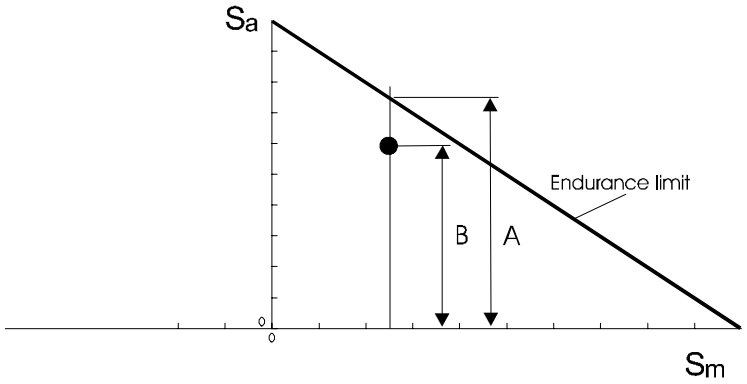


Figure 5.21 Factor of strength on stress amplitude for a fixed mean stress.

This method of determining stress-based safety factors, or allowable stresses, is valid for either

(i) constant amplitude loading at any specified endurance

(ii) more complex stress histories for infinite life design.

In Figure 5.22, the loading consists of 1 occurrence of the largest cycle, and (say) 100 occurrences of the next smallest cycle, shown grey. The design life is (say) 10^5 repeats of the loading.

Under the applied loading, the smaller (grey) cycles would be assumed to be non-damaging. The Goodman analysis would then use the ratio A/B to estimate the factor of strength. However, scaling the applied loading by this value would make the smaller cycles damaging. As there are many more of these, the safety factor would be greatly overestimated, and the analysis would be unsafe.

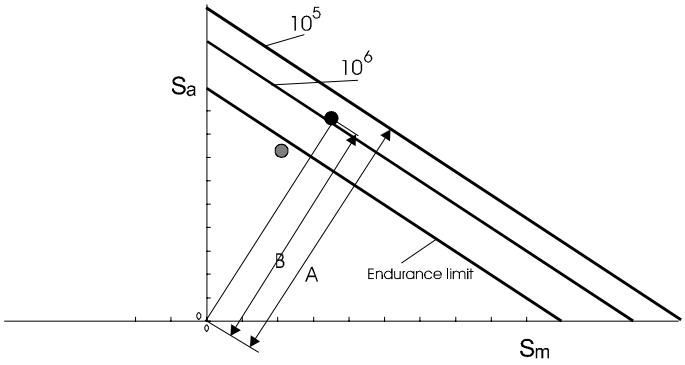


Figure 5.22 Error in calculating stress-based safety factors for complex stress histories.

5.5 Application to other materials

Applying fatigue data from one material to another is still an area of uncertainty and one which can lead to serious errors.

The static strength or ultimate tensile stress (UTS) of a material is one of the most easily measured properties. It was - and is - tempting to express fatigue strength as a proportion of UTS. The S-N curve is then plotted, not as a function of stress amplitude, S_a, but as a function of $\frac{S_a}{UTS}$. The potential danger is illustrated in *Figure 5.23* which shows the relationship between fatigue strength and tensile strength for strain-hardening aluminium alloys. An approximate relationship clearly exists, but the scatter bands are wide, and the relationship becomes non-linear at values of UTS greater than 200 MPa.

For highly notched components there is little relationship between fatigue strength and UTS *(Figure 5.24)*.

Clearly, taking fatigue test data from a lower strength material and extrapolating to a high strength material can lead to very non-conservative estimates of fatigue strength. S-N curves presented on the basis of UTS from reputable organisations will usually provide details of the materials and the range of UTS values covered by the data. The danger is that the S-N curves will be used as general curves and extrapolated outside their range of validity.

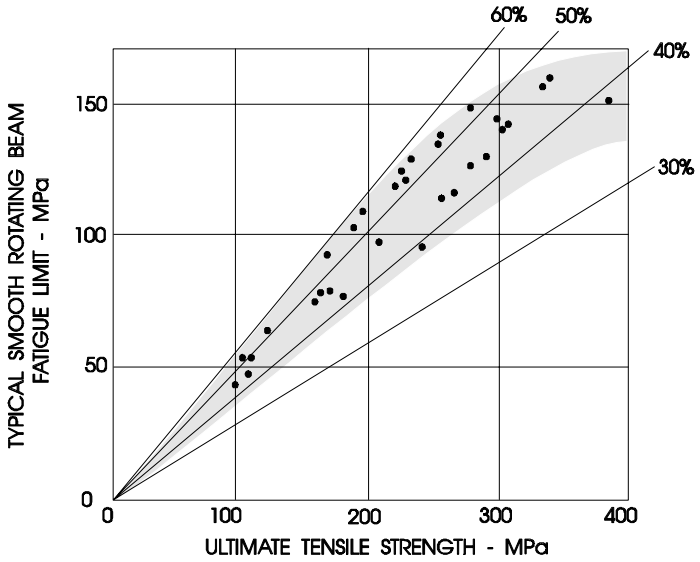


Figure 5.23 Fatigue limit vs UTS for smooth aluminium alloy specimens (from ref. 5.3)

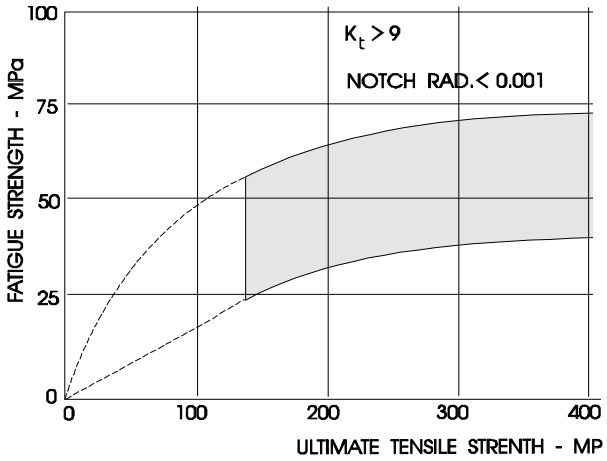


Figure 5.24 Fatigue limit vs UTS for highly notched aluminium alloy specimens (from ref. 5.3)

Some approximate relationships have been proposed for steel smooth polished specimens at R= -1.

1. For steels with a UTS of less than 1000 MPa, the endurance limit stress amplitude will be approximately 50% of the UTS, with most values falling between 30% and 60%. For values of UTS greater than 1000 MPa, the endurance limit stress amplitude will be typically 500 MPa.
2. A typical stress amplitude at 1000 cycles is 90% of the UTS.

However, Devlukia (Ref. 5.4) using data for a range of automotive steels showed that these relationships are generally non-conservative.

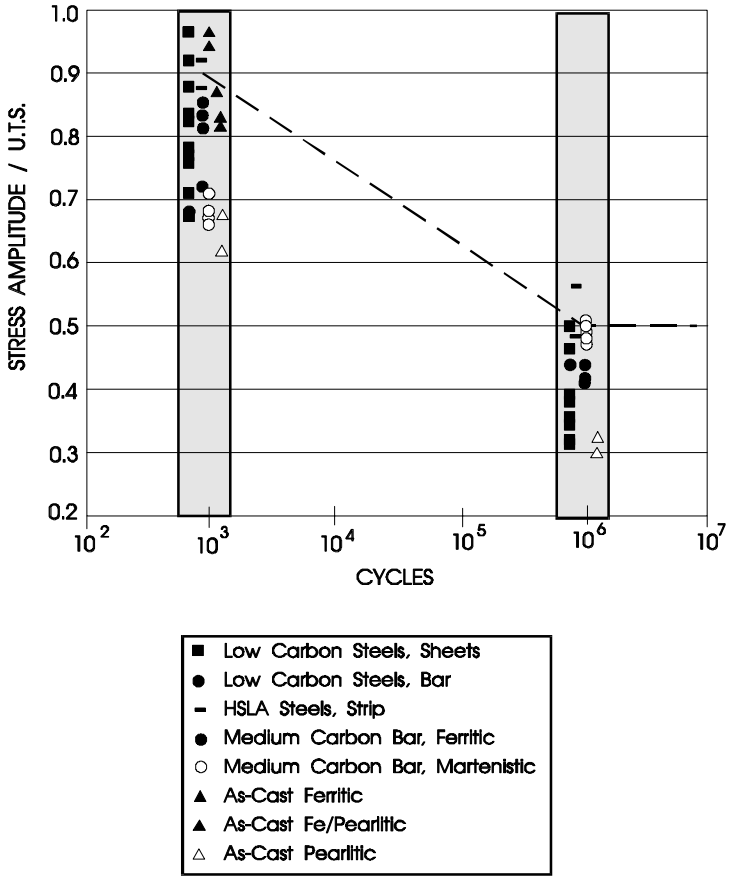


Figure 5.24 Relationship between fatigue strength and tensile strength for automotive steels (from ref. 5.4)

5.6 Applicability of S-N curves

This chapter defined three desirable generalisations - to apply fatigue data to

other shapes of component

more complex load histories

other materials

S-N curves for smooth specimens can be applied to components which contain geometric changes of shape, by using the elastic stress concentration factor. The method works best at long endurances, where there is little plasticity at critical locations.

Allowance can be made for the effect of mean stress by using a Goodman, Gerber or other relationship. The methods work best for small mean stresses, and for smooth specimens. Because local mean stresses can be quite different from nominal mean stresses following yielding at a notch, the methods are not recommended for components containing geometric notches if any cycles in the service history produce yielding at the notch.

These methods do become inaccurate if smooth specimen S-N curves obtained at zero mean stress are applied to notched components at non-zero mean stress, particularly at shorter lives - the effects of plasticity result in too many empirical corrections. Because of these shortcomings, S-N curves have been superseded by local stress-strain methods for the analysis of smooth specimens and components with geometric changes of shape.

However, for complex components, such as structural joints made with welds, bolts and rivets, the S-N curve approach is very powerful.

The reasons are -

The S-N curves are obtained by testing components (not smooth specimens) so they may be used directly or with only minor corrections.

Structural joints have a high effective stress concentration, so the effect of mean stress is small. (Refer to the notched curve in *Figure 5.24*)

The engineering stress is usually calculated assuming elastic conditions, and can therefore be calculated without reference to detailed cyclic material properties.

Within certain limits, S-N curves for steels and steel components can be expressed in a generalised form, although the potential errors are great.

It must be noted that for S-N curves obtained from axially loaded tests on smooth specimens, the stress amplitude does not make any assumption of elasticity. Stresses at lower endurances will not be elastic stresses. Similarly, the stresses used on Goodman and Gerber diagrams are not elastically calculated stresses. Stress results from elastic FEA must not be used with S-N curves for anything other than high cycle fatigue, and must not be used with Goodman/Gerber diagrams if there is the possibility that the combination of stress amplitude and mean stress will produce inelastic stresses. Usually this will limit the use of Goodman/Gerber diagrams to mean stresses close to zero.

5.7 References

5.1 Peterson R E

Analytical Approach To Stress Concentration Effect In Fatigue Of Aircraft Materials.

Westinghouse Research Labs Report.

5.2 E.S.D.U. Data Sheets

Engineering Sciences Data Unit, London

5.3 Fatigue Design Handbook

SAE AE-4 Society of Automotive Engineers Inc 1968, and subsequent issues

5.4 Devlukia J

Fatigue Studies Related to the Automotive Industry

PhD Thesis, Sheffield University, 1994

6 Stress concentrations

6.1 Introduction

In Chapters 2 and 5 it was shown that an elastic stress concentration factor can be used to apply smooth specimen test data to notched components. This Chapter describes stress concentrations in more detail.

Static design methods to prevent ultimate failure may use an average stress over a critical cross-section. Gross yielding is assumed to redistribute the loads over the section until it is uniformly stressed. Under normal service loads gross yielding should never occur, and stresses will be elastic over most of the component. The exceptions are areas of local high stress caused by details such as holes, grooves and fillet radii.

Taking as an example a thin flat plate containing a circular hole (Figure 6.1):

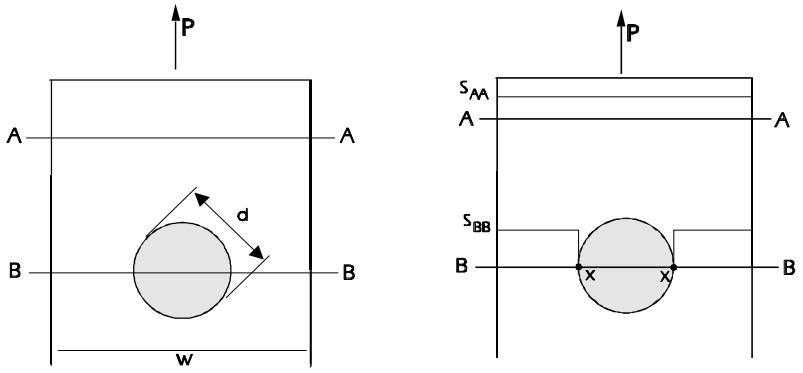


Figure 6.1 Average stress distribution at a circular hole

At section AA undisturbed by the presence of the hole, the elastic stress is uniformly distributed, and is

$$S_{AA} = \frac{P}{A_{AA}} \qquad \text{where } A_{AA} \text{ is the cross-section area at AA}$$

At section BB, the average stress will be higher, because the presence of the hole reduces the cross sectional area, and is

$$S_{BB} = \frac{P}{A_{BB}} \qquad \text{where } A_{BB} \text{ is the cross-section area at BB}$$

The ratio of the areas is

$$\frac{A_{BB}}{A_{AA}} = \frac{w - d}{w}$$

However, the presence of the hole also causes a local disturbance to the stress distribution at section BB, so that the stress is not uniformly distributed across the section. It can be shown mathematically that the maximum stress at the points X, at the edge of the hole, will be 2 to 3 times as high as the average stress across BB (the value depends on the geometry of the plate) and will be distributed as shown in Figure 6.2.

The area under the curves of the average stress distribution and actual stress distribution must be the same, in order to maintain the balance of forces at section BB, so the stress at the edge of the plate will be less than the average value of S_{BB}.

For elastic stresses the ratio of the maximum stress σ to the average stress is called the stress concentration factor, K_t, where the suffix t denotes a theoretical or mathematical value.

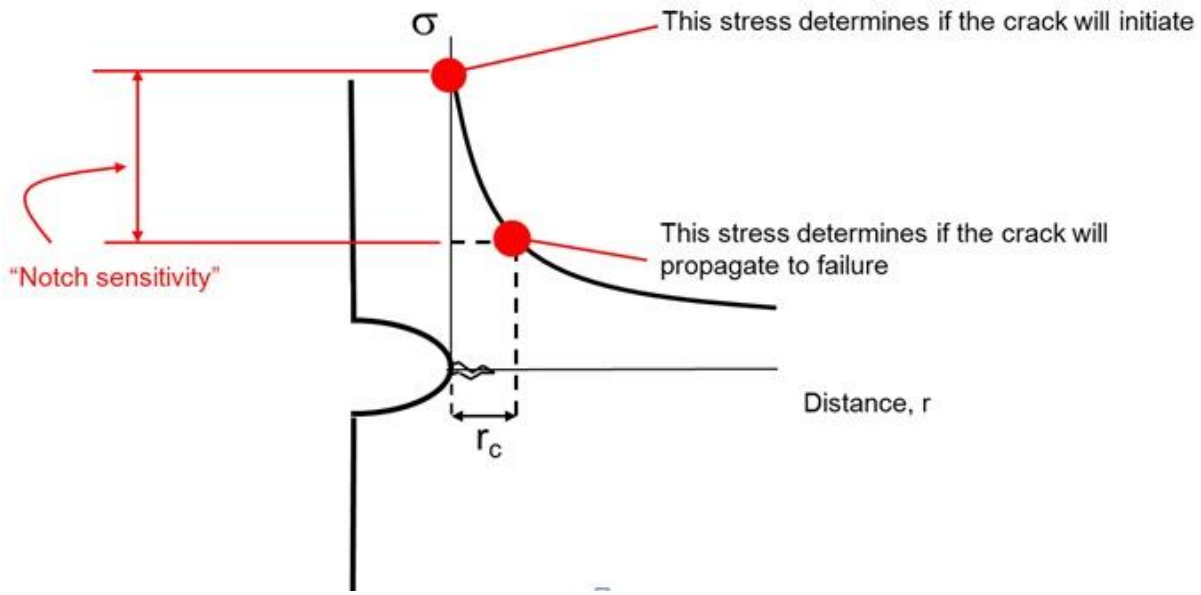


Figure 6.2 Actual stress distribution

Two values of stress concentration factor may be calculated for the example shown in *Figure 6.2*.

Based on the 'gross' stress at AA $K_{t\ gross} = \frac{\sigma}{S_{AA}}$

Based on the 'nett' stress at BB $K_{t\ nett} = \frac{\sigma}{S_{BB}}$

where S_{AA} and S_{BB} are the gross and net section stresses respectively.

These values of stress concentration factor are based on purely geometric effects, and are calculated for *elastic conditions*. The method of calculation may use two-dimensional analysis, with the stresses uniformly distributed through the thickness of the plate. These plane stress conditions apply to thin sections.

Figure 6.3 shows how the value of gross and nett section K_t varies with plate geometry, (from Ref. 6.1).

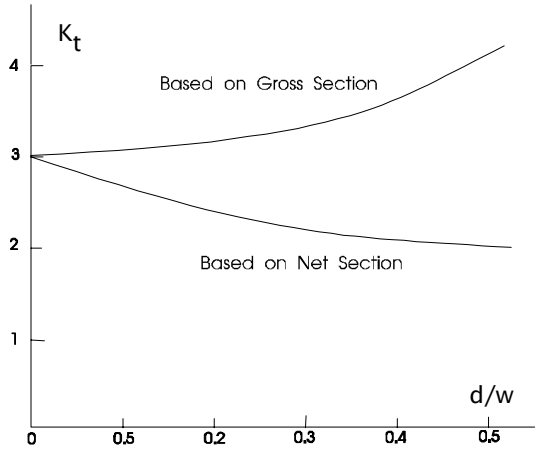


Figure 6.3 Effect of geometry on the value of stress concentration factor for a circular hole in a thin plate

The example of the plate with a circular hole emphasises two important points - that the stress concentration is a very local phenomenon, and that the value of K_t is always related to a stress at some location. There is no such thing as a 'stress concentration factor', but only a 'stress concentration factor with respect to the stress calculated (or measured) at a specified location'.

The values of stress concentration factor given in standard references are usually based on nett section stress. Strain gauge measurement will often be on the gross section some distance from the notch. Many published graphs of stress concentration factors must be corrected for the ratio of gross to nett section stress before they can be used with measured strain gauge signals. Different corrections will apply for direct stress and bending stress, and the proportion of direct and bending stress may be difficult to estimate in practical strain gauge measurement

Figure 6.4 shows a physical explanation for a stress concentration. Under the action of an applied uniaxial stress a circular hole is elongated. At point X, a small strip of material will have its radius of curvature increased, producing an increased tensile stress. At point Y, the radius of curvature is reduced, producing a compressive stress at this location.

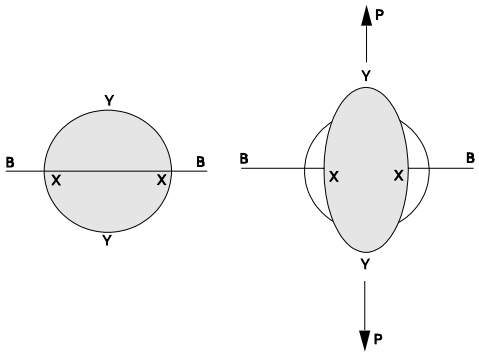


Figure 6.4 Elongation of a circular hole under an applied load

For a hole in an infinite plate under uniaxial loading, the stress concentration factor K_t varies as a continuous function around the hole, with a maximum value of 3 at point X and a minimum value of -1 at point Y.

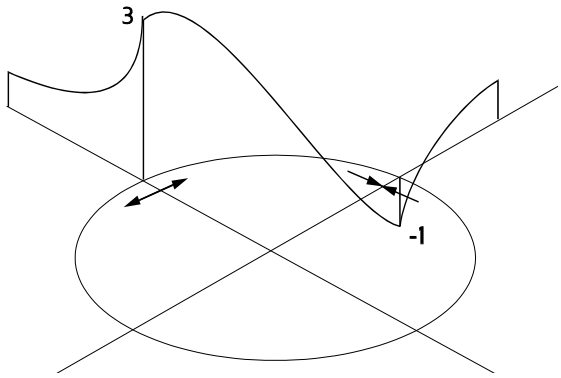


Figure 6.5 Variation of stress concentration around a circular hole

The variation of the stress concentration factor across section XX is shown in *Figure 6.6* (from Ref. 6.4).

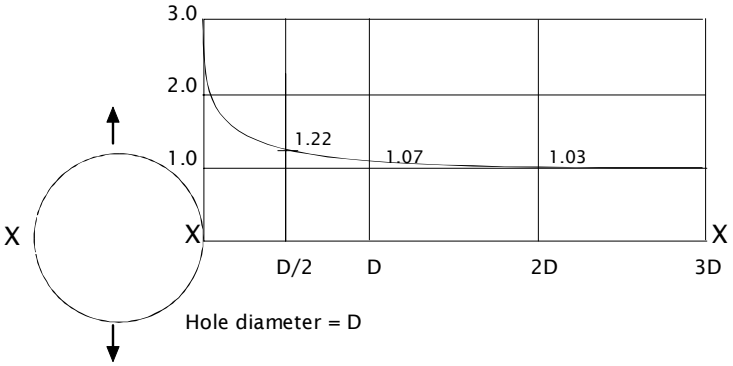
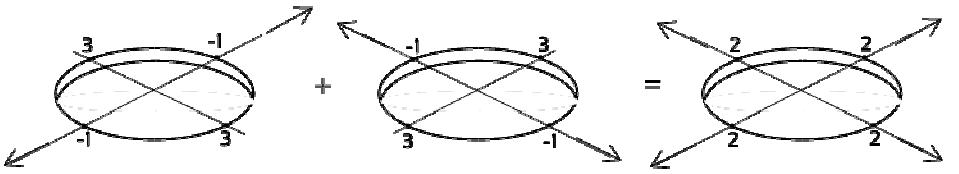


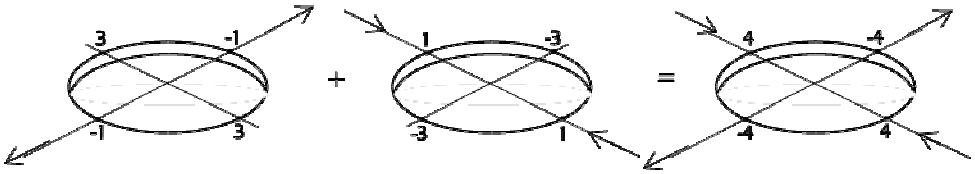
Figure 6.6 Variation of stress concentration factor for a circular hole in an axially loaded plate

From *Figure 6.5*, with the values of K_t = 3 on the axis perpendicular to the applied stress, and K_t = -1 on the axis parallel to the direction of applied stress, other conditions can be derived by superimposition, because the stresses are elastic.

For **equal biaxial tension** K_t has a maximum value of 2. For **equal biaxial tension/compression** K_t has a maximum value of 4.



Equal biaxial tension



Equal biaxial tension/compression

Figure 6.7 Stress concentration factors obtained by superimposition

For holes, grooves and fillet radii in thicker materials, the assumption of plane stress does not apply. For a circular hole in an infinite plate in uniaxial tension, K_t varies through the thickness (Ref. 6.4), with a maximum value of 3.1 at the mid thickness and a value of 2.8 at the plate surface *(Figure 6.8)*. As plane stress conditions do not apply the stress state at the notch will be multiaxial and uniaxial fatigue data cannot be used.

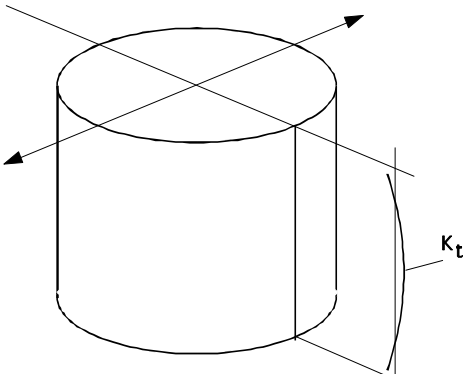


Figure 6.8 Variation of stress concentration through a thick plate

Stress raisers such as holes or grooves may interfere if they are close together, changing both the value of K_t and the position of the maximum value *(Figure 6.9)*

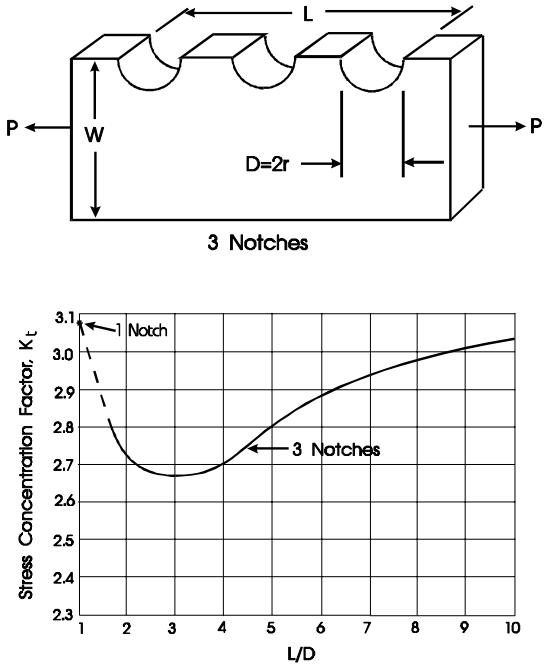


Figure 6.9 Stress concentration factors for adjacent grooves

6.2 Fatigue strength reduction factor

The elastic stress concentration factor K_t is used to calculate the increase in stress at a notch. The *actual* reduction in fatigue strength produced by a stress concentration is usually less than that calculated using K_t,. The actual reduction in fatigue strength is defined by K_f, the *'fatigue strength reduction factor'*. In *Figure 6.10*

if $\frac{\Delta S_{smooth}}{2}$ is the stress amplitude for a smooth specimen at some high endurance, (say) 10^7 cycles,

and $\frac{\Delta S_{notched}}{2}$ is the stress amplitude for a notched specimen at the same endurance,

then K_f is the ratio of the two stress amplitudes, i.e. $K_f = \frac{\Delta S_{smooth}}{2} \Big/ \frac{\Delta S_{notched}}{2}$

K_f represents the real difference in fatigue strength between smooth and notched specimens. K_f is discussed in two sections – local plasticity at the notch, and the notch sensitivity of a material.

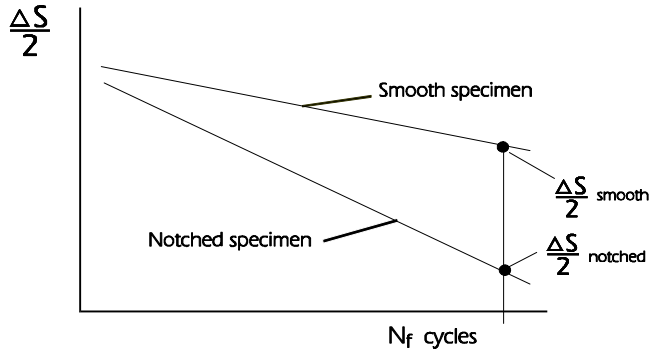


Figure 6.10 Fatigue strength reduction factor

6.2.1 Effect of local plasticity

As values of K_t are calculated for elastic conditions, plasticity at the stress concentration will reduce the apparent value of K_t. This effect is observed during tests to produce S-N curves. *(Figure 6.10)*

An explanation for this apparent reduction in K_t at shorter endurance was explained in terms of local yielding in Chapter 2. When using S-N curves, the methods described in Chapter 5 to construct a notched S-N curve from a smooth-specimen curve can be used, with K_f replacing K_t. For local strain analysis it has been proposed (Ref. 6.6) that K_f can be used instead of K_t for Neuber's rule (Section 2.9.1).

6.2.2 Notch sensitivity

Material and size effects also influence the value of K_f for a given value of K_t. This effect is termed the '*notch sensitivity*' of a material, and is defined as

$$\text{notch sensitivity,} \quad q = \frac{K_f - 1}{K_t - 1}$$

Notch sensitivity is a measure of the extent to which a theoretical stress concentration is realised as a real reduction in fatigue strength. For many steels in general engineering use, q has a value between 0.9 and 1.0, so K_f will be between 90% and 100% of K_t for the types of stress concentrations found in typical engineering components, so using K_t instead of K_f is not excessively conservative.

Several attempts have been made to explain the fact that K_f is generally less than K_t.

As there is a pronounced stress gradient at a notch (see *Figure 6.11*), Neuber (Ref. 6.8) considered the material at the notch to be composed of thin layers, each layer incapable of supporting a stress gradient. The stress gradient through the component would then consist of a number of steps. As the theoretical stress distribution is a smooth curve, and as the total force across the notched section is obtained by integrating the stress distribution, then the maximum stress in the surface layer is less than the value calculated from the smooth curve, and so $K_f < K_t$. Neuber proposed that the thickness of the layers was a material property, and so the difference between K_f and K_t will vary with material. The 'layer' concept also explains why the difference between K_f and K_t increases with increasing K_t. As K_t increases, the stress gradient increases, and if the layer thickness is constant for a given material, the integration of the stresses in each layer implies an increased difference between K_f and K_t.

In *Figure 6.11*, the curve is the theoretical stress distribution, and the steps show the 'actual' Neuber stress distribution. The thickness of the steps has been very much enlarged in the diagram.

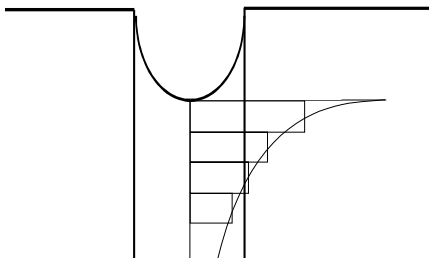


Figure 6.11 The Neuber 'layer' concept

For parallel side grooves, Neuber proposed the relationship

$$K_f = 1 + \frac{K_t - 1}{1 + \sqrt{\frac{\rho n}{r}}}$$

where $2\rho n$ is the layer thickness and r is the radius of the groove.

Kuhn and Hardrath (Ref. 6.10) developed an empirical relationship between ρn and UTS for steels which showed that ρn would be an impossibly large value for steels of low UTS, implying that the 'layer' concept is only a theoretical concept. For aluminium alloys a value of 0.02 has been suggested for ρn.

It has also been proposed that K_f is less than K_t because K_f is the ratio of the un-notched and notched fatigue strengths. If the material contains flaws, the un-notched specimen is already notched. Grey irons contain sharp graphite flakes, and are relatively insensitive to geometric notches. Using the concept of an effective flaw size, Peterson (Ref. 6.9) has proposed a relationship:

$$K_f = 1 + \frac{K_t - 1}{1 + \frac{a}{r}}$$

where a is a material property and r is the notch root radius. Typical values for a are shown in *Figure 6.13*.

Harris (Ref. 6.11) proposed the relationship:

$$K_f = e^{\frac{-r}{\rho H}} + K_t \left(1 - e^{\frac{-r}{\rho H}}\right)$$

where ρH is a material constant, and suggested

$$\rho H = 10^{-2} \left(\frac{84.3}{\text{UTS}}\right)^2 \text{ inches for steels (UTS in ksi)}$$

and $\rho H = 0.004$ inches for aluminium alloys.

Heywood (Ref. 6.14) proposed the relationship

$$K_f = \frac{K_t}{1 + 2\sqrt{\frac{a_H}{r}}}$$

where ah varies with material and with type of notch (hole, groove, etc.). If the notch root radius r is measured in inches, and UTS in ksi, the values proposed for steels are:

$$a_H = \left(\frac{5}{UTS}\right)^2 \quad \text{for a hole}$$

$$a_H = \left(\frac{4}{UTS}\right)^2 \quad \text{for a shoulder}$$

$$a_H = \left(\frac{3}{UTS}\right)^2 \quad \text{for a groove}$$

The actual size of the notch can also effect the value of K_f. As an example, K_t for a circular hole in an infinite plate is 3.0, from mathematical calculation. For steels, this value applies for holes of a diameter normally encountered in engineering components. However, as the radius of the hole becomes very small, the value of K_f will reduce. This has been explained in terms of the volume of stressed material, in the same way as the size effect illustrated in Chapter 1. For a shaft with a semi-circular groove, the stress concentration factor is shown in *Figure 6.12*.

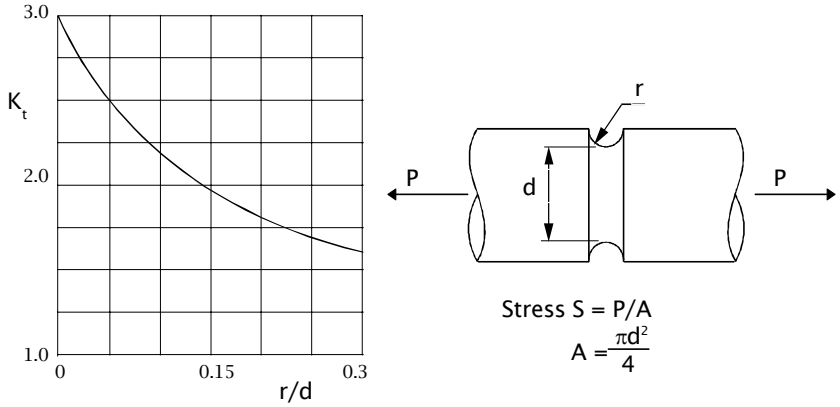


Figure 6.12 Stress concentration factor for a semi-circular groove in a shaft

There is no absolute size given in this diagram, and the data will apply to grooves of 'typical' engineering size (5 - 50mm perhaps). For very small radii the volume of highly stressed material is much less, and the fatigue strength reduction factor K_f will be less than K_t . This is illustrated in *Figure 6.13*.

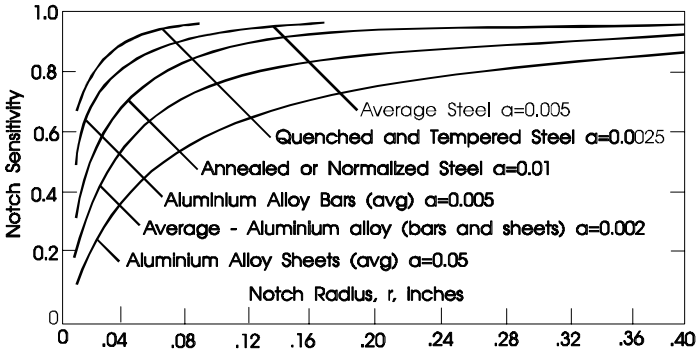


Figure 6.13 Variation in k_f with notch size (from ref. 6.5)

Kuguel (Ref. 6.12) proposed the relationship for the highly stressed volume of material at the root of the notch: $\sigma_{max} = B.V.^{-a}$

B and *a* are material constants and *V* is chosen (arbitrarily) as the volume of material which experiences more than 95% of the maximum notch stress.

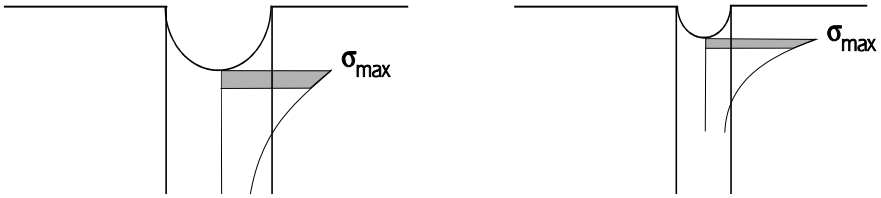


Figure 6.14 Volume of highly stressed material varies with the physical size of the notch

A comparison of experimental values of K_f with predictions from the Neuber, Peterson, Harris, Heywood and Kuguel relationships is given in (Ref. 6.13). The data exhibited considerable scatter. However, for steels, Heywood's relationship was generally the most accurate with an average error of 19%, and half the predictions were non-conservative. For aluminium alloys all the methods gave inaccurate predictions of K_f.

Identical design details under bending and axial loads will give different values of stress concentration factor. An example is shown in *Figure 6.15*.

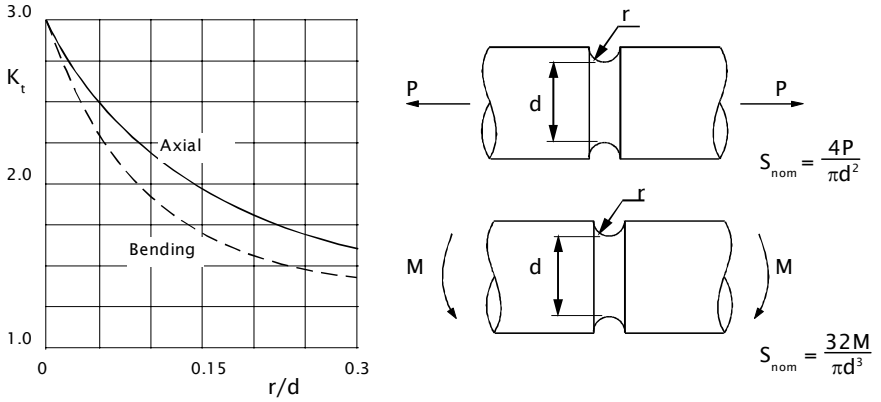


Figure 6.15 Stress concentration for a semi-circular groove under axial and bending loads

Figure 6.16 shows values of the stress concentration factor for a groove with a semi-circular root. This shows very high values of K_t for small root radii. These values will not be achieved in practice, because there is a limit to the maximum stress concentration factor that can be achieved. This may be explained by comparing the effects of a blunt notch and a crack of the same depth.

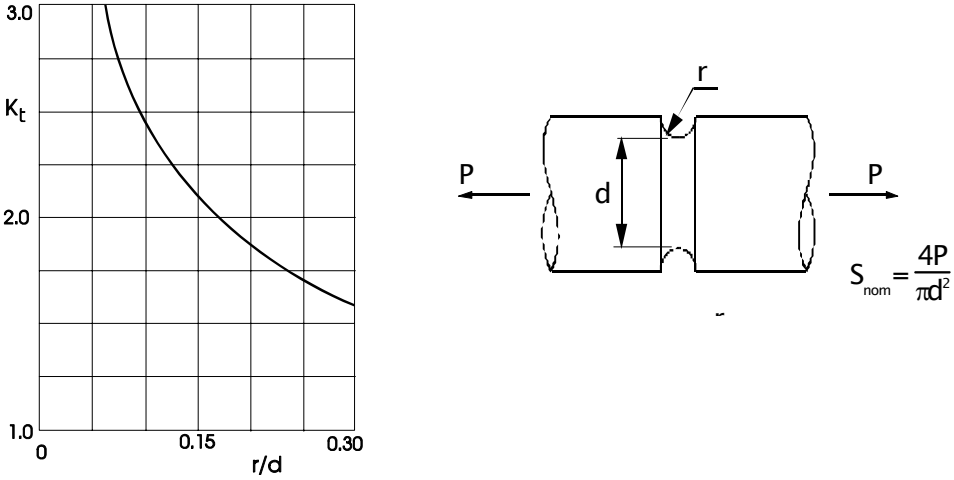


Figure 6.16 Stress concentration factor for a deep groove

The fatigue limit for the notched component is notched fatigue limit $= \frac{\text{smooth specimen fatigue limit}}{K_t}$

Figure 6.17 shows the fatigue limit stress amplitude reducing as K_t increases, i.e., as the notch radius reduces.

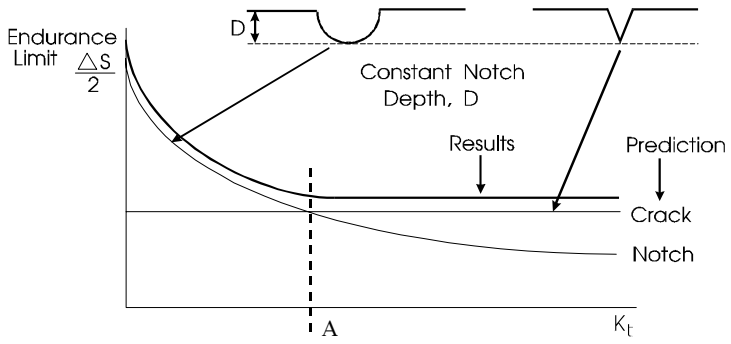


Figure 6.17 Fatigue limit calculated for a blunt notch and a crack of the same depth

As the root radius is made smaller, for the same depth of groove, the shape of the groove approaches a crack. *Figure 6.17* shows the fatigue limit which would be calculated for a crack using fracture mechanics concepts. Crack growth is determined by ΔK, the range of the stress intensity factor at the crack tip. ΔK is calculated from the nominal stress ΔS using

$$\Delta K = \beta \Delta S \sqrt{\pi a}$$

where a is the crack depth and β is a factor to allow for the geometry of the component. (Fracture mechanics is described in Chapter 9). Fatigue cracks will not propagate when ΔK is less than a threshold value ΔK_{th} , and hence below a threshold value of ΔS i.e. ΔS_{th}. ΔS_{th} is therefore the endurance limit stress for the crack. This must be the lowest value of endurance limit stress that can be achieved in practice for this depth of notch.

It can be seen from *Figure 6.17* that the geometric approach gives the endurance limit stress for blunt notches, until the value reduces to the endurance limit stress calculated from the fracture mechanics approach for a sharp crack. The intersection of the 'notch' and 'crack' curves (denoted by A in *Figure 6.17*) gives the maximum value of stress concentration that can be achieved. This value is material dependant.

Figure 6.17 suggests that there is an upper limit to the value of stress concentration that can be achieved for a particular material. Frost (Ref 6.15), and Frost and Dugdale (Ref 6.16), looked at this phenomena for steels and for aluminium alloys. They showed that if the *total fatigue life* is used, there is indeed an upper limit to the value of stress concentration factor. If the life to crack *initiation* is used, there is no such limit. This implies that cracks initiate in response to the surface stresses or strains, with little or no effect of stress gradient. However, in the presence of a sharp notch, the stress gradient may ensure that the crack stops propagating because it grows into a low stressed region. *Figure 6.18* is taken from the work of Frost and Dugdale.

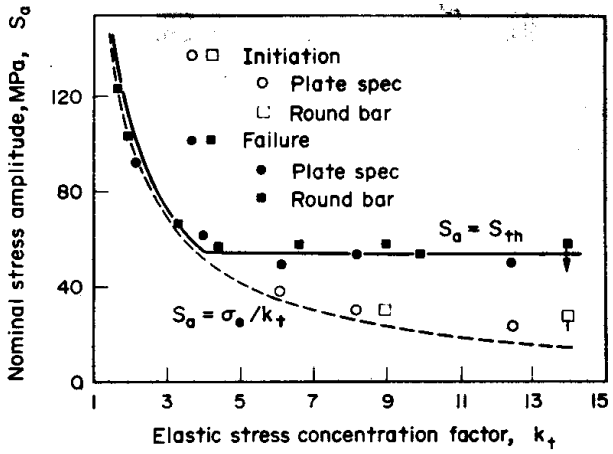


Figure 6.18 Effect of stress concentration factor on life to crack initiation, and on life to failure.

This suggests that, if lives to crack initiation are being calculated, the effect of stress gradient can be ignored.

This approach will give realistic life estimates for most conventional materials. For sharp notches in grey cast iron (fillet radii in crank shafts, for example) the results may be excessively conservative. Grey cast iron is essentially pre-cracked because of the graphite morphology, and infinite life design therefore relies on cracks not propagating.

A diagram such as *Figure 6.17* is required to determine whether a geometric detail behaves as a blunt notch (in which case stress gradient can be ignored) or as a crack (in which case the 'blunt notch' approach will give an excessively conservative estimate of the endurance limit. However, this requires that the stress concentration factor be known. This may be difficult for a complex shaped component, and may not be possible at all when a finite element model is being analysed. 'Critical distance' methods (see for example Taylor (Refs 6.17 – 6.19)) provide a simple but elegant solution to this problem.

6.3 The Critical Distance methods (TCD)

Cracks often initiate from the surface of a component, and the life to crack initiation is determined by the surface stresses and strains. In their simplest form, critical distance methods propose that the crack will propagate if the stress amplitude at some point below the surface is greater than a certain value (*Figure 6.20*). At present critical distance methods are applied to the design of infinite life components. The benefits are clear. If we are prepared to accept cracking, we can base our design (or calculated safety factors) on the subsurface stresses instead of the surface stresses.

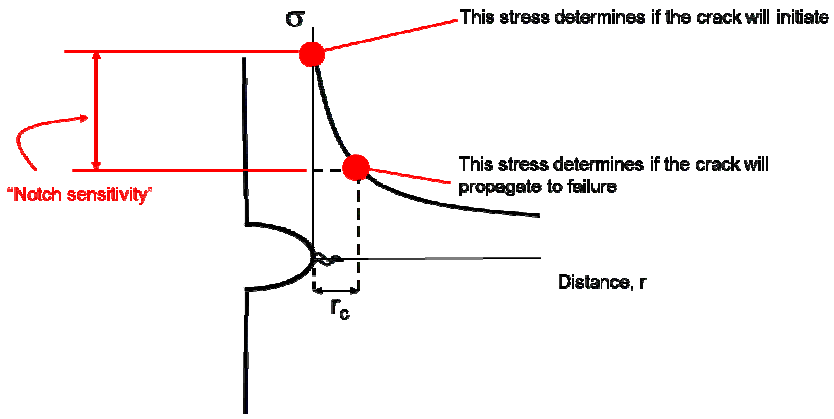


Figure 6.20 The critical distance method

We therefore need to know (i) the distance r_c in *Figure 6.20* and (ii) the allowable stress at this distance.

The distance r_c is a material property. The allowable stress is the endurance limit for the material if a crack is present. Referring to *Figure 6.17*, this is the 'sharp crack' endurance limit stress amplitude.

Using fracture mechanics (see Chapter 10) the stress intensity factor for a crack in a stress field is defined as

$$K = \sigma\sqrt{\pi a}$$

where σ is the applied stress undisturbed by the crack

and a is the crack length

The crack will not propagate if the range $\Delta K = \Delta\sigma\sqrt{\pi a}$ is less than a threshold value ΔK_{TH}.

The distance $r_c = L/2$, where L is the 'critical distance' for the material. L can be obtained from testing, or it can be estimated from

$$L = \frac{1}{\pi}\left(\frac{\Delta K_{th}}{\Delta\sigma_o}\right)^2$$

where $\Delta\sigma_o$ is the stress amplitude at the constant amplitude endurance limit from a conventional smooth-specimen uniaxial stress S-N curve at zero mean stress.

Two critical distance methods are commonly used. The point method (PM) takes the stress at the distance r_c from the surface. The stress is usually obtained from a finite element model (*Figure 6.21*). The line method (LM) averages the stress over a distance (*Figure 6.21*). It may in some circumstances be less sensitive to finite element mesh density than the point method.

Other methods based on stressed area or stressed volume have also been proposed. The additional complexity does not generally lead to increased accuracy. A method based on crack modelling has also been developed but is not described here.

Critical distances can vary from less than 0.1mm for high strength steels, to 4mm for some grey irons. Susmel (Ref 6.20) gives critical distance values for approximately 100 metals.

For fatigue analysis from finite element models, the procedure is

- Calculate the safety factor based on crack initiation using the surface stresses. The safety factor is the allowable stress for infinite life divided by the actual stress.

- Take stress gradient information from the finite element model. For the point method, use the stress at a distance r_c below the surface to calculate if the crack will propagate. If this analysis shows that the crack will not propagate a higher safety factor can be calculated, or the stresses could be increased to produce a more optimised component.

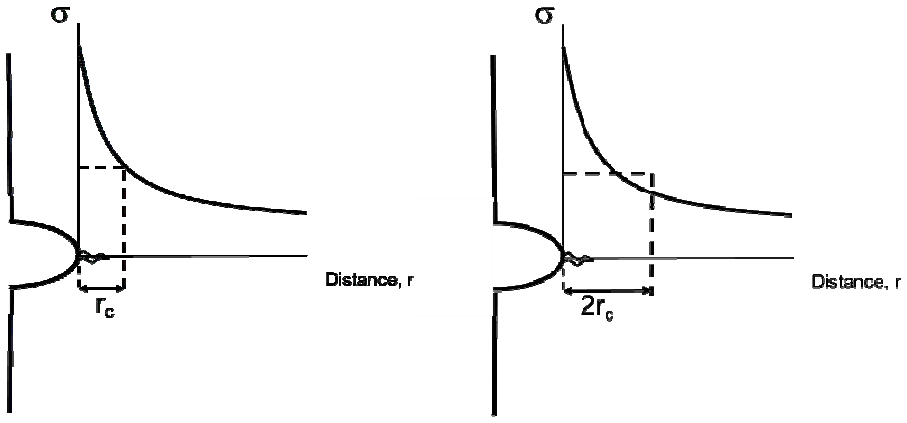


Figure 6.21 The point method (left) and line method (right)

For sharper notches (i.e. at higher values of K_t) there will be a bigger difference between the stresses at the surface and the stresses at the critical distance. This difference will be greater for lower strength materials because the critical distance is greater. Critical distance methods are therefore most applicable to relatively sharp notches in lower strength materials, particularly cast irons.

The benefit of using critical distance methods is that higher stresses may be used. If the crack will not propagate it may be possible to increase the stresses to a value where the crack will just not propagate. However, the designer is then moving from a 'crack initiation' design criterion to one in which cracks are allowed.

Critical distance methods provide an analytical solution to the various problems discussed earlier in this chapter. The difference between axial and bending fatigue strengths is included because the gradients are different and so the stresses at the critical distance are different. The difference between the stresses at the surface and at the critical distance is the materials 'notch sensitivity', and critical distance methods really replace the concept of notch sensitivity factor. The 'size effect' in fatigue is also included in the critical distance methods. In *Figure 6.21*, if the component is scaled to a smaller size the stress distribution will also be scaled, but the critical distance remains the same, so the stresses at the critical distance will be smaller.

Critical distance methods are described in a new book (Ref 6.21). They have been shown to be successful in allowing for both notch sensitivity and size effects in fatigue.

Two important consequences of the critical distance methods are noted in (Ref 6.21). The first concerns methods such as shot peening or cold rolling that improve fatigue strength by introducing residual compressive surface stresses. As the life to failure is determined by the stress some distance below the surface of a notch, it follows that the compressive residual stresses must be present to a depth below the surface at least equal to the distance r_c.

The second consequence concerns re-design of notches to increase the fatigue life by reducing the notch stresses. For design to failure (as opposed to crack initiation), the re-design of the notch must ensure that stresses are reduced at the critical distance below the surface, not just at the surface itself.

6.4 Summary

Elastic stress concentration factors for a wide range of standard cases may be obtained (Refs. 6.1-6.3). With care, very accurate values can be obtained from these references, and if a required detail is not covered, good approximations can often be obtained by interpolating between a more severe and a less severe condition, or by superimposing two or more cases.

Published values of K_t are often based on nett section, and may need to be corrected to gross section values when used with strain gauge measurements.

The actual difference in fatigue strength between smooth and notched components may be less than indicated by K_t . The actual fatigue strength reduction factor is denoted by K_f. The difference between K_t and K_f at low endurance results from plasticity at the notch. Materials also have different values of notch sensitivity. Several attempts have been made to predict values of the notch sensitivity factor q . For steels, the methods give average errors of 20% or more, and for aluminium alloys the methods are inaccurate.

Fatigue cracks initiate in response to surface stresses and strains. The effect of stress gradient need not be included in crack initiation calculations for conventional steels and aluminium alloys. It may be necessary to include the effects of stress gradient/notch sensitivity in the analysis of cast irons, particularly grey iron. Critical distance methods are successful in including stress gradient/notch sensitivity effects when working from finite element models.

6.5 References

6.1. Pilkey, W D

Peterson's Stress Concentration Factors, 2nd ed.

John Wiley and Sons, 1997

6.2 Lipson C and Juvenal R C

Handbook Of Stress And Strength - Design And Material Application

Published by MacMillan, 1963

6.3. E.S.D.U. DATA SHEETS

Engineering Sciences Data Unit, London

6.4 Sternberg E and Sadowsky M A

Three Dimensional Solution For The Stress Concentration Around A Circular Hole In A Plate Of Arbitrary Thickness.

A.S.M.E. J. Appl. Mech March 1949

6.5 Peterson R E
Analytical Approach To Stress Concentration Effect In Fatigue Of Aircraft Materials.
Westinghouse Research Labs Report

6.6 Topper T H, Wetzel R M, Morrow J
Neuber's Rule Applied To Fatigue Of Notched Specimens
J.Mater Vol 4, No 1, 1969, pp200-209

6.7 Fatigue Design Handbook Vol 4
Society of Automotive Engineers.

6.8 Neuber H
Theory of Notch Stresses: Principles for Exact Stress Calculations
Edwards J W, Ann Arbor, Michigan 1946

6.9 Peterson R E
'Notch Sensitivity'
Metal Fatigue Chapter 13, Sines and Waisman, Editors, McGraw-Hill Book Co., In., 1959.

6.10 Kuhn P, Hardrath H F
'An Engineering Method for Estimating Notch-Size Effect in Fatigue Tests on Steel'
National Advisory Committee for Aeronautics, Technical Note 2805, October 1952.

6.11 Harris H J
Metal Fatigue
Pergamon Press, New York, 1961

6.12 Kuguel R
'The Highly Stressed Volume of Material as a Fundamental Parameter in the Fatigue Strength of Metal Members'
University of Illinois, T&AM Report No. 169, June 1960.

6.13 Topper T H, Morrow J (Editors)
Simulation of the Fatigue Behaviour at the Notch Root in Spectrum Loaded Notched Members
University of Illinois, January 1970.

6.14 Heywood R B
Design by Photoelasticity
Chapman and Hall Ltd., London, 1952, p. 348

6.15 Frost N E
Notch Effects And The Critical Alternating Stress Required To Propagate A Crack In An Aluminium Alloy Subject To Fatigue Loading.
Journal of Mechanical Engineering Science 2, 109-119, 1960

6.16 Frost N E and Dugdale D S
Fatigue Tests On Notched Mild Steel Plates With Measurements Of Fatigue Cracks.
Journal of the Mechanics and Physics of Solids 5:182-192, 1957

6.17 Taylor D and O'Donnell M
Notch Geometry Effects in Fatigue : A Conservative Design Approach.
Engineering Failure Analysis, Vol. 1, pp 275-287. 1994

6.18 Taylor D and Lawless S
Prediction Of Fatigue Behaviour In Stress-concentrators Of Arbitrary Geometry.
Engineering Fracture mechanics, Vol. 53, No. 6, pp929-939, 1996

6.19 Taylor D
Crack Modelling : A Technique For The Fatigue Design Of Components
Engineering Failure Analysis, Vol. 3, No. 2, pp129-136, 1996

6.20 Susmel L
Multiaxial Notch Fatigue
Woodhead Publishing, 2009

6.21 Taylor D
The Theory of Critical Distances. A New Perspective in Fracture Mechanics.
Published by Elsevier, 2007.

7 Biaxial fatigue

7.1 Introduction

The fatigue theories introduced in the earlier chapters can be applied only to conditions of uniaxial stress. Many components experience biaxial stresses, often as a result of combined bending and torsion. Powered axles on road and rail vehicles, and engine crankshafts, are obvious examples.

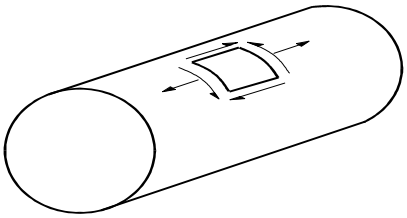


Figure 7.1 Bending and shear stresses on an axle

The use of uniaxial fatigue methods to analyse biaxially stressed components can give very optimistic life estimates. In (Ref. 7.1) a welded steel bracket from a passenger car subjected to multiaxial loading developed fatigue cracks at a life much shorter than that predicted by uniaxial local strain fatigue analysis. The component had also been tested under two different service duties and uniaxial analysis failed to reproduce the relative severity of the two duties.

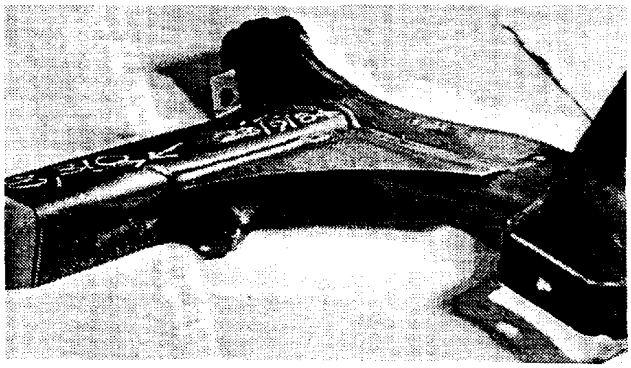


Figure 7.2 Car bracket failure

Figure 7.3 shows a test specimen used in a research program on multiaxial-axial fatigue (Ref. 7.2). The specimen was tested under a typical service loading history.

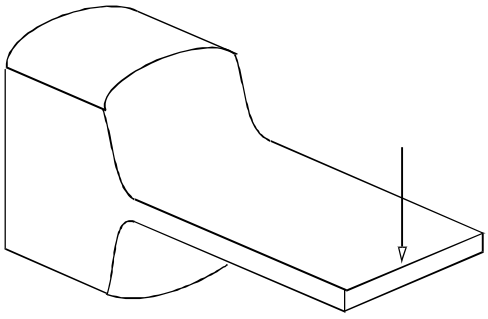


Figure 7.3 A test specimen for multiaxial fatigue

An example result was a test life of 600 repeats of the load history. A life estimation using uniaxial local strain fatigue analysis from a single strain gauge in the fillet radius gave a calculated life of 5000 repeats. The error in calculated life - non-conservative by a factor of 8 - is because biaxial stresses are produced in the fillet radius. Transverse strains across the width of the specimen at the notch cannot be developed because the surrounding lower-stressed material adjacent to the notch provides a restraint. This restraint produces a transverse stress. This is an example a simple uniaxial loading producing a biaxial stress state at a notch.

These two examples show that multiaxial stresses are important in fatigue, and that ignoring them can lead to life estimates which are non-conservative by large factors - factors of 20 or more have been reported (Ref. 7.2).

Multiaxial fatigue is the subject of much current research. However, a consensus on methods of analysis is emerging. Some of the methods are described in this chapter.

Any proposed method of analysing multiaxial fatigue must be consistent with observations of fatigue crack initiation. In brittle metals crack initiation is controlled by the plane which experiences the maximum direct stress or strain. For ductile metals cracks initiation is related to the shear stress or shear strain. Section 7.2 develops the equations for calculating direct and shear stress and strain.

7.2 Stress and strain relationships

7.2.1 Uniaxial and multiaxial stress and strain

The conventional test specimen for materials' testing is shown in *Figure 7.4*.

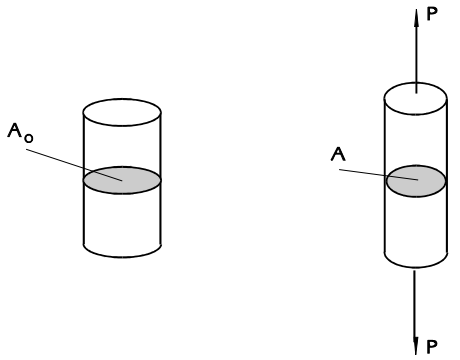


Figure 7.4 Uniaxial fatigue test specimen

If this is re-drawn as a rectangular section *(Figure 7.5)* then it can be seen that as the specimen is extended in the longitudinal direction, it contracts in the two perpendicular directions. The fact that the specimen is free to contract means that no stress is produced on these planes, and so the specimen is in a state of uniaxial stress. There are strains in all three directions, and so the strains are multiaxial.

If the specimen is changed to one which is wide but still thin *(Figure 7.6)* the specimen is not free to contract in width (except at the very edge) because it is restrained by the lower-stressed material at the ends of the specimen, and so a stress is produced in the transverse direction. The specimen is therefore in a state of biaxial stress. As there is no stress in the out-of-plane direction, this condition is one of plane stress. There is an out-of-plane strain, so this is not a plane strain condition.

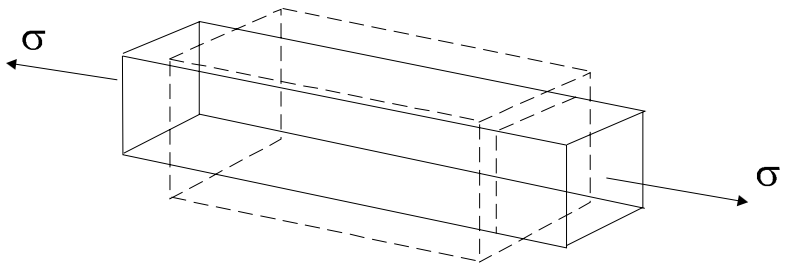


Figure 7.5 Uniaxial stress

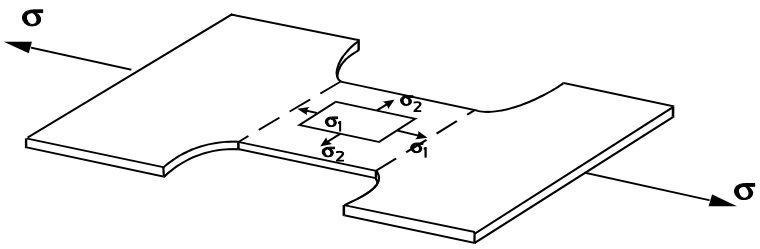


Figure 7.6 Biaxial plane stress

Fatigue cracks usually initiate from the surface of a component, where there is no out-of-plane stress. Much multiaxial fatigue research is therefore the study of fatigue plane stress biaxial stresses.

7.2.2 Equations for plane stress

Plane stress conditions apply on the surface of components. The general state of stress is shown in *Figure 7.7*, with direct stress and shear stresses on the x and y faces.

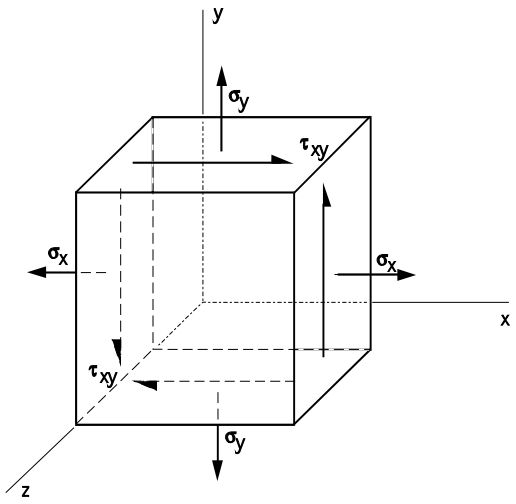
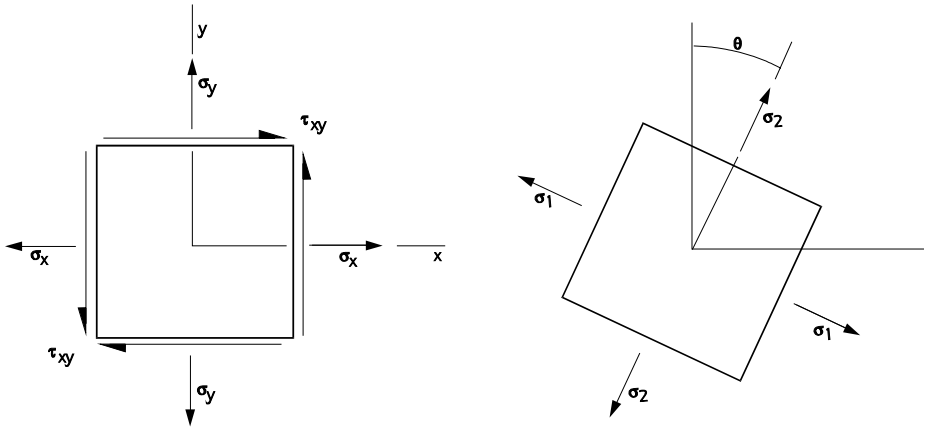


Figure 7.7 General plane stress

As there are no stresses on the third face, the diagram can be simplified to two dimensions.



σ_1, σ_2 are principal stresses

Figure 7.8 Principal stresses

The element can be rotated until the shear stresses fall to zero. The direct stresses are then principal stresses, σ_1 and σ_2.

$$\sigma_{1,2} = \frac{\sigma_x + \sigma_y}{2} \pm \sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + \tau_{xy}^2} \qquad (7.1)$$

As this is a solid 3-dimensional body, there is a third principal stress, σ_3, which acts parallel to the z-axis, and has a value of zero on the surface of a component.

The maximum shear stress is given by

$$\tau_{max} = \sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + \tau_{xy}^2} \qquad (7.2)$$

The maximum and minimum shear stresses occur on planes at 45° to the principal planes, and are equal and opposite in value. From equation 7.1

$$\sigma_1 = \frac{\sigma_x + \sigma_y}{2} + \sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + \tau_{xy}^2}$$

and

$$\sigma_2 = \frac{\sigma_x + \sigma_y}{2} - \sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + \tau_{xy}^2}$$

so that

$$\sigma_1 - \sigma_2 = 2\sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + \tau_{xy}^2}$$

From equation 7.2, the right hand term is equal to $2\tau_{max}$

and so $$\tau_{max} = \frac{\sigma_1 - \sigma_2}{2} \qquad (7.3)$$

These relationships can be illustrated using Mohr's circle *(Figure 7.9)*. Equation 7.1 shows that the principal stresses are given by

a constant $$\frac{\sigma_x + \sigma_y}{2}$$

and a term $$\sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + \tau_{xy}^2}$$

These terms are analogous to the equation for a circle in a plane where the horizontal axis represents the direct stresses and the vertical axis represents the shear stresses. The two terms are the centre of the circle and its radius.

The principal stresses occur on the planes where the shear stresses are zero, and so the values of the principal stresses are σ_1 and σ_2 at the intersections of the circle and the horizontal axis.

The maximum shear stress $$\tau_{max} = \frac{\sigma_1 - \sigma_2}{2} \quad (7.4)$$

occurs on planes at θ=45° to the planes of the maximum principal stresses.

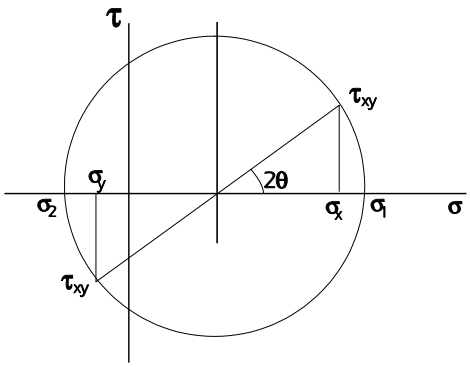


Figure 7.9 Mohr's circle for stress

A direct stress acts in a direction normal to the planes of maximum and minimum shear stress, and is given by

$$\sigma_n = \frac{\sigma_1 + \sigma_2}{2} \quad (7.5)$$

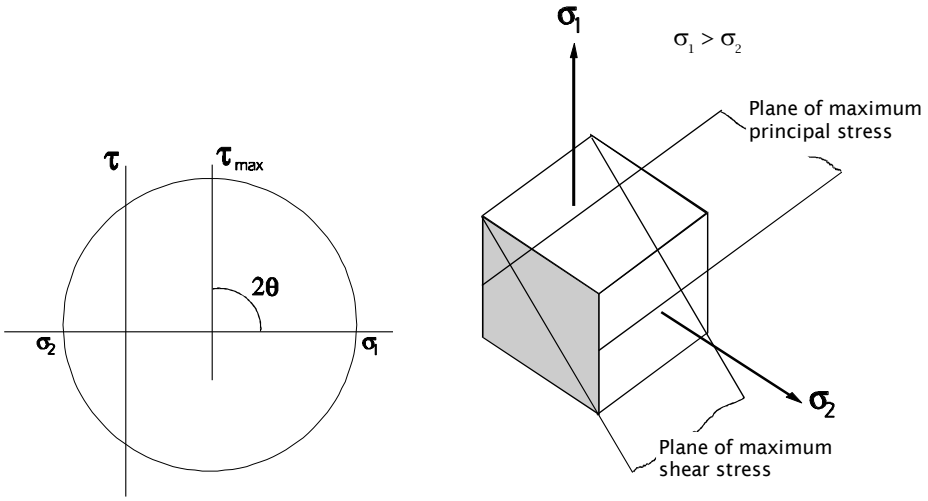


Figure 7.10 Planes of maximum principal stress and maximum shear stress

These equations consider only the stresses in the x-y plane of *Figure 7.7*. For a 3-dimensional element, with σ_z = 0, equation 7.3 can be used to calculate the maximum shear stresses on planes at 45° to each of the principal planes. These are:

$$(\tau_{12})_{max} = \frac{+\sigma_1 - \sigma_2}{2} \qquad (\tau_{13})_{max} = \frac{+\sigma_1}{2} \qquad (\tau_{23})_{max} = \frac{+\sigma_2}{2} \qquad (7.6)$$

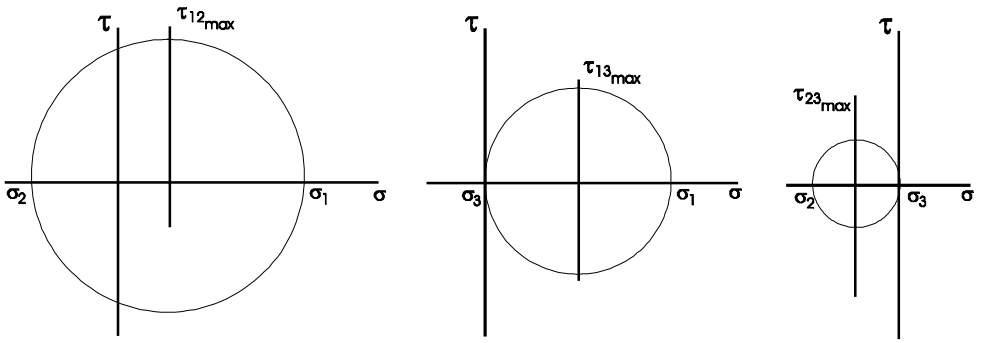


Figure 7.11 Mohr's circles for the three planes

The three Mohr's circles can be combined into a single diagram:

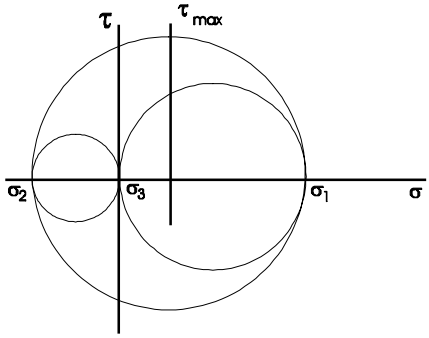


Figure 7.12 Mohr's combined circle for the three planes

7.2.3 Strain equations for plane stress

At the surface of a component there will generally be a strain perpendicular to the surface, so the strain state is one of triaxial strain. For the more simple case of plane strain, i.e. no strain perpendicular to the surface, a two-dimensional diagram can be used to represent the three-dimensional element *(Figure 7.13)*

The element can be rotated so that the shear strains fall to zero, and the direct strains are then the principal strains.

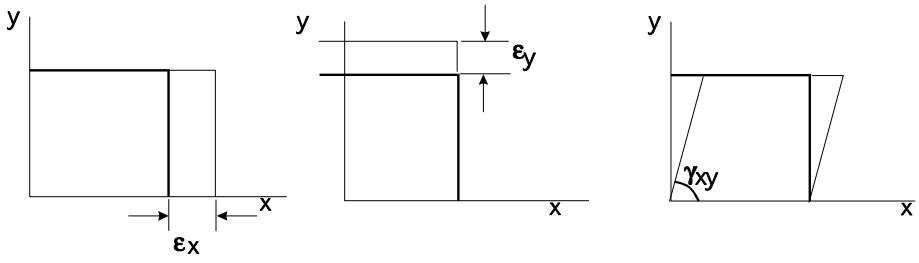


Figure 7.13 Direct and shear strains

For plane strain conditions with strains ε_x , ε_y and γ_{xy} , the principal strains are

$$\varepsilon_{1,2} = \frac{\varepsilon_x + \varepsilon_y}{2} \pm \sqrt{\left(\frac{\varepsilon_x - \varepsilon_y}{2}\right)^2 + \left(\frac{\gamma_{xy}}{2}\right)^2} \qquad (7.7)$$

This equation is similar to equation 7.1 for principal stresses, with the stresses replaced by strains, except that $\frac{\gamma_{xy}}{2}$ replaces the shear stress τ_{xy}.

By comparison with the stress equations, Mohr's circle can be used to describe the strain relationships.

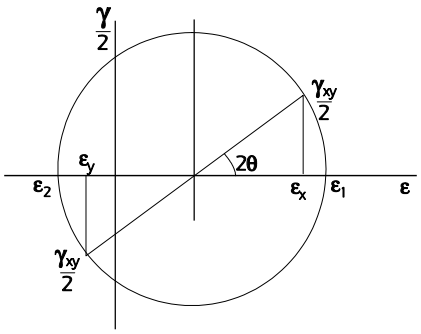


Figure 7.14 Mohr's circle for strain

From *Figure 7.14*, the maximum shear strain occurs on planes at $2\theta = 90^o$, or $\theta = 45^o$, to the planes of the principal strains, and is

$$\frac{\gamma_{max}}{2} = \frac{\varepsilon_1 - \varepsilon_2}{2}$$

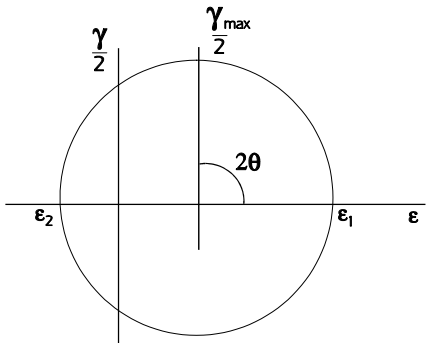


Figure 7.15 Principal strains and maximum shear strain

The strain normal to the plane of maximum shear strain is

$$\varepsilon_n = \frac{\varepsilon_1 + \varepsilon_2}{2} \qquad (7.8)$$

For the general case of strains on all three planes, Mohr's circles can be drawn for each plane:

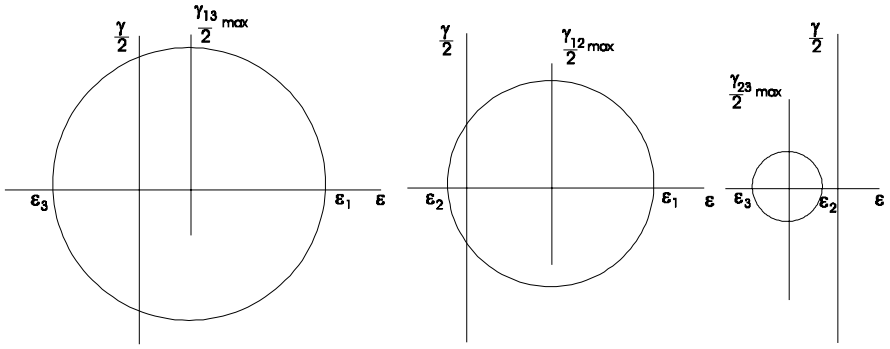


Figure 7.16 Mohr's strain circles for the three planes

The three diagrams can be combined, as shown in *Figure 7.17*.

If $\varepsilon_1 > \varepsilon_2 > \varepsilon_3$, the plane of maximum principal strain is perpendicular to the strain ε_1 *(Figure 7.18)*

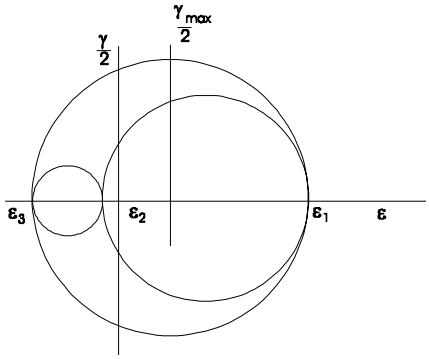


Figure 7.17 Mohr's combined strain circles for the three planes

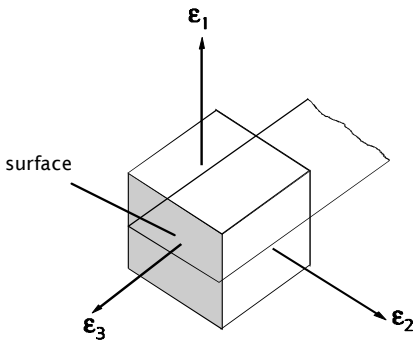


Figure 7.18 Plane of maximum direct strain

The three shear strain planes are shown in *Figure 7.19*. If $\varepsilon_1 > \varepsilon_2 > \varepsilon_3$, the plane in the upper diagram will have the maximum shear strain. Its magnitude is

$$\frac{\gamma_{max}}{2} = \frac{\varepsilon_1 - \varepsilon_3}{2} \qquad (7.9)$$

and the strain normal to this plane is

$$\varepsilon_n = \frac{\varepsilon_1 + \varepsilon_3}{2} \qquad (7.10)$$

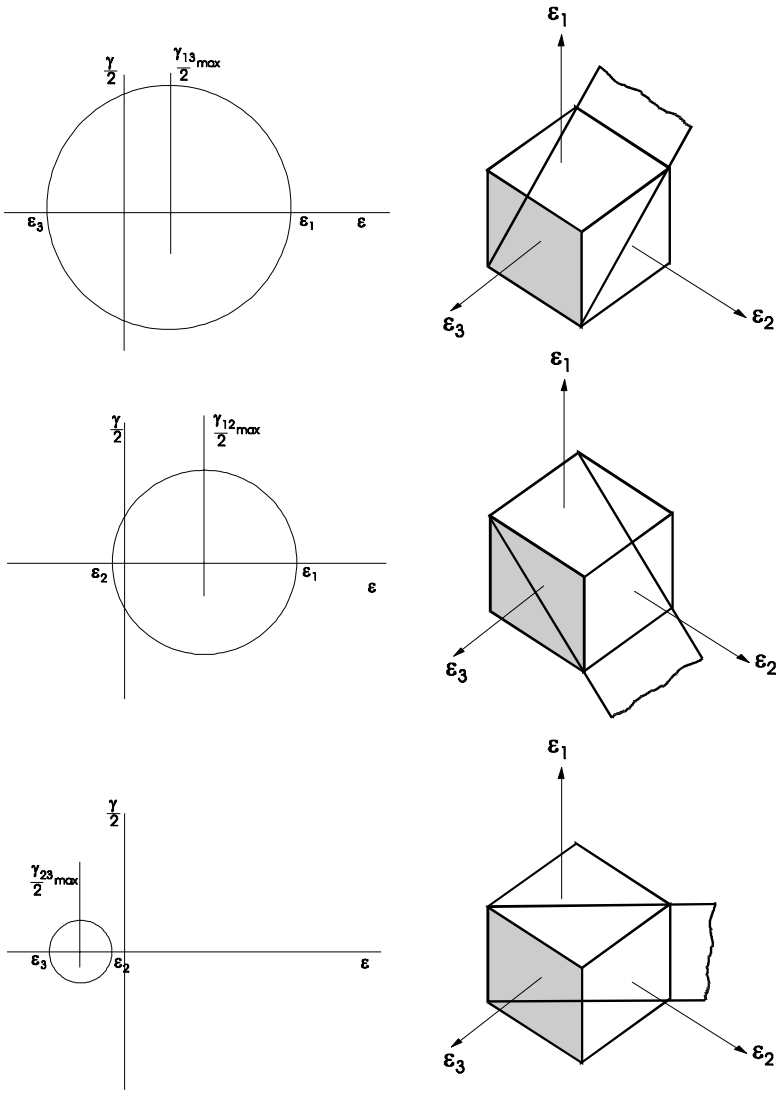


Figure 7.19 Planes of maximum shear strain

7.2.4 Relationships between stress and strain

For plane stress conditions there are direct stresses in the x- and y-planes only.

Considering each stress in turn, σ_x produces a strain $\varepsilon_x = \frac{\sigma_x}{E}$ in the x-direction, and strains $-\nu\varepsilon_x$ in the y- and z- directions.

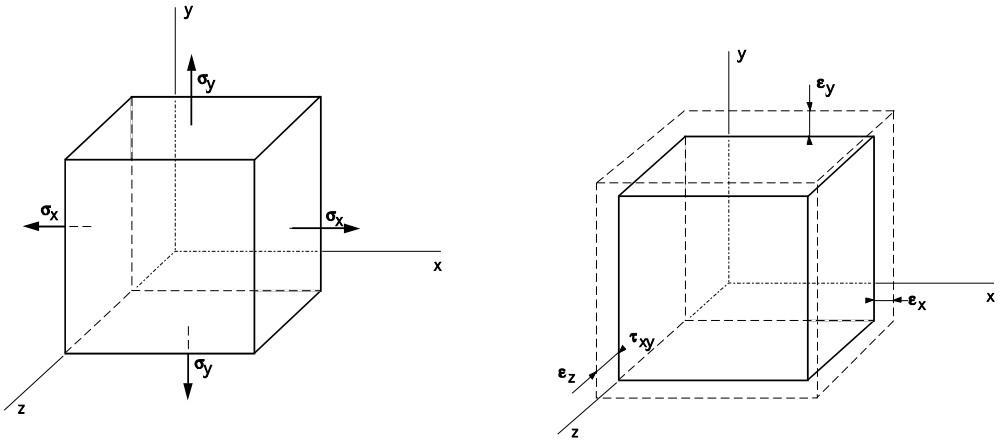


Figure 7.20 Stresses and strains for plane stress

Similarly σ_y produces a strain $\varepsilon_y = \frac{\sigma_y}{E}$ in the y-direction, and strains $-\nu\varepsilon_y$ in the x- and z- directions. The total direct strains are then

$$\varepsilon_x = \frac{1}{E}(\sigma_x - \nu\sigma_y)$$

$$\varepsilon_y = \frac{1}{E}(\sigma_y - \nu\sigma_x)$$

$$\varepsilon_z = \frac{-\nu}{E}(\sigma_x + \sigma_y)$$

(7.11)

From these equations, the stresses can be expressed in terms of strain

$$\sigma_x = \frac{E}{1 - \nu^2}(\varepsilon_x + \nu\varepsilon_y) \qquad \sigma_y = \frac{E}{1 - \nu^2}(\varepsilon_y + \nu\varepsilon_x) \tag{7.12}$$

The strain ε_z in terms of ε_x and ε_y is

$$\varepsilon_z = \frac{-\nu}{1 - \nu}(\varepsilon_x + \varepsilon_y)$$

For the principal strains, the strain perpendicular to the surface of a component is

$$\varepsilon_3 = \frac{-\nu}{1 - \nu}(\varepsilon_1 + \varepsilon_2)$$

(7.13)

7.3 Static yield criteria

Fatigue is a phenomenon associated with local yielding. Early methods of multiaxial fatigue analysis were based on the criteria for yielding under multiaxial stresses.

For uniaxial stress, the material yields at the monotonic yield stress, for a single load application. For biaxial stresses a number of criteria have been proposed for estimating the combination of stresses which will just cause yielding. Two criteria, proposed by Tresca and by von Mises, are shown in *Figure 7.21*.

σ_1 and σ_2 are principal stresses, and σ_y is the uniaxial yield stress.

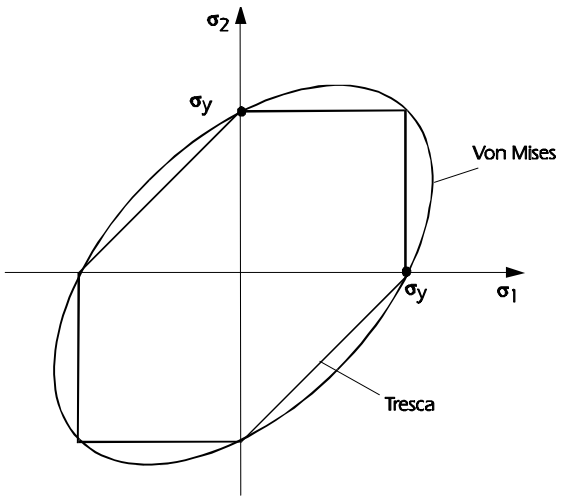


Figure 7.21 Tresca and von Mises yield criteria

Within the von Mises ellipse or the Tresca polygon the material is elastic. Outside these boundaries, the stresses will cause yielding. Experimental results support the von Mises criterion.

A triaxial stress system can be described in terms of the average value of the principal stresses, and the difference between the average value and the individual principal stresses *(Figure 7.22)*. The average value of the principal stresses is called the hydrostatic stress, or spherical stress. It corresponds to a uniform pressure (positive or negative) on each face. The difference between an individual stress and the hydrostatic stress is called the deviatoric stress. Deviatoric stresses are usually denoted by S. It is generally considered that hydrostatic stresses do not contribute to yielding. As yielding is essential for fatigue crack initiation, it has been suggested that crack initiation must be a result of the deviatoric stresses only. (Hydrostatic stresses are now considered to have an effect similar to the effect of mean stress in uniaxial fatigue.)

The hydrostatic stress is the average of the principal stresses

$$\sigma_m = \frac{\sigma_1 + \sigma_2 + \sigma_3}{3} \qquad (7.14)$$

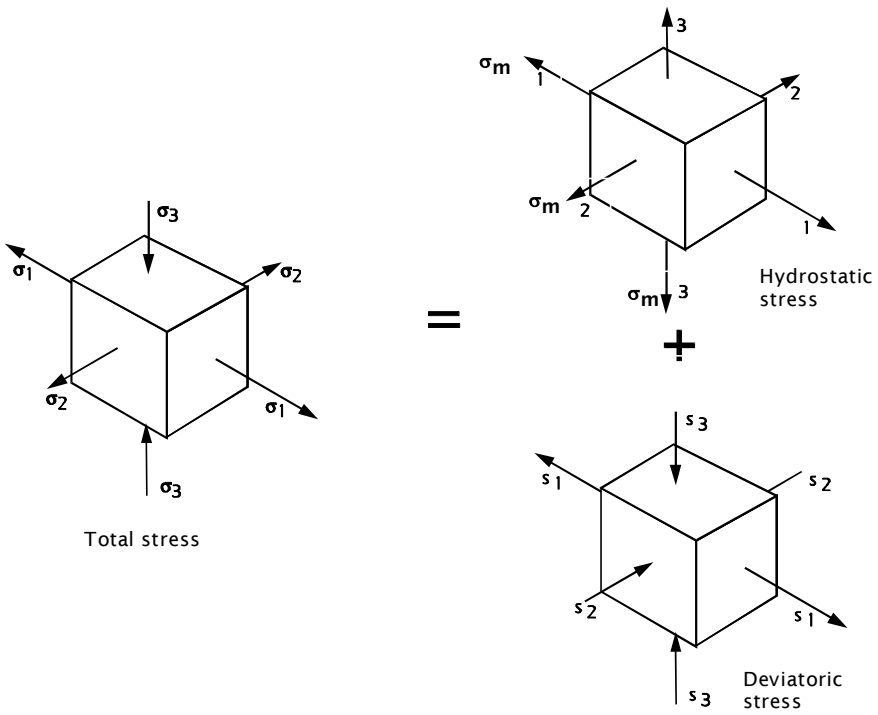


Figure 7.22 Hydrostatic and deviatoric stresses

The hydrostatic stress causes volumetric change, and the deviatoric stresses cause distortion, or shear. The total strain energy U is the sum of the volumetric strain energy U_V and shear strain energy U_S, i.e.

$$U = U_V + U_S$$

It can be shown that $$U_V = \frac{(1-2\nu)(\sigma_1 + \sigma_2 + \sigma_3)^2}{6E}$$

and that $$U = \frac{1}{2E}(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)) - \frac{\nu}{E}(\sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_1\sigma_3))$$

The shear strain energy is then $U_S = U - U_V$

or $$U_S = \frac{(1+\nu)}{6E}\left((\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right)$$

The value of U_S for yielding under uniaxial stress is found by putting σ_1 equal to the uniaxial yield stress σ_Y, and $\sigma_2 = \sigma_3 = 0$,

giving $$U_S = \frac{(1+\nu)\sigma_Y^2}{3E}$$

and hence $$(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 = 2(\sigma_Y)^2 \qquad (7.15)$$

This is the von Mises yield criterion for multiaxial stresses. *(Figure 7.23)*

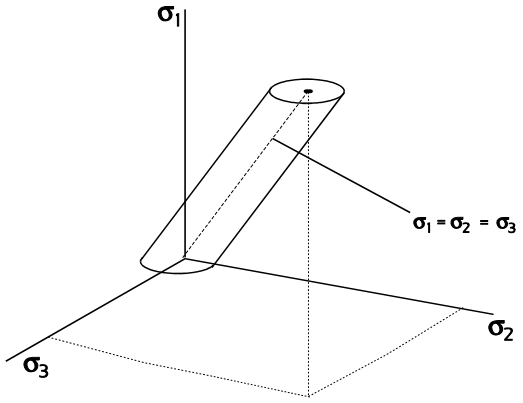


Figure 7.23 von Mises yield criteria in 3-D total stress axes

The ellipse is the von Mises yield ellipse, and the axis of the cylinder represents the hydrostatic stress $\sigma_1 = \sigma_2 = \sigma_3$. As the ellipse has the same size irrespective of its position along the cylinder (i.e. irrespective of the hydrostatic stress) the hydrostatic stress can be ignored in assessing if the material has yielded. This produces a diagram using deviatoric stresses :

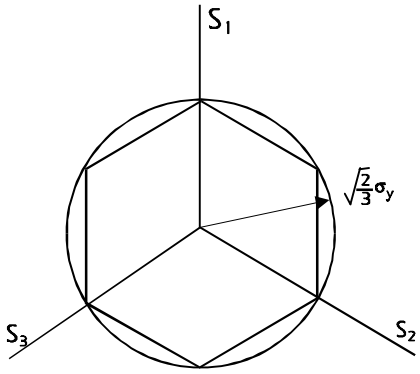


Figure 7.24 von Mises and Tresca yield criteria (deviatoric stresses)

7.4 Multiaxial fatigue using equivalent stress or strain

7.4.1 Introduction

To re-cap, fatigue cracks usually initiate from the surface of a component. It has been shown that the stress distribution on a surface is plane stress. There is usually an out-of-plane strain at the surface, so the strain distribution is triaxial. Multiaxial fatigue theories concentrate on this condition.

The strain-life equation for uniaxial stress is

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c \qquad (7.17)$$

where $\Delta\varepsilon$ is the applied strain range
$2N_f$ is the endurance in reversals
σ'_f is the fatigue strength coefficient
ε'_f is the fatigue ductility coefficient

b is the fatigue strength exponent

c is the fatigue ductility exponent

The following points are important in deriving multiaxial fatigue criteria

(1) The fatigue life for uniaixal stresses must be calculated correctly.

(2) The predicted direction of crack initiation must agree with observations.

(3) An analysis which uses standard uniaxial materials properties has economic advantages.

One approach to multiaxial fatigue is to maintain the form of equation (7.17), thus satisfying item (3), but to use some other strain parameter on the left-hand side and make corresponding changes to the constants on the right hand side. The modifications to the right hand side of the equation are made by considering the uniaxial plane stress condition, so that the proposed biaxial equation gives the same endurance for the uniaxial condition as does the uniaxial equation itself (thus satisfying item (1) above.

This procedure is illustrated in the following sections.

The notation used is that ε_1, ε_2 and ε_3 are the principal strains, with ε_3 as the out-of-plane strain. From equation 7.13

$$\varepsilon_3 = \frac{-\nu}{1-\nu}(\varepsilon_1 + \varepsilon_2) \qquad (7.18)$$

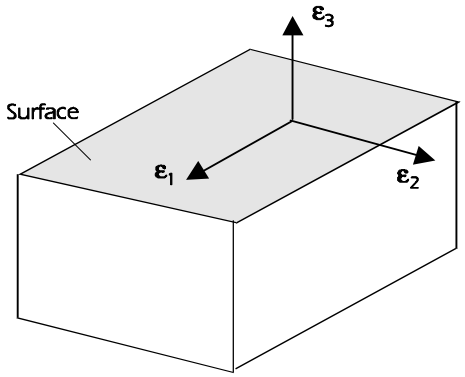


Figure 7.25 Notation for principal strains

7.4.2 Principal strain criterion

This criterion proposes that fatigue cracks initiate on planes which experience the largest amplitude of principal strain. For uniaxial stress the maximum principal strain is the axial strain in the direction of the applied stress.

Replacing the axial strain in equation (7.17) by the principal strain will give a multiaxial fatigue equation (7.19) which gives the same calculated life for uniaxial stresses as does equation (7.17).

$$\boxed{\frac{\Delta\varepsilon_1}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c} \qquad (7.19)$$

For high cycle fatigue the equation reduces to

$$\frac{\Delta\varepsilon_1}{2} = \frac{\sigma'_f}{E}(2N_f)^b$$

or its equivalent in terms of true stress : $\frac{\Delta\sigma_1}{2} = \sigma'_f(2N_f)^b$

At very long endurance the relationship between stress and strain is almost linear, so elastic local stresses at the notch can be substituted for true stresses without significant error.

The principal strain criterion (7.19) is recommended for the analysis of brittle materials, for example cast irons and very high strength steels. It tends to give unsafe life estimates for ductile metals.

7.4.3 Principal stress criterion

Early attempts to analyse biaxial fatigue were based on principal stresses, using a conventional S-N curve. For a general state of stress the principal stress is calculated using the standard equation

$$\sigma_{1,2} = \frac{\sigma_x + \sigma_y}{2} \pm \sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + \tau_{xy}^2}$$

For a fatigue cycle, the stress range of σ_1 , or the stress amplitude $\sigma_1 /2$, would be used with a stress-life curve obtained by testing an axially loaded specimen. The (false) assumption in this procedure is that the fatigue life is always determined by the amplitude of the largest principal stress σ_1 , and therefore that the second principal stress σ_2 has no effect on fatigue life.

Consider a simple circular shaft loaded in pure torsion. If τ is the torsion stress, then the principal stress is calculated from equation 7.1. As there are no axial stresses σ_x and σ_y , the principal stresses are :

$$\sigma_{1,2} = \pm \sqrt{\tau_{xy}^2}$$

i.e. the maximum principal stress is equal to the torsional stress. A fatigue cycle of $\pm \tau_{xy}$ will produce a principal stress cycle of $\pm \sigma_1 = \pm \tau_{xy}$. *The use of the principal stresses therefore predicts that the fatigue strength in torsion is the same as the fatigue strength under axial loading.* This is not supported by test data, as *Figure 7.27* shows.

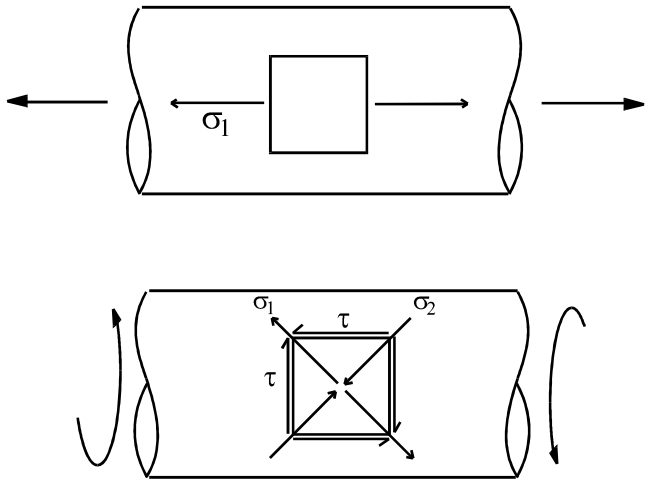


Figure 7.26 Principal stresses for a shaft under axial load and torsion load.

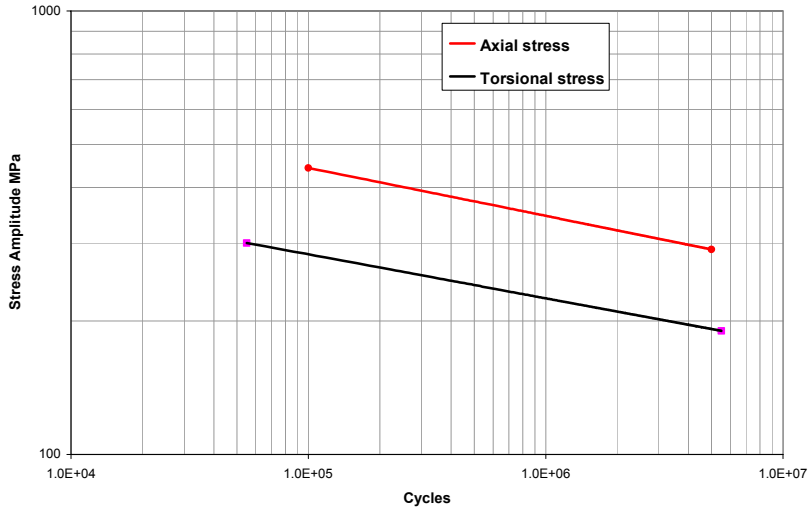


Figure 7.27. Stress-life curves for axial and torsion loading

Figure 7.27 shows the results of fatigue tests on a commonly used steel. It is clear that the torsional fatigue strength is much lower than the axial fatigue strength - the allowable principal stress in torsion is approximately 60% of the allowable axial stress. Calculating fatigue lives using principal stress will clearly be grossly optimistic for torsion loading, and allowable torsional fatigue stresses will be overestimated by a factor of 1/0.6 = 1.66. This could mean the difference between identifying and missing a potential fatigue 'hot spot'.

It has been shown over the past 20 years that principal stresses should only be used for fatigue analysis of 'brittle' metals, for example cast irons and very high strength steels. *A fatigue analysis using principal stresses gives very unsafe fatigue life predictions for more ductile metals including most commonly used steels.*

7.4.4 Maximum shear strain criterion

It has been observed that fatigue cracks often initiate on shear planes. The maximum shear strain criterion proposes that cracks will initiate on planes which experience the maximum shear strain amplitude.

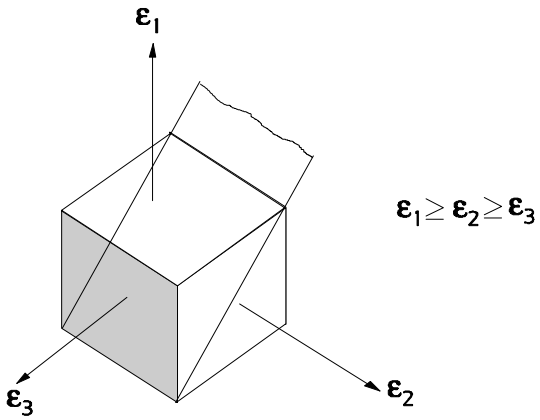


Figure 7.28 Plane of maximum shear strain

If $\varepsilon_1 \geq \varepsilon_2 \geq \varepsilon_3$, crack initiation takes place on the plane at 45° to the plane perpendicular to ε_1 *(Figure 7.28)*.

The maximum shear strain can be calculated from the principal strains using Mohr's strain circle *(Figure 7.29)*. If $\varepsilon_1 > \varepsilon_2 > \varepsilon_3$ the maximum shear strain is given by

$$\frac{\gamma_{max}}{2} = \frac{\varepsilon_1 - \varepsilon_3}{2} \qquad (7.20)$$

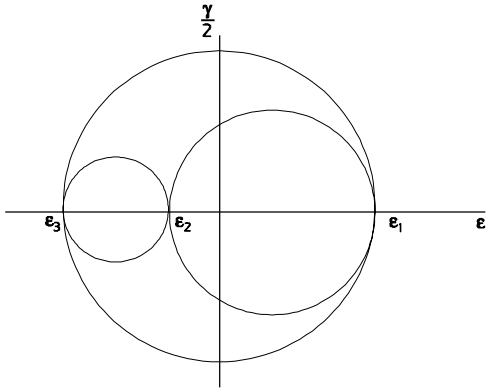


Figure 7.29 Mohr's strain circle

If the conventional strain-life equation

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c \qquad \text{(from 7.17)}$$

is written with the shear strain amplitude on the left hand side

$$\frac{\Delta\gamma}{2} = C1\frac{\sigma'_f}{E}(2N_f)^b + C2\,\varepsilon'_f(2N_f)^c \qquad (7.20a)$$

the constants C1 and C2 can be evaluated by making the two equations give the same endurance for uniaxial stresses.

For uniaxial stress, with an axial strain ε_1 , the principal strains $\varepsilon_2 = \varepsilon_3 = -\nu\varepsilon_1$,

so from (7.20),

$$\gamma_{max} = \varepsilon_1 - \varepsilon_3 = \varepsilon_1 - (-\nu\varepsilon_1) = (1+\nu)\,\varepsilon_1$$

For uniaxial stresses, the shear strain amplitude is $(1+\nu)$ times the direct strain amplitude, so equation (7.21a) can be re-written as

$$\frac{\Delta\gamma}{2} = (1+\nu_e)\frac{\sigma'_f}{E}(2N_f)^b + (1+\nu_p)\,\varepsilon'_f(2N_f)^c$$

For elastic strains Poisson's ratio ν_e is approximately 0.3 and for purely plastic strains, ν_p is 0.5

The uniaxial strain-life equation (7.17), expressed in terms of γ_{max} , is therefore

$$\frac{\Delta\gamma_{max}}{2} = 1.3\frac{\sigma'_f}{E}(2N_f)^b + 1.5\,\varepsilon'_f(2N_f)^c$$

elastic strain plastic strain

(7.21)

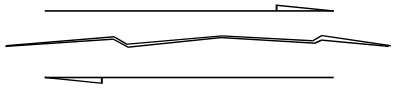
This equation tends to give conservative life estimates for ductile metals, but can give unsafe life estimates for brittle metals.

7.4.5 Comparison of allowable direct strain and shear strain

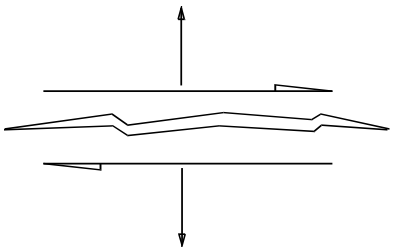
Standard references (for example Ref. 7.8) suggest that for a given fatigue life, the allowable shear stress is 0.58 times the allowable direct stress. Re-writing equation (7.21) in terms of stresses shows that the allowable shear stress is approximately 0.6 times the allowable direct stress. The allowable shear strain is 1.3 to 1.5 times the allowable direct strain.

7.4.6 McDiarmid's criterion

Findlay proposed that fatigue damage under multiaxial conditions must be a function of both shear and direct stress. For a small crack initiation site, the physical dislocation under shear stress will be as shown below. In engineering terms it can be supposed that some of the shear forces will be reacted by friction and by contact at mating faces, reducing the shear forces at the tips, and so reducing the rate at which damage develops at the tips.



An additional direct stress will open the dislocation, reducing the friction and the forces reacted by mating faces, and increasing the development of damage at the tips.



Fatigue damage must therefore be a function of both shear and direct stress.

Based on Findlay's proposal, McDiarmid's criterion (Ref. 7.3) is

$$\frac{\tau_o}{t} + \frac{\sigma_{n\,max}}{2\sigma_T} = 1 \qquad (7.22)$$

where τ is the shear stress amplitude on the plane of maximum shear stress

σ_n is the normal stress on this plane

t is the fatigue strength in torsion at the specified endurance (usually the endurance limit)

σ_t is the tensile strength of the material

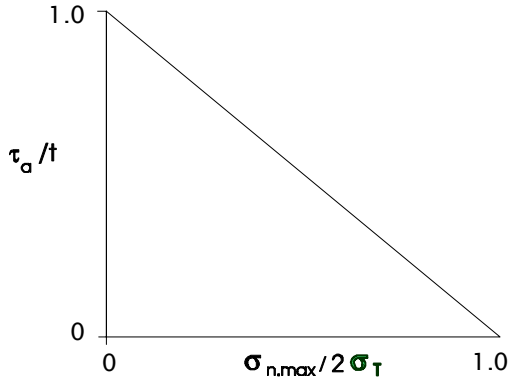


Figure 7.30 McDiarmid's criterion

For pure torsion, there is no normal stress on the plane of maximum shear stress, so the value of $\frac{\tau_o}{t}$ at $\sigma_{n\ max} = 0$ represents a pure torsion test. McDiarmids parameter therefore requires torsion test data. The diagram in *Figure 7.30* is shows the combinations of torsion stress amplitude and normal stress which will give the same endurance as the pure torsion stress amplitude.

McDiarmid's criterion has been assessed against published and other test data, and given good agreement at high endurance.

For in-phase combined bending and torsion the method gives the following results:

For materials with a ratio of bending fatigue strength b to torsion fatigue strength t in the range 1.55-1.75, allowable stresses are predicted to within 5%.

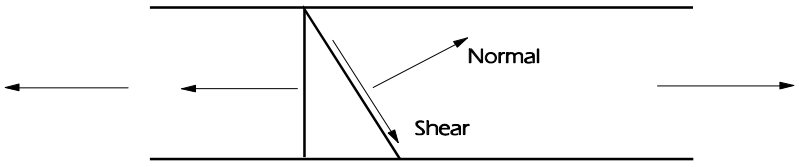
For b/t > 1.75, the method is more than 5% conservative

For b/t < 1.55, the method is more than 5% optimistic. Shear-stress-based criteria should not be used for these materials.

For out-of-phase loading, estimation of endurance limit stresses is generally within ±10%.

7.4.7 Brown-Miller combined strain criterion

The Brown-Miller equation proposes that the maximum fatigue damage occurs on the plane which experiences the maximum shear strain amplitude, and that the damage is a function of both this shear strain and the strain normal to this plane.



If the maximum shear strain is γ_{max}

and the strain normal to the maximum shear strain is ε_n

then from Mohr's strain circle

$$\frac{\gamma_{max}}{2} = \frac{\varepsilon_1 - \varepsilon_3}{2} \qquad \text{and} \qquad \varepsilon_n = \frac{\varepsilon_1 + \varepsilon_3}{2}$$

For uniaxial plane stress, $\varepsilon_2 = -\nu\varepsilon_1$ $\varepsilon_3 = -\nu\varepsilon_1$ so that

$$\gamma_{max} = \varepsilon_1 - \varepsilon_3 = (1+\nu)\varepsilon_1 \quad \text{and} \quad \varepsilon_n = \frac{\varepsilon_1 + \varepsilon_3}{2} = \frac{(1-\nu)\varepsilon_1}{2}$$

The conventional strain-life equation

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c$$

is re-written with the shear strain amplitude and normal strain amplitude on the left hand side

$$\frac{\Delta\gamma_{max}}{2} + \frac{\Delta\varepsilon_n}{2} = C1\ \frac{\sigma'_f}{E}(2N_f)^b + C2\ \varepsilon'_f(2N_f)^c$$

elastic plastic

For elastic stresses, $\nu_\varepsilon = 0.3$ and so

$$\gamma_{max} = (1+\nu_\varepsilon)\varepsilon_1 = 1.3\ \varepsilon_1 \quad \text{and} \quad \varepsilon_n = \frac{(1-\nu_\varepsilon)\varepsilon_1}{2} = 0.35\ \varepsilon_1$$

The constant C1 is therefore 1.3 + 0.35 = 1.65

For the plastic value of Poissons ratio of 0.5, the same mathod give C2 = 1.75

and the complete Brown-Miller strain-life equation is then

$$\boxed{\frac{\Delta\gamma_{max}}{2} + \frac{\Delta\varepsilon_n}{2} = 1.65\ \frac{\sigma'_f}{E}(2N_f)^b + 1.75\ \varepsilon'_f(2N_f)^c} \qquad (7.23)$$

This formulation of the Brown-Miller parameter was developed by Kandil, Brown and Miller (Ref. 7.4), and differs from that given in the original paper (Ref. 7.5).

The constants C1=1.65 and C2 = 1.75 were derived on the assumption that cracks initiate on the plane of maximum shear strain. For complex variable amplitude loading, it was found that better agreement with test results was obtained if it was assumed that the most damaged plane was the one that produced the highest value of $\frac{\Delta\gamma_{max}}{2} + \frac{\Delta\varepsilon_n}{2}$. The constants C1 and C2 will have slightly different values on this plane. However, the values shown in equation (7.23) are used almost universally.

The Brown-Miller equation gives the most realistic life estimates for ductile metals, and tends to be non-conservative for brittle metals.

For design calculations, test data was plotted as constant life curves for different combinations of in - phase direct and shear strain.

An expression $$\left(\frac{\gamma_{max}}{g}\right)^i + \left(\frac{\varepsilon_n}{h}\right)^i = 1$$

was fitted to the data. The results showed that the values of *g*, *h*, and *i*, varied with endurance. It was suggested that a value of *i* = 2 would give conservative life estimates, and the values of g and h for any given endurance can then be determined from conventional uniaxial strain-controlled smooth-specimen tests. Hence the expression becomes the Brown-Miller ellipse.

$$\left(\frac{\gamma_{max}}{g}\right)^2 + \left(\frac{\varepsilon_n}{h}\right)^2 = 1 \qquad (7.24)$$

The Γ-plane diagram in *Figure 7.30* is a plot of the plastic component of normal strain vs the plastic component of maximum shear strain. It refers only to type B cracks, which are those which extend into a component. Type A cracks are shallow cracks on the surface. They can occur under pure torsion loading and would give longer lives than type B cracks. Type B cracks are clearly the most dangerous. If a life contour for a type A crack is added to the diagram it can intersect a type B life contour which represents a shorter life, implying that two values of endurance can result from the same strain distribution, with the type A crack giving the longer endurance.

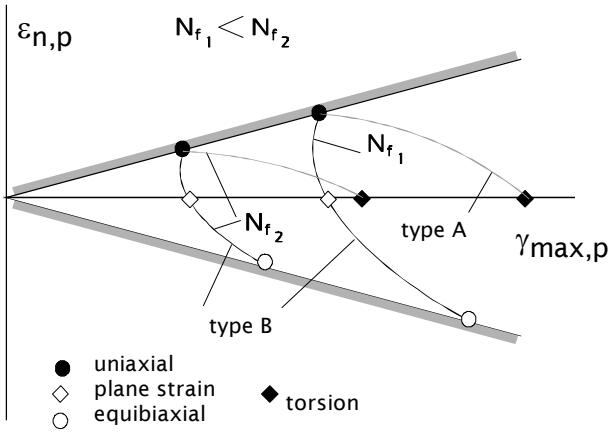


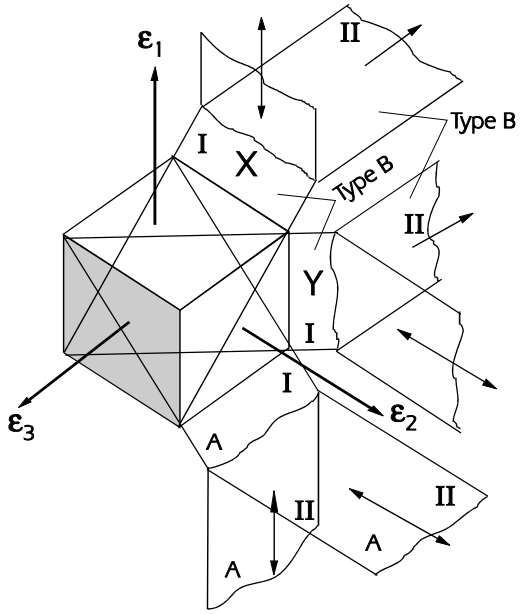
Figure 7.31 Γ-plane for type A and type B cracks

Figure 7.32 shows the possible crack growth directions, from (Ref. 7.6).

Stage I growth is on planes of maximum shear strain, and stage II growth is on planes at 45^o to these planes. For isotropic materials,

growth along X predominates if $\varepsilon_1 \geq \varepsilon_2 \geq \varepsilon_3$

growth along Y is possible if $\varepsilon_2 = \varepsilon_1$



The shaded face is the surface of the component.

Figure 7.32 Possible crack orientations

7.4.8 von Mises Equivalent Strain

Because the von Mises criterion provides an estimate of the onset of yielding, it has been proposed as a criterion for fatigue life estimation.

The strain-life equation for uniaxial stress is:

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c$$

From Section 7.3 the equation for von Mises strain, calculated from principal strains, is

$$\varepsilon_{eff} = \beta\left((\varepsilon_1 - \varepsilon_2)^2 + (\varepsilon_2 - \varepsilon_3)^2 + (\varepsilon_3 - \varepsilon_1)^2 \right)^{\frac{1}{2}} \qquad (7.25)$$

The value of β is chosen so that ε_{eff} has the same value as the strain ε_1 for the uniaxial stress condition.

For uniaxial stresses, $\varepsilon_2 = \varepsilon_3 = -\nu\varepsilon_1$ so that $\varepsilon_{eff} = \sqrt{2}\,(1+\nu)\,\varepsilon_1$

For elastic conditions, $\nu_e = {}^1/_3$ so that $\beta = \frac{3}{4\sqrt{2}} = 0.53$

For fully plastic conditions, $\nu_p = {}^1/_2$ so that $\beta = \frac{\sqrt{2}}{3} = 0.47$

For general elastic-plastic stresses, the following expression for β has been proposed for in-phase loading

$$\beta = \frac{1}{(1+\nu^*)\sqrt{2}} \quad \text{where} \quad \nu^* = \frac{\nu_e\,\varepsilon_e + \nu_p\,\varepsilon_p}{\varepsilon_t}$$

and ε_e, ε_p are the elastic and plastic strains and ε_t is the total strain.

The strain-life equation in terms of von Mises equivalent strain is

$$\boxed{\frac{\Delta\varepsilon_{eff}}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c} \qquad (7.26)$$

where ε_{eff} is given by equation (7.25).

For design analysis based on calculated stresses, at high endurance where the plastic component is small, the von Mises equivalent stress is

$$\sigma_{eff} = \frac{1}{\sqrt{2}}\left((\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right)^{\frac{1}{2}}$$

$$\frac{\Delta\sigma_{eff}}{2} = \sigma'_f(2N_f)^b \qquad (7.27)$$

A major problem with the practical application of von Mises criteria to measured signals is that the von Mises stress or strain is always positive, even for negative values of stress or strain, and so Rainflow cycle counting cannot be applied directly. Some approximations have been proposed, such as assigning the sign of the largest stress or strain to the von Mises stress or strain, or alternatively to assign the sign of the hydrostatic stress or strain to the von Mises stress or strain. These are termed *'signed von Mises'* criteria. The different methods of determining the sign can give significantly different life estimates.

The von Mises stress is one of a family of stresses, including octahedral shear stress, which are functions of $\left((\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right)$ The octahedral shear stress interpretation implies that fatigue damage is oriented towards the octahedral shear planes.

von Mises criteria correlate poorly with test data, particularly for biaxial stresses when the two in-plane principal stresses change their orientation during the fatigue loading.

7.4.9 SAE notched shaft test results

The SAE multiaxial test programme uses a 40mm diameter notched shaft with 5mm fillet radii, machined from SAE1045 steel. The specimens were tested under pure bending loads, pure torsion loads, and combined bending-torsion with various proportions of bending and torsion. The tests use variable amplitude loading signals derived from service measurements.

The test results have been compared with life estimates made from measured strains at the notch (Ref. 7.10). The strain measurement showed that circumferential strains were not fully suppressed on this specimen. Life predictions were made for the two limiting conditions - fully suppressed and not suppressed at all.

Comparisons of predicted and test lives for the principle strain, maximum shear strain, Brown-Miller, and von Mises criteria are shown in *Figures 7.33 - 7.36*, reproduced from (Ref. 7.10)

The maximum principal strain criterion produced life estimates which were non-conservative, particularly at lower values of endurance, and the scatter was large. The maximum shear strain criterion produced more conservative life estimates, with similar scatter. The von Mises strain criterion produced non-conservative life estimates at long and short lives. The Brown-Miller parameter was the most successful, with the least scatter, and most life predictions were within a factor of ± 3 of the test lives.

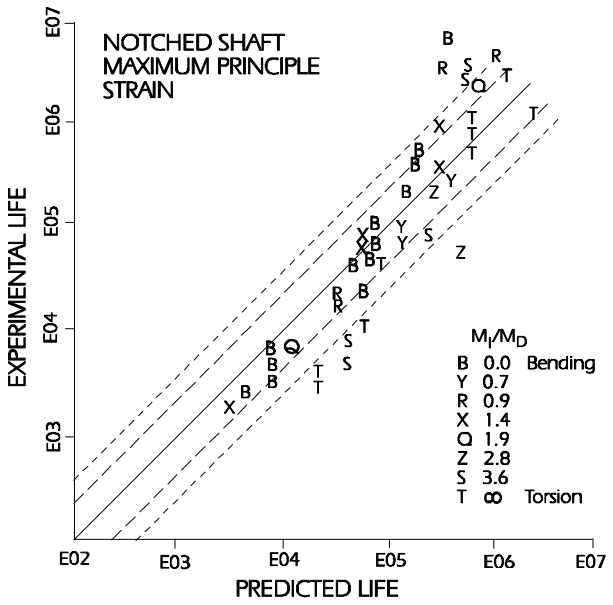


Figure 7.33 SAE notched shaft, principal strain theory

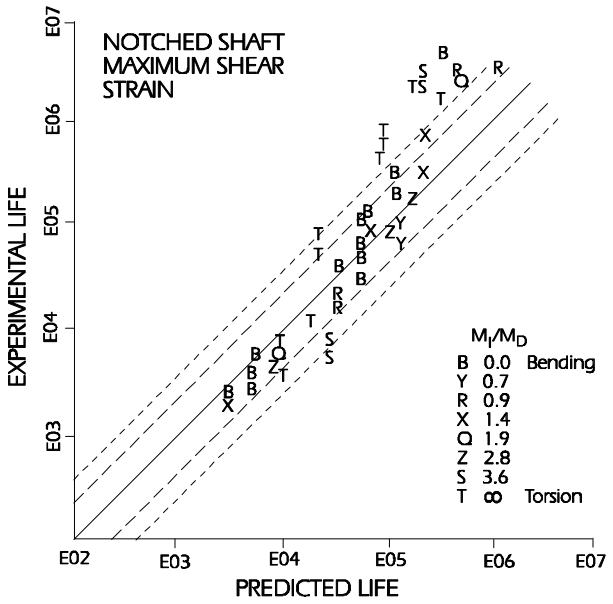


Figure 7.34 SAE notched shaft, maximum shear strain theory

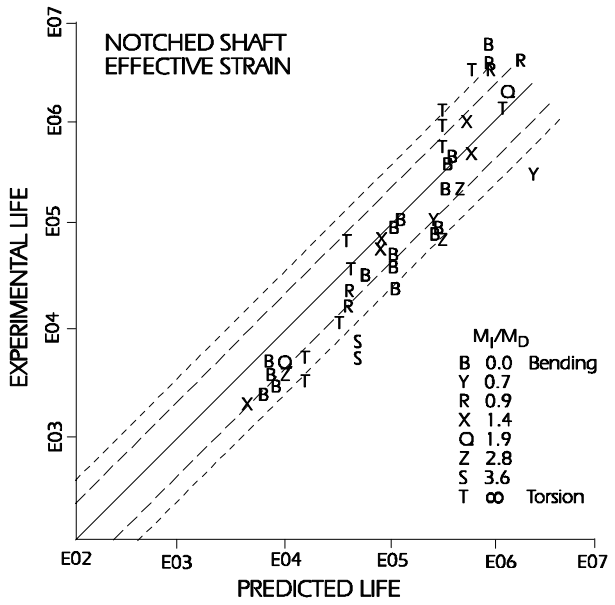


Figure 7.35 SAE notched shaft, von Mises strain theory

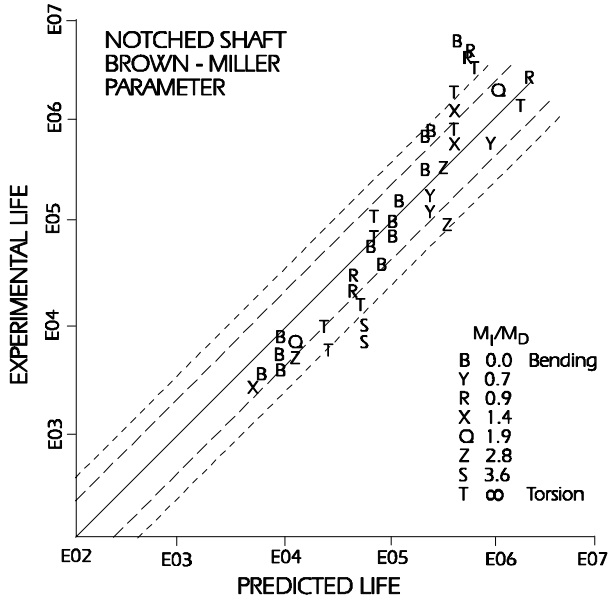
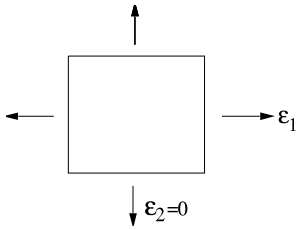


Figure 7.36 SAE notched shaft, Brown-Miller parameter

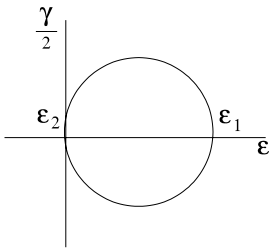
Example 7.1

Direct strain with fully restrained transverse strains.



ε_1 cycles from 0 to 800 $\mu\varepsilon$

Mohr's circle is

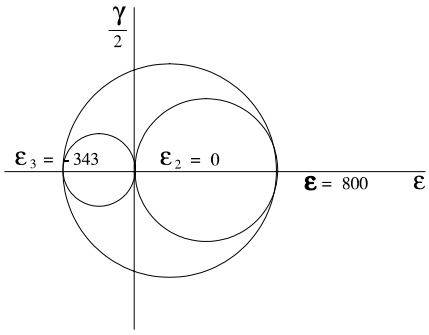


The out-of-plane strain ε_3 is given by

$$\varepsilon_3 = \frac{-\gamma}{1-\gamma}(\varepsilon_1 + \varepsilon_2)$$

and taking $\gamma = -0.3$ $\varepsilon_3 = -343$

Mohr's circles for the three planes are:



The maximum shear strain is given by

$$\frac{\gamma_{max}}{2} = \frac{\varepsilon_1 - \varepsilon_3}{2}$$

$= 572$

so $\gamma_{max} = 1140\mu\varepsilon$

The normal strain $\varepsilon_n = \frac{\varepsilon_1 + \varepsilon_3}{2} = 229\ \mu\varepsilon$

To calculate the endurance, the three biaxial fatigue equations will be simplified to:

Principal strain criterion: $\frac{\Delta\varepsilon_1}{2} = \frac{\sigma'_f}{E}(2N_f)^b$

Maximum shear criterion: $\frac{\Delta\gamma_{max}}{2} = 1.3\frac{\sigma'_f}{E}(2N_f)^b$

Brown - Miller criterion: $\frac{\Delta\gamma_{max}}{2} + \frac{\Delta\varepsilon_n}{2} = 1.65\frac{\sigma'_f}{E}(2N_f)^b$

The material properties are σ'_f = 1000 MPa, b = -0.1 and E = 2x10^5 MPa

The calculated values of endurance are:

For the principal strain criterion:

$$\frac{0.0008}{2} = \frac{1000}{2x10^5}(2N_f)^{-0.1}$$

so $2N_f = 9.31 \times 10^{10}$

For the maximum shear criterion

$$\frac{0.00114}{2} = \frac{1.3 \times 1000}{2 \times 10^5}(2N_f)^{-0.1}$$

so $2N_f = 3.72 \times 10^{10}$

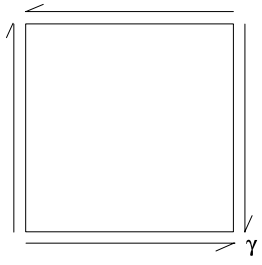
For the Brown - Miller criterion

$$\frac{0.00114}{2} + \frac{0.000229}{2} = 1.65 \times \frac{1000}{2x10^5}(2N_f)^{-0.1}$$

so $2N_f = 6.47 \times 10^{10}$

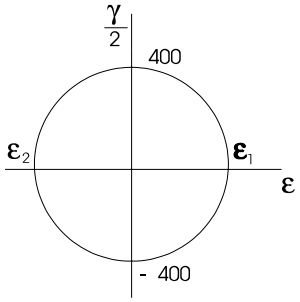
Example 7.2

Pure torsion



γ cycles from 0 to 800 με

Mohr's circle is



The principal strains are

$$\varepsilon_1 = 400\ \mu\varepsilon\ , \quad \varepsilon_2 = -400\ \mu\varepsilon$$

The maximum shear strain $\gamma_{max} = 800\ \mu\varepsilon$

The normal strain $\varepsilon_n = 0$

Taking the same equations and materials data as Example 7.1

For the principal strain criterion

$$\frac{0.0004}{2} = \frac{1000}{2 \times 10^5} (2N_f)^{-0.1}$$

so $2N_f = 9.54 \times 10^{13}$

For the maximum shear criterion

$$\frac{0.0008}{2} = 1.3 \times \frac{1000}{2 \times 10^5} (2N_f)^{-0.1}$$

so $2N_f = 1.28 \times 10^{12}$

For the Brown - Miller criterion

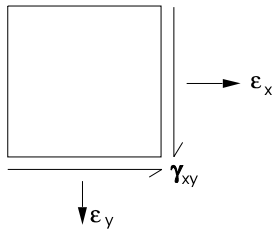
$$\frac{0.0008}{2} + 0 = 1.65 \times \frac{1000}{2 \times 10^5} (2N_f)^{-0.1}$$

so $2N_f = 1.39 \times 10^{13}$

Example 7.3

Torsion combined with direct strain and restrained transverse strain.

This example combines examples 7.1 and 7.2

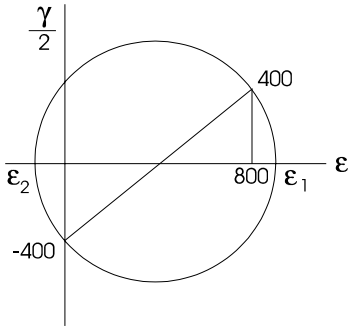


ε_x cycles between 0 and 800 με

γ_{xy} cycles between 0 and 800 με in phase

$\varepsilon_y = 0$

Mohr's circle is



The radius of the circle is $(400^2 + 400^2)^{1/2}$ = 566 με

So ε_1 = 400 + 566 = 966

ε_2 = 400 - 566 = -166

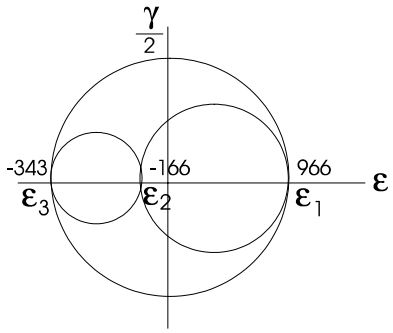
$\frac{\gamma_{max}}{2}$ = 566 με so γ_{max} = 1130 με

The out-of-plane strain ε_3 is

$$\varepsilon_3 = \frac{-\gamma}{1-\gamma}(\varepsilon_1 + \varepsilon_2) = \frac{-0.3}{1-0.3}(966 - 166)$$

$$= -343\ \mu\varepsilon$$

Mohr's circles for the 3 planes are



The maximum shear strain

$$\frac{\gamma_{max}}{2} = \frac{966 + 343}{2} = 655\ \mu\varepsilon$$

so $\gamma_{max} = 1310\ \mu\varepsilon$

$$\varepsilon_n = \frac{966 - 343}{2} = 312\ \mu\varepsilon$$

Using the same equations and materials data as examples 7.1 and 7.2

For principal strain criterion

$$\frac{0.000966}{2} = \frac{1000}{2 \times 10^5} (2N_f)^{-0.1}$$

so $2N_f = 1.41 \times 10^{10}$

For the maximum shear criterion

$$\frac{0.001310}{2} = 1.3 \times \frac{1000}{2 \times 10^5}$$

so $2N_f = 9.26 \times 10^9$

For the Brown - Miller criterion

$$\frac{0.001310}{2} + \frac{0.000312}{2} = 1.65 \times \frac{1000}{2 \times 10^5} (2N_f)^{-0.1}$$

so $2N_f = 1.19 \times 10^{10}$

7.5 Critical plane analysis

7.5.1 Introduction

For many components subjected to combined direct and shear stresses, the phase relationship between the stresses is not constant. Examples are rotating shafts in bending with torsion forces applied at a frequency different from the rotational frequency. Machine tools, and powered axles with ABS braking or traction control, are examples. In these cases it is not obvious which plane will experience the most severe combination of strains and hence the highest fatigue damage.

Figure 7.37 shows an idealised time history of bending and shear strains from a powered railway axle with traction control experiencing wheel slip. The cyclic bending strains are caused by the rotation of the axle. The shear strain cycles at a higher frequency as the traction control repeatedly applies and removes power to the axle. The angle θ of the maximum principal strain to the axis of the axle is also shown.

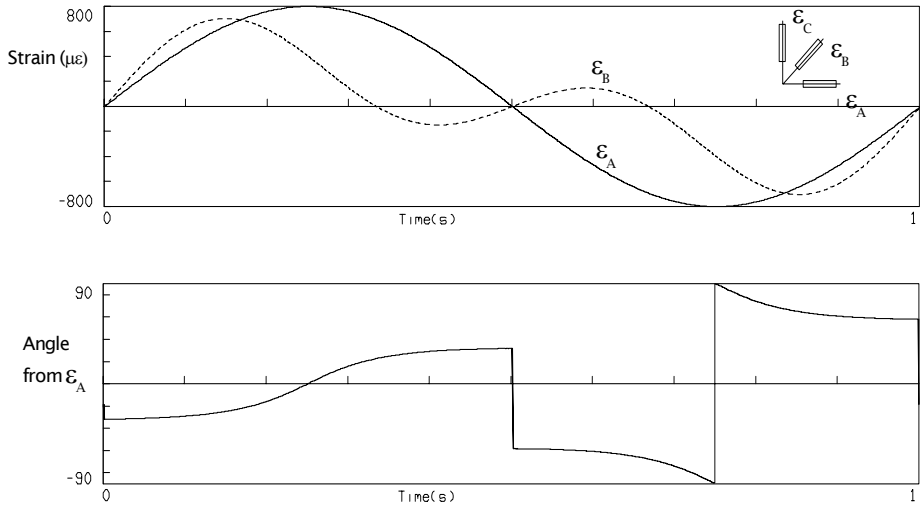


Figure 7.37 Direction of principal and shear strains on an axle

In this case it is not obvious which plane experiences the largest strain cycles. If a fatigue analysis was carried out using the principal strain-life equation (7.19), the plane on which the principal strain acts changes its orientation throughout the loading cycle, as shown in *Figure 7.37*. The plane of maximum shear strain is at 45° to the principal strain plane.

Critical plane methods resolve the strains onto a number of planes, and calculate the damage on each plane. This form of analysis must be applied for criteria such as principal stress/strain, maximum shear stress/strain, and the Brown-Miller criterion, for complex strain signals with varying phase relationships.

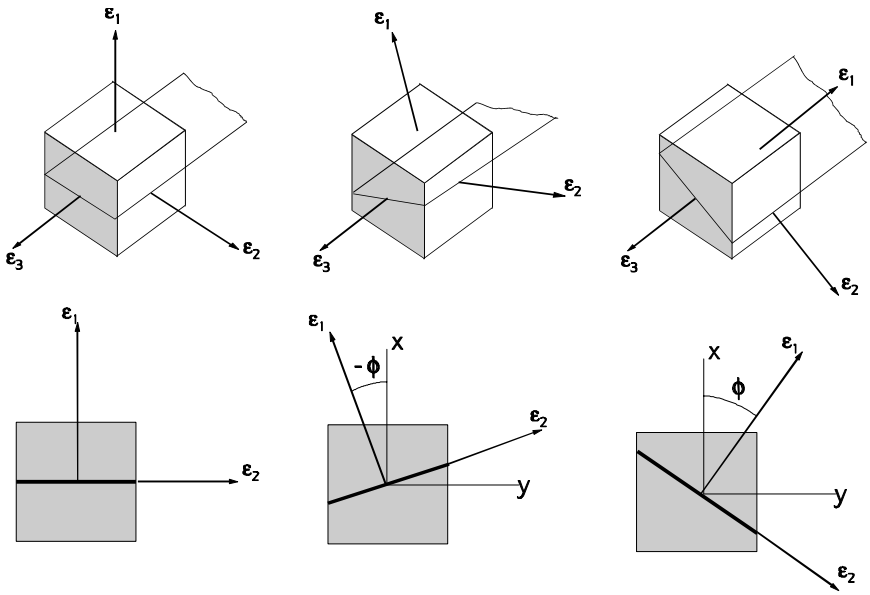


Figure 7.38 Rotation of the principal plane

The orientation of a plane is defined by the angle that the normal to the plane makes with an x-y-z co-ordinate system orientated as *Figure 7.39*. The plane has an angle ϕ from the x-axis towards the y-axis, and an angle θ from the z-axis towards the x-y plane.

For the principal planes shown in *Figure 7.38*, the angle $\theta = 90^{o}$, and ϕ can vary between 0 and 180^{o} *(Figure 7.39)*. For the planes of maximum shear strain, the three possible planes are shown in *Figure 7.40*. Directional cosine matrices are used to resolve the strains onto the chosen plane, and the time history of shear or normal strain on the plane is cycle counted.

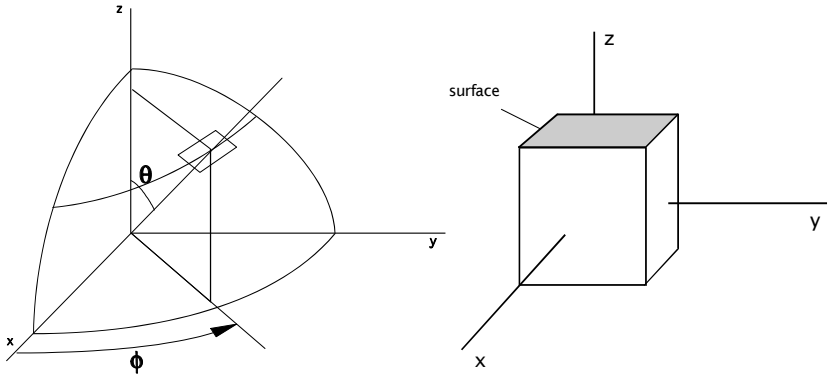


Figure 7.39 Definition of plane orientation to the x-y-z axes

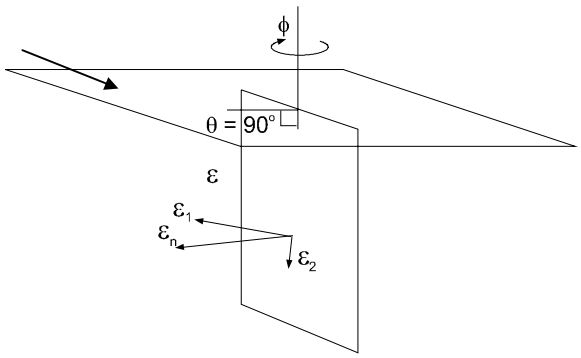


Figure 7.40 (a) Critical plane orientation for a principal strain analysis (brittle metals)

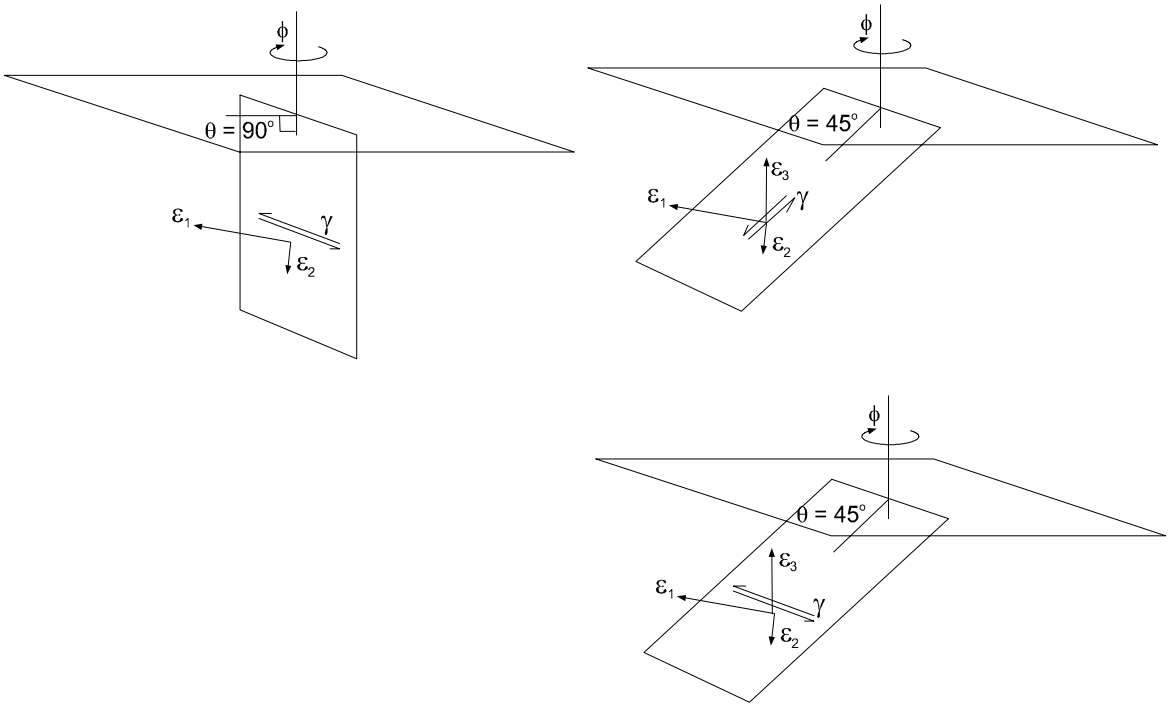


Figure 7.40 (b) Critical plane orientation for a shear strain and Brown-Miller analysis (ductile metals)

The results of critical plane analysis can be plotted to show the orientation of the most damaged plane. *Figure 7.41* shows results from an axle with pure bending and out-of-phase bending and torsion strains.

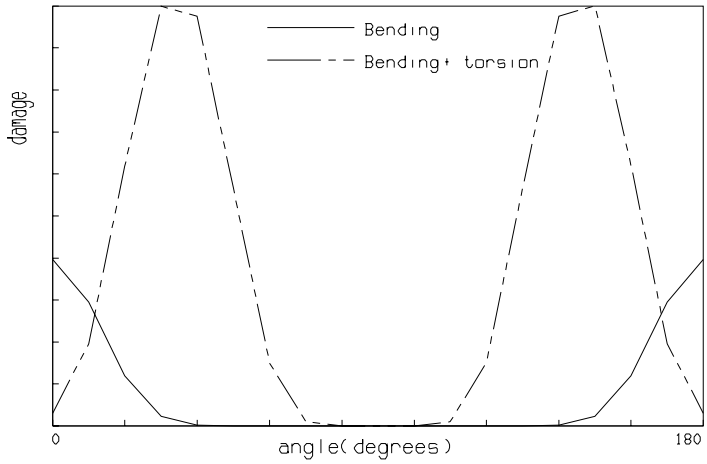


Figure 7.41 Planes of maximum fatigue damage

A 10° increment between planes is often used. *Figure 7.42* shows the effect on calculated life of varying this increment. All results are referred to the life calculated for a 1° increment. A 10° increment produces an error in calculated life of less than 2% compared with a 1° increment.

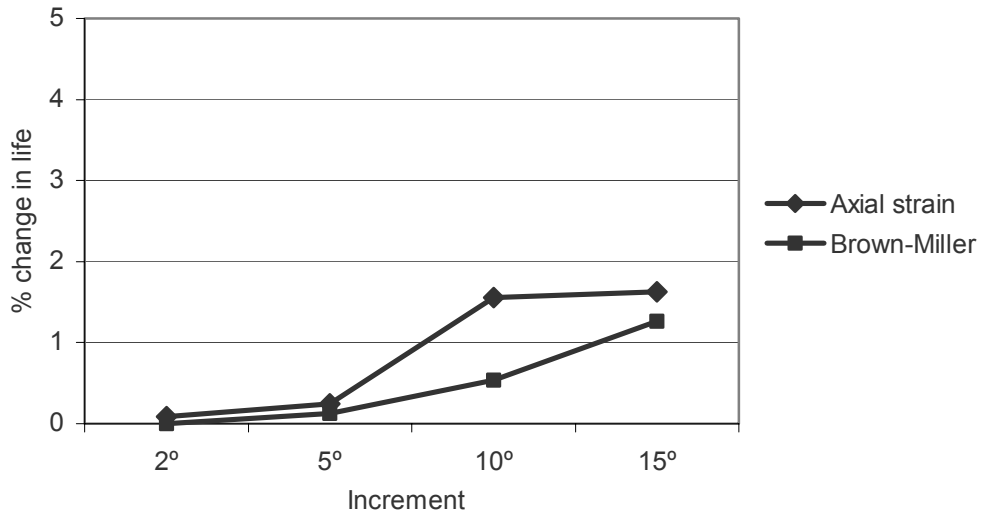


Figure 7.42 Effect of angle increment on calculated fatigue life

7.6 Effects of mean stress

7.6.1 Fatigue life algorithms

Socie and Bannantine (Ref. 7.2) have proposed the following variation on the Smith-Watson-Topper mean stress correction.

$$\frac{\Delta\varepsilon_1}{2}\sigma_{1max} = \frac{\sigma_f'^2}{E}(2N_f)^{2b} + \varepsilon_f'\sigma_f'(2N_f)^{b+c} \qquad (7.28)$$

where ε_1 = maximum principal strain

and $\sigma_{1,max}$ = maximum normal stress on the plane of maximum principal strain

For many years it was considered that mean shear stresses did not influence the fatigue life. For materials where the damage development is dominated by shear strains, Fatemi and Socie proposed the following relationship

$$\frac{\Delta\gamma_{max}}{2}\left(1+\frac{\sigma_n}{\sigma_y}\right) = \frac{\tau'_f}{G}(2N_f)^b + \gamma'_f(2N_f)^c \qquad (7.29)$$

where $\frac{\Delta\gamma_{max}}{2}$ = maximum shear strain amplitude

σ_n = maximum normal stress on the plane of maximum shear strain amplitude

and σ_y = uniaxial yield stress

Equation (7.29) requires torsion test data. However, Bannantine (Ref. 7.2) showed that the torsion constants τ'_f and γ'_f can be approximated by $\tau'_f = \sigma'_f/\sqrt{3}$ and $\gamma'_f = \varepsilon'_f\sqrt{3}$

The loss of accuracy was considered acceptable. However, there are more recent developments in mean stress corrections for shear-dominated multiaxial fatigue crack initiation, and the Fatemi-Socie equation (7.29) is becoming less used.

The Brown-Miller equation can be modified by including the Morrow mean stress correction

$$\frac{\Delta\gamma_{max}}{2}+\frac{\Delta\varepsilon_n}{2} = 1.65\,\frac{(\sigma'_f - \sigma_{n,m})}{E}(2N_f)^b + 1.75\,\varepsilon'_f(2N_f)^c \qquad (7.30)$$

where $\sigma_{n,m}$ is the mean normal stress on the plane.

Recently, Chu, Conle and Bonnen (Ref 7.14) have shown improved correlation if the mean shear stress is included, and have proposed the following extension to the Brown-Miller equation, using a mean stress correction similar to a Smith-Watson-Topper correction

$$\frac{\Delta\gamma.\,\tau_{max}}{2}+\frac{\Delta\varepsilon_n.\sigma_{n\,max}}{2} = f(2N_f) \qquad (7.31)$$

where τ_{max} is the maximum shear stress

and $\sigma_{n\,max}$ is the maximum normal stress.

7.6.2 Multiaxial cyclic hardening

These criteria differ from the strain-based multiaxial fatigue criteria already described in that they require stresses in addition to strain amplitudes. This requires a method of calculating true cyclic elastic-plastic stresses at any point in a complex strain history under multiaxial cyclic strain conditions.

For uniaxial stress, if yielding occurs, the stress developed for any strain is *path dependent*. This means that the stress is not determined only by the applied strain, but by the time history of applied strains. This is illustrated in *Figure 7.43(a)* which shows a simple strain excursion, and the stress associated with it.

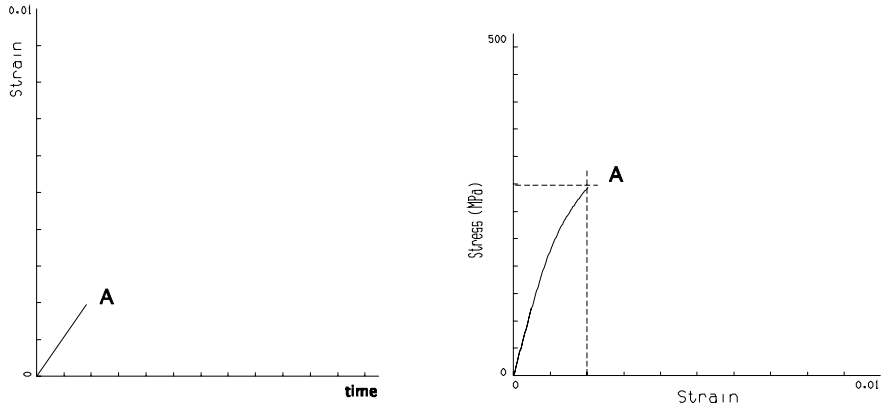


Figure 7.43(a) Strain excursion to point A (SAE1045 data)

Figure 7.43(b) shows the same applied strain, preceded by a larger tensile strain excursion.

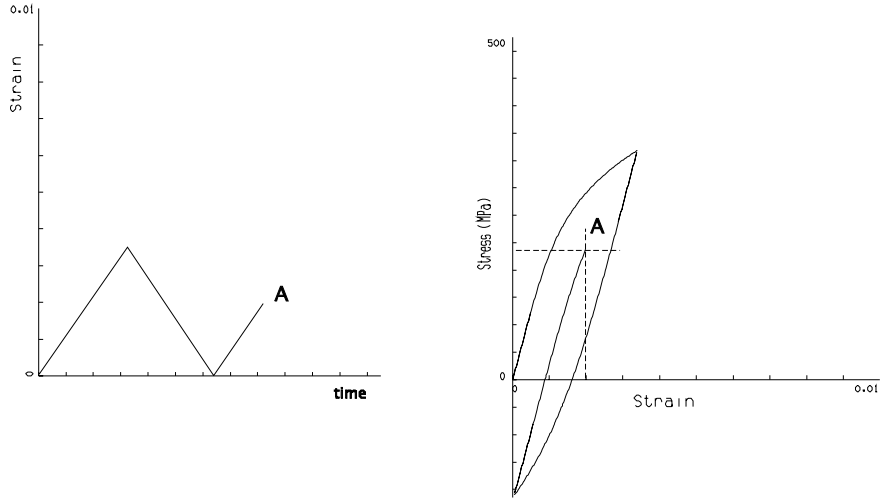


Figure 7.43(b) Strain excursion A preceded by a larger strain excursion (SAE1045 data)

In each case the final strain excursion is the same, but the associated peak stress is quite different. This sequence effect, or path dependency, is allowed for in uniaxial analysis, and must be included in any proposed multi-axial analysis.

Considering again the uniaxial stress condition, with a sequence of applied strains. The largest strains are associated with the largest hysteresis loop (the outside loop). *Figure 7.44* also shows the stress-strain response for the two smaller strain excursions AA and BB and it can be seen that they differ in two aspects: the amount of elastic behaviour, and the shapes of the stress-strain curves.

The amount of elastic behaviour is determined by how close the strain is to the yield point for the curve. In the case illustrated, this is how close the strain is to the outside loop. In general it is determined by how close the strain is to the previous position of the yield point.

It can also be seen from *Figure 7.44* that the stress-strain curve for BB is much flatter than that for AA, BB is closer to the maximum strain produced by the strain history. If the maximum strain is a limit, then the shape of the stress-strain curve is determined by its distance from the limit. For uniaxial conditions there are two limit points, which are the maximum and minimum points in the signal, and correspond to the outside loop. For a strain excursion such as AA or BB, the shape of an up-going stress-strain curve is determined by its distance from the maximum limit, and the shape of a down-going stress-strain curve is determined by its distance from the minimum limit.

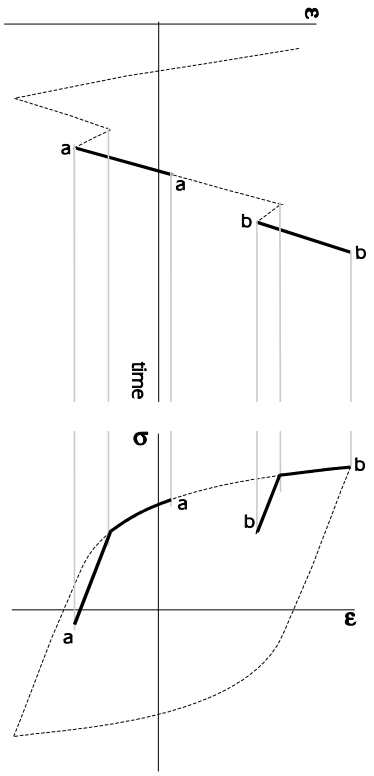


Figure 7.44 Uniaxial stress-strain response

These concepts of path dependency and of stress-strain curves whose shape is governed by their proximity to the limit conditions need to be translated into their multiaxial equivalents.
From Section 7.2.2, for elastic behaviour:

$$\varepsilon_x = \frac{1}{E}[\,\sigma_x - \nu(\sigma_y + \sigma_z)\,]$$

$$\varepsilon_y = \frac{1}{E}[\,\sigma_y - \nu(\sigma_z + \sigma_x)\,]$$

$$\varepsilon_z = \frac{1}{E}[\,\sigma_z - \nu(\sigma_x + \sigma_y)\,]$$

$$\gamma_{yz} = \frac{\tau_{yz}}{2G}$$

$$\gamma_{zx} = \frac{\tau_{zx}}{2G}$$

$$\gamma_{xy} = \frac{\tau_{xy}}{2G}$$

The hydrostatic stress is $\sigma_m = \dfrac{\sigma_x + \sigma_y + \sigma_z}{3}$

and the equations can be re-written as

$$\varepsilon_x = \frac{1}{2G}(\sigma_x - \sigma_m) + \frac{(1 - 2n)}{E}\sigma_m$$

$$\varepsilon_y = \frac{1}{2G}(\sigma_y - \sigma_m) + \frac{(1 - 2n)}{E}\sigma_m$$

$$\varepsilon_z = \frac{1}{2G}(\sigma_z - \sigma_m) + \frac{(1 - 2n)}{E}\sigma_m$$

$$\gamma_{yz} = \frac{\tau_{yz}}{2G}$$

$$\gamma_{zx} = \frac{\tau_{zx}}{2G}$$

$$\gamma_{xy} = \frac{\tau_{xy}}{2G}$$

The terms $(\sigma_x - \sigma_m)$ etc. are the deviatoric stresses S_x etc. and as the deviatoric shear stresses and strains are equal to the total shear stresses and strains, the complete elastic stress-strain relationships can be written simply as

$$\varepsilon^e_{ij} = \frac{S_{ij}}{2G} + \frac{(1 - 2n)}{E}\sigma_m$$

For the plastic component of strain, the Reuss flow rule assumes that the plastic component of strain increment is proportional to the deviatoric and shear stresses:

$$\frac{d\varepsilon^p_x}{S_x} = \frac{d\varepsilon^p_y}{S_y} = \frac{d\varepsilon^p_z}{S_z} = \frac{d\gamma^p_{xy}}{S_{xy}} = \frac{d\gamma^p_{yz}}{S_{yz}} = \frac{d\gamma^p_{xz}}{S_{xz}} = d\lambda$$

or $$\frac{d\varepsilon^p_{ij}}{S_{ij}} = d\lambda \qquad \text{or} \qquad d\varepsilon^p_{ij} = d\lambda\ S_{ij}$$

This equation can only be used to calculate *increments* of plastic strain, because the factor $d\lambda$ which relates strain to deviatoric stress will change with stress. Total plastic strain is obtained by integrating the equation along the strain path.

Isotropic hardening

Most metals strain harden under cyclic yielding, with the new yield stress being a function of the applied strain history. One theory is that the von Mises yield surface will *expand* to allow for this change. This is *isotropic hardening,* and can be described by a modified von Mises equation, i.e.

$$S_{ij}\,S_{ij} = \frac{2}{3}\sigma^2_y \quad \text{becomes} \qquad S_{ij}\,S_{ij} = \frac{2}{3}\underline{\sigma}^2$$

where $\underline{\sigma}$ is the largest value of stress reached in the load path. In uniaxial terms, this is illustrated in Fig 7.45.

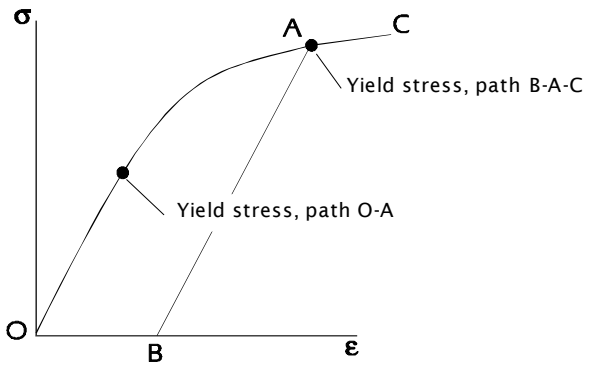


Figure 7.45 Increase in yield stress as a result of yielding on a preceding load path

On the first loading, from O to A, the material yields at the yield stress σ_y. If the load is removed, from A to B, and then re-applied from B to C, the material yields at point A, the previous maximum stress. Point A is the new yield stress $\underline{\sigma}$. Isotropic hardening proposes that the von Mises yield surface expands to include this point.

Kinematic hardening

An alternative is to assume that the yield surface does not change in size, but translates to intersect the new yield stress. This is kinematic hardening, and is expressed by

$$(S_{ij} - S^c_{ij})(S_{ij} - S^c_{ij}) = \frac{2}{3}\sigma^2_y$$

where S^c_{ij} is the centre of the yield surface.

Working with deviatoric stresses, the yield point becomes a yield surface, analogous to the von Mises yield criterion, and the two limit points in the uniaxial case become a limit surface. The shape of the stress-strain curve is determined by the closeness of the yield surface to the limit surface. The amount of elastic behaviour is determined by how close the start point is to the yield surface.

To calculate if yielding has occurred at any point in a strain history, the von Mises relationship can be used. This states that yielding occurs if

$$\frac{1}{2}\left[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right] = (\underline{\sigma})^2$$

stress allowing for where σ_1, σ_2, σ_3 are the principal stresses and $\underline{\sigma}$ is the current yield hardening.

In terms of deviatoric stresses, the von Mises relationship is

$$\frac{3}{2}\left[S_1^2 + S_2^2 + S_3^2\right] = (\underline{\sigma})^2$$

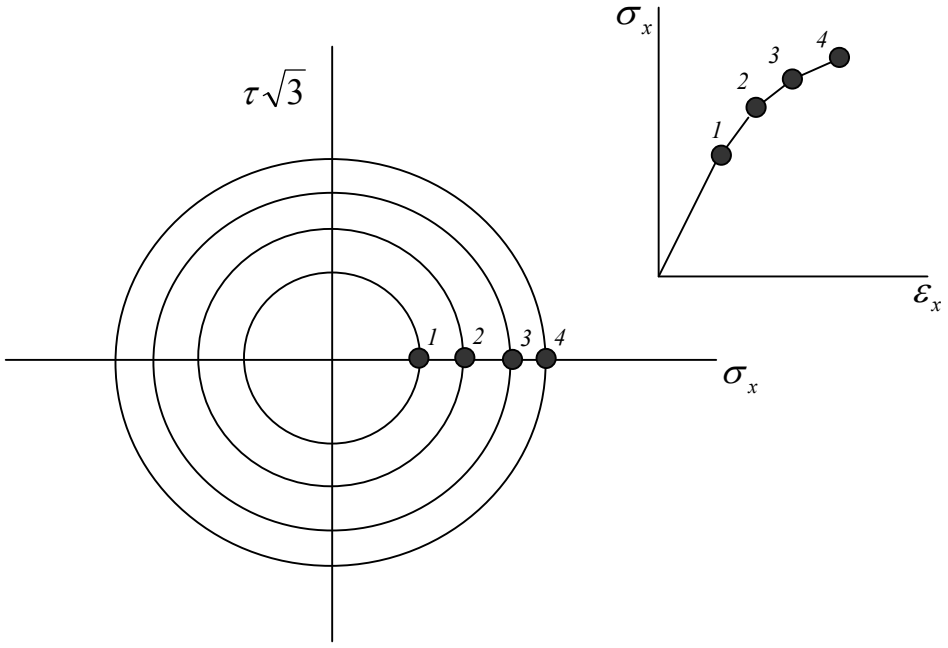
where S_1, S_2, S_3 are the deviatoric stresses.

If the combination of the three deviatoric stresses at any instant in a measured strain history exceeds $(\underline{\sigma})^2$ then the material has yielded. The incremental plasticity rule can be used to calculate the stresses, and the hardening rule used to position the yield surface. The total strain is then used in the fatigue life calculation.

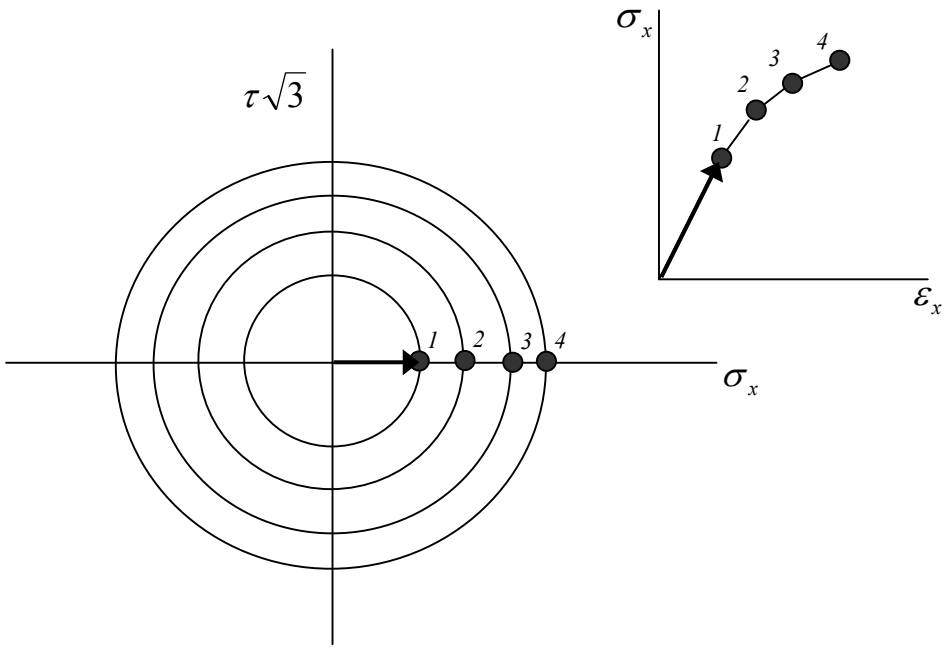
Mroz (Ref 7.11) proposed that multiaxial kinematic hardening could be simulated by using a series of yield surfaces, each representing a different amount of yielding. The behaviour of the Mroz model is illustrated below.

Consider the simplified case of a single axial stress σ_x, and a torsion stress τ. For the axial stress alone, yielding will occur at the uniaxial yield stress σ_y. For the torsion stress, the von Mises equation predicts that yielding will occur at a stress equal to $\sigma_y / \sqrt{3}$

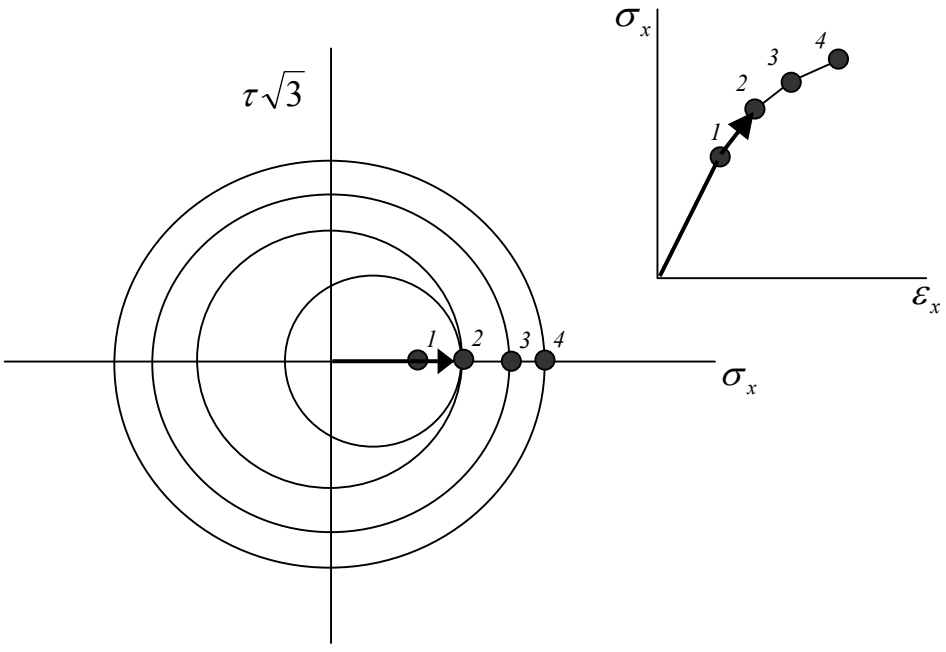
On a stress space with axes of σ_x and $\tau\sqrt{3}$ yielding under the axial stress alone, and the torsion stress alone, will then occur at the same numerical value. A series of concentric circles in this stress space will represent different amounts of yielding.



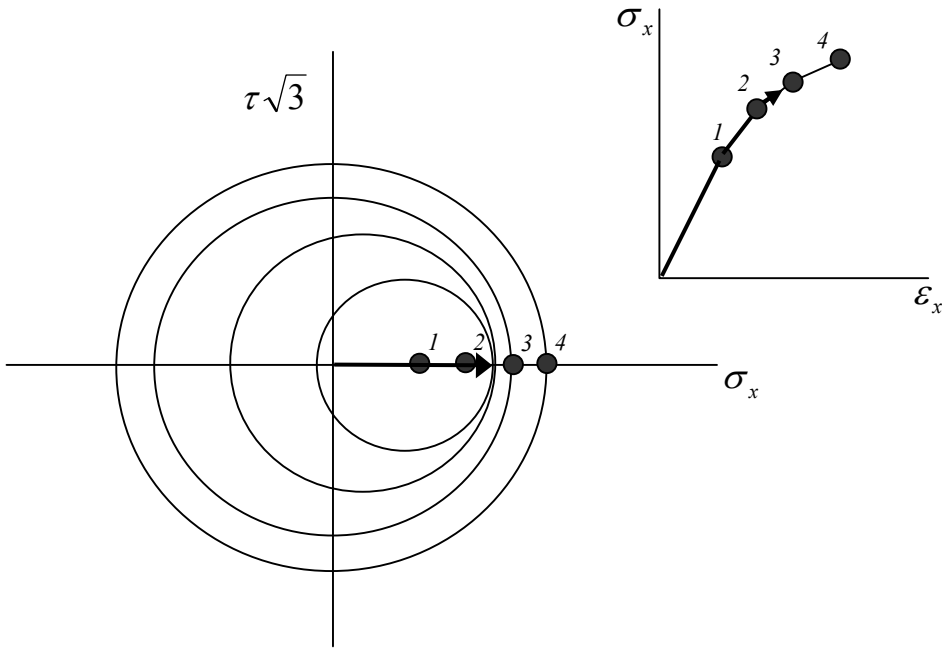
A stress in the direction of σ_x will be elastic up to point *1*.



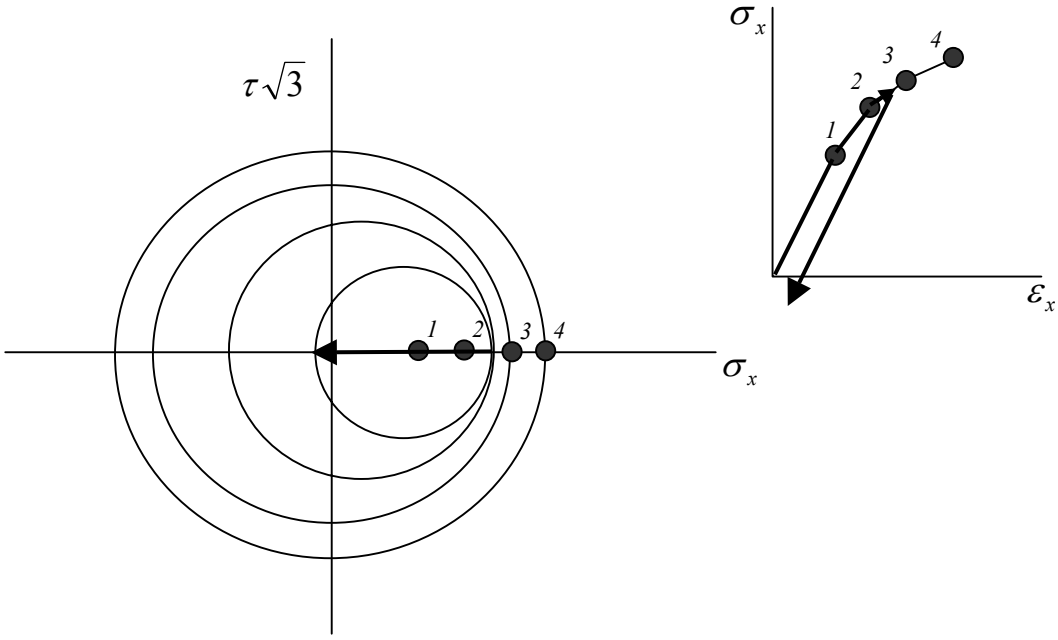
Increasing σ_x will move the inner yield surface in the direction of σ_x. This represents a certain amount of yielding.



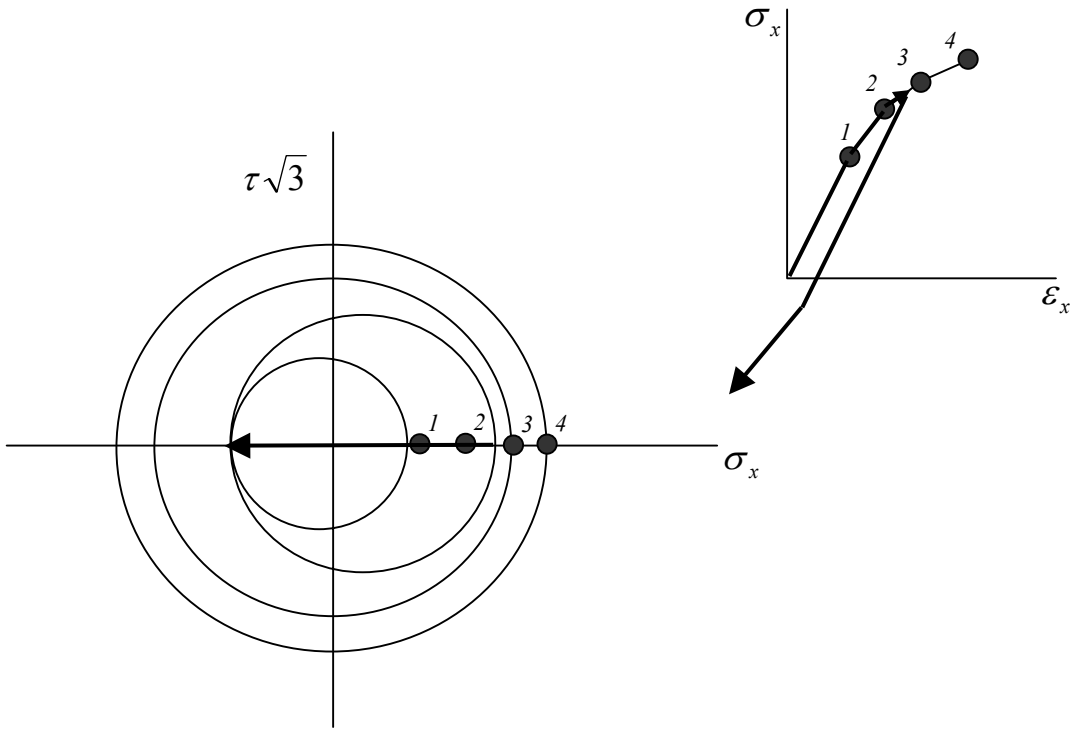
Further increasing σ_x will translate both the inner surfaces in the direction of σ_x. This represents further yielding.



If σ_x is now reversed, the response is elastic within the inner surface. The elastic part of the stress-strain curve is therefore twice as large as the initial event. This is consistent with the material behaviour described in Section 2.



Further increasing σ_x in this direction will cause both the inner surfaces to translate. Again, surface 2 can move through twice the distance before it reaches surface 3, simulating the Masing behaviour described in Section 2.



This simple description of Mroz kinematic hardening can be extended to include a full stress tensor of three direct and three shear stresses. It may also be extended to include a continuous stress-strain response, rather than the step-wise linearisation shown in the diagrams (see for example Ref 7.12).

The size of the inner yield surface must be chosen with care. It is very much smaller than the conventional 0.2% proof stress.

Lamba (Ref 7.13) developed a two-surface variant of the Mroz model. This uses only the inner and outer yield surfaces, with an algorithm to construct the stress-strain response.

Kinematic hardening is generally used for fatigue, because it is compatible with the material behaviour described for uniaxial fatigue in Section 2. It is almost certainly a simplification of more complex material behaviour.

Bannantine combined the Lamba model with critical plane analysis, using the direct-strain life equation (7.28) or shear-strain life equation (7.29). Correlation with test data (Ref. 7.9) is shown in Table 7.1. Signals were variable amplitude loading from measured service histories, and the test results are compared to life predictions made using a uniaxial Smith-Watson-Topper analysis. The flat specimen had simple bending, and in-phase bending and torsion loading. The notched shaft had end load and torsion, with random phase relationships. All lives are repeats of the test signal.

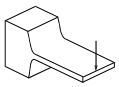
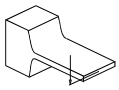
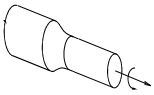
	TEST LIFE	PREDICTION	
		MULTIAXIAL	UNIAXIAL
	600	770	5000
	200	110	450
	4000	12000	53000
	1700	7000	30000
	1000	3000	19000

Table 7.1 Results from a multiaxial fatigue test programme from (Ref. 7.9)

7.7 Wang-Brown criterion

One objection to critical plane analyses is that, in resolving all stresses or strains onto a plane, it is implied that crack initiation will take place along a single plane. In fact, crack initiation may take place incrementally on several planes, depending on the instantaneous direction of the principal strains. Critical plane analyses should therefore tend to be non-conservative. The Wang-Brown criterion does not require a critical plane procedure.

The Wang-Brown criterion is:

$$\frac{\Delta\gamma_{max}}{2} + S\varepsilon_n^* = C_1 \frac{\sigma'_f}{E}(2N_f)^b + C_2\, \varepsilon'_f(2N_f)^c \qquad (7.30)$$

where $\frac{\Delta\gamma_{max}}{2}$ = the maximum shear strain amplitude

ε_n^* = the maximum normal strain excursion

$C1 = 1 + \nu_e + (1 - \nu_e)S$

and $C1 = 1 + \nu_p + (1 - \nu_p)S$

For proportional loading this equation is the same as the Brown-Miller parameter. Because it treats fatigue damage as a scalar quantity it requires additional rules to identify fatigue cycles for random loading.. This requires the calculation of a pseudo-stress, and a new multiaxial cycle counting method to determine the effective value of ε_n^* to be associated with the shear strain amplitude (Ref. 7.7).

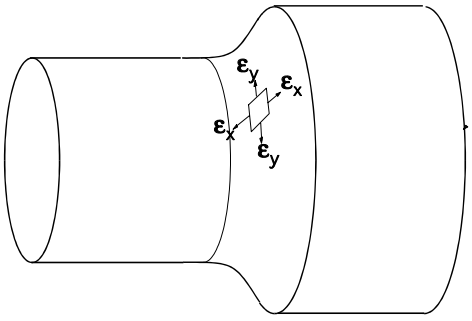
The parameter S is an extra materials property which represents the influence of the normal strain on fatigue crack initiation, and can be obtained from axial and torsion fatigue tests. For En15R steel, S = 1.38. This requirement for an extra materials property is a disadvantage in applying the Wang-Brown method, and in general the additional materials data does not seem to offer additional accuracy when compared with the conventional Brown-Miller parameter.

7.8 Comparison of strain-based criteria

7.8.1 Notched shaft in bending

Two of the criteria described above will be compared using as an example a notched shaft in bending.

The bending strain in the shaft will produce an axial strain ε_x at the notch, and a circumferential strain ε_y. The magnitude of ε_y will depend on the absolute size of the component, the sharpness of the fillet radius between the two shaft diameters, and amount of yielding at the notch. For a small component with a low value of stress concentration, and at low values of ε_x with little yielding, ε_y will have a value close to $-\nu\varepsilon_x$. For the opposite condition, with a large component, a sharp notch and large values of ε_x then there will be suppression of the circumferential strains, and ε_y may be close to zero.



If there is no torsion applied to the shaft, ε_x and ε_y are principal strains ε_1 and ε_2 .

The maximum shear criterion and Brown-Miller criterion can be compared using both $\varepsilon_2 = -\nu\varepsilon_1$ and $\varepsilon_2 = 0$. The comparison can be made at both high and low endurance, using the appropriate section of the full strain-life equation.

The equations are summarised:

	HIGH CYCLE	LOW CYCLE
Uniaxial	$\frac{\Delta\varepsilon}{2} = \frac{\sigma_f'}{E}(2N_f)^b$	$\frac{\Delta\varepsilon}{2} = \varepsilon_f'(2N_f)^c$
Max shear	$\frac{\Delta\gamma_{max}}{2} = 1.3\frac{\sigma_f'}{E}(2N_f)^b$	$\frac{\Delta\gamma_{max}}{2} = 1.5\,\varepsilon_f'(2N_f)^c$
Brown-Miller	$\frac{\Delta\gamma_{max}}{2}+\frac{\Delta\varepsilon_n}{2} = 1.65\frac{\sigma_f'}{E}(2N_f)^b$	$\frac{\Delta\gamma_{max}}{2}+\frac{\Delta\varepsilon_n}{2} = 1.75\,\varepsilon_f'(2N_f)^c$

If there is no suppression of circumferential strains

$$\varepsilon_2 = -\nu\varepsilon_1 \quad \varepsilon_3 = -\nu\varepsilon_1$$

If the circumferential strains are fully suppressed

$$\varepsilon_2 = 0 \qquad \varepsilon_3 = \frac{-\gamma}{1-\gamma}(\varepsilon_1 - \varepsilon_2) = \frac{-\gamma}{1-\gamma}\varepsilon_1$$

As $\gamma_{max} = \varepsilon_1 - \varepsilon_3$ and $\varepsilon_n = \frac{\varepsilon_1 + \varepsilon_3}{2}$

then these two sets of equations can be combined to give the following table

FORMULA FOR γ_{max} AND ε_n
FOR THE TWO CIRCUMFERENTIAL STRAIN CONDITIONS

	$\varepsilon_2 = -\nu\varepsilon_1$	$\varepsilon_2 = 0$
γ_{max}	$(1+\nu)\varepsilon_1$	$\frac{1-\nu}{2}\varepsilon_1$
ε_n	$\frac{1}{1-\nu}\varepsilon_1$	$\frac{1}{2}\left(\frac{1-2\nu}{1-\nu}\right)\varepsilon_1$

Taking $\nu = 0.3$ for the high cycle case, and $\nu = 0.5$ for the low cycle case:

VALUES OF γ_{max} AND ε_n
FOR THE TWO CIRCUMFERENTIAL STRAIN CONDITIONS

	HIGH CYCLE ($\nu = 0.3$)		LOW CYCLE ($\nu = 0.5$)	
	$\varepsilon_2 = -\nu\varepsilon_1$	$\varepsilon_2 = 0$	$\varepsilon_2 = -\nu\varepsilon_1$	$\varepsilon_2 = 0$
γ_{max}	$1.3\varepsilon_1$	$1.43\varepsilon_1$	$1.5\varepsilon_1$	$2.0\varepsilon_1$
ε_n	$0.35\varepsilon_1$	$0.29\varepsilon_1$	$0.25\varepsilon_1$	$0\varepsilon_1$

Using materials properties for MANTEN steel, and the high cycle formulae for endurance, the uniaxial strain-life equation shows that a strain amplitude of $\pm 930\mu\varepsilon$ gives an endurance of $2N_f = 2 \times 10^7$ reversals.

If this strain amplitude is measured at the notch on the shaft, with the circumferential strains fully suppressed, the life estimates for the shaft would be:

LIFE ESTIMATES FOR THE NOTCHED SHAFT,
CIRCUMFERENTIAL STRAINS FULLY SUPPRESSED, HIGH CYCLE

CRITERION	ENDURANCE $2N_f$ REVERSALS	RATIO TO UNIAXIAL PREDICTION
Uniaxial	2×10^7	1
Max shear	7×10^6	1/2.85
Brown-Miller	8×10^6	1/2.5

Using the low cycle formulae for endurance, the uniaxial strain-life equation shows that a strain amplitude of $\pm 30000\mu\varepsilon$ gives an endurance of $2N_f = 100$ reversals. If this strain amplitude is measured at the notch on the shaft, with the circumferential strains fully suppressed, the life estimates for the shaft would be:

LIFE ESTIMATES FOR THE NOTCHED SHAFT,
CIRCUMFERENTIAL STRAINS FULLY SUPPRESSED, LOW CYCLE

CRITERION	ENDURANCE $2N_f$ REVERSALS	RATIO TO UNIAXIAL PREDICTION
Uniaxial	100	1
Max shear	54	1/1.85
Brown-Miller	75	1/1.33

There is evidence that these life estimates may be optimistic. Bannantine's test results (Ref. 7.9) on a machined specimen of SAE1045 steel, similar to that shown in *Figure 7.3*, were described earlier. Rosette strain measurement at the notch showed very small transverse strains (analogous to the fully suppressed circumferential strains in the example above). The test lives were about $^1/_8$ of the lives predicted by a uniaxial analysis.

7.8.2 Formulae for combined direct and shear strain

The following formulae may be useful for fatigue calculations for combined tension/torsion loading if the tension/torsion is in phase or 90^{o} out-of-phase. Taken from (Ref. 7.10).

WITH NO SUPPRESSION OF TRANSVERSE OR CIRCUMFERENTIAL STRAINS :

for in-phase loading

$$\gamma_{max} = \sqrt{\left[\varepsilon_x (1+\nu))^2 + (\gamma_{xy})^2 \right]}$$

for 90^{o} out-of-phase loading

$$\gamma_{max} = \text{greater of} \begin{cases} \varepsilon_x (1+\nu) \\ \gamma_{xy} \end{cases}$$

WITH TRANSVERSE OR CIRCUMFERENTIAL STRAINS FULLY SUPPRESSED

for in-phase loading

$$\gamma_{max} = \text{greater of} \begin{cases} \sqrt{\left[\varepsilon_x^{\,2} + (\gamma_{xy})^2 \right]} \\ \left(\frac{1+\nu}{1-\nu} \varepsilon_x + \sqrt{\left[\varepsilon_x^{\,2} + (\gamma_{xy})^2 \right]} \right) \end{cases}$$

for 90^{o} out-of-phase loading

$$\gamma_{max} = \text{greater of} \begin{cases} \gamma_{xy} \\ \frac{\varepsilon_x}{1-\nu} \end{cases}$$

In each case :

γ_{max} = max notch shear strain amplitude

ε_x = stabilised notch direct strain amplitude

γ_{xy} = stabilised notch torsional shear strain amplitude

For the Brown-Miller parameter the following relationships are also required:

WITH NO SUPPRESSION OF TRANSVERSE OR CIRCUMFERENTIAL STRAINS :

for in-phase loading

$$\varepsilon_n = \frac{1-\nu}{2} \varepsilon_x$$

for 90^{o} out-of-phase loading

$$\varepsilon_n = \text{greater of} \begin{cases} \frac{1}{2}\sqrt{\left[\varepsilon_x (1+\nu))^2 + (\gamma_{xy})^2 \right]} \\ \varepsilon_x \end{cases}$$

WITH TRANSVERSE OR CIRCUMFERENTIAL STRAINS FULLY SUPPRESSED

for in-phase loading

$$\varepsilon_n = \text{greater of} \begin{cases} \dfrac{\varepsilon_x}{2} \\ \dfrac{1}{4}\left(\dfrac{1-3\nu}{1-\nu}\,\varepsilon_x + \sqrt{\left[\varepsilon_x^{\,2} + (\gamma_{xy})^2\right]} \right) \end{cases}$$

for 90^o out-of-phase loading

$$\varepsilon_n = \text{greater of} \begin{cases} \varepsilon_x \\ \dfrac{1-2\nu}{2(1-\nu)}\,\varepsilon_x \end{cases}$$

7.9 Dang Van criterion

Dang Van has proposed a method for assessing high cycle fatigue for components subject to multiaxial stresses. The method does not calculate fatigue lives, but calculates whether a component will have Infinite life. It is therefore a pass/fail criterion.

On a microscopic scale, fatigue damage initiates in plastically deformed grains, and the stress state in individual grains will deform plastically even if the engineering stress is elastic. The restraining effect of adjacent grains will be to produce residual stresses in the plastically deformed grain. The Dang Van theory relates fatigue damage to the stress state in these grains.

The microscopic stress is related to the engineering stress by

$$\sigma_{ij}(P,t) = \Sigma_{ij}(P,t) + S_{ij}(P,t)$$

where

σ_{ij} = microscopic stress tensor

Σ_{ij} = macroscopic stress tensor

S_{ij} = residual stress tensor

and P,t are position and time co-ordinates

At long lives, applied stresses are low and cyclic loading causes the microscopic stresses to converge to a specific state and the residual stresses to stabilise (and become time-independent).

The Dang Van criterion can be expressed as

$$\tau = \alpha S = \tau o \quad \text{at a specific endurance (e.g. the fatigue limit)}$$

where

τ = local shear stress

S = hydrostatic stress

$\alpha, \tau o$ are material-specific constants

This equation is similar to a Haigh mean stress correction, i.e.

cyclic stress + mean stress = constant given at any given endurance

and the equation can be plotted *(Figure 7.46)*.

Two commonly available test cases are constant amplitude cyclic loading at zero mean stress (R = -1), and at zero minimum stress (R = 0).

For these cases

$$S = \Sigma_{max}/3$$

$$\tau = (\Sigma \text{ - mean stress})/2$$

These cases give two points on the Dang Van diagram which allow the strain line to be constructed. A second line is a mirror image on this one.

For complex biaxial stresses, the stabilising residual stress is calculated by using one repeat of the signal, using kinematic and isotropic hardening. The microscopic stress tensor is then the difference between the macroscopic and residual stress tensors. Its time history can be plotted on the Dang Van diagram, and if its locus is within the two straight lines, the component will have infinite life.

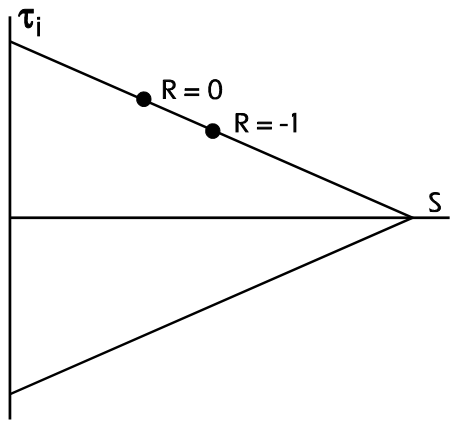


Figure 7.46 The Dang Van plot for the endurance limit

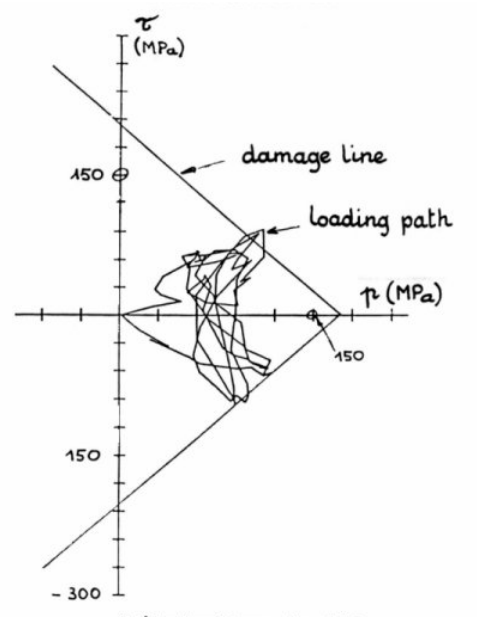


Figure 7.47 Result from a Dang Van analysis

The distance between the loading path and the Dang Van line can be interpreted as a stress-based factor of strength. In *Figure 7.48*, the ratio *b/a* is the factor of safety. Alternatively, this ratio may be calculated using a vertical line. Factors of safety less than 1, i.e. where the locus of the 'Dang Van' tensor crosses the infinite life line, are not reliable. This is because the Dang Van criteria applies to infinite life, and non-infinite life implies yielding on a macro scale, which invalidates the Dang Van criterion.

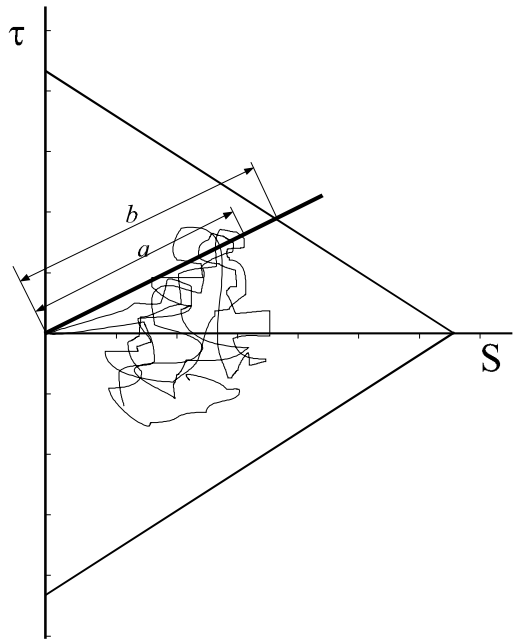


Figure 7.48 Factor of safety from the Dang Van analysis.

7.10 Summary of fatigue-life relationships

The following biaxial fatigue life relationships have been described. The von Mises criteria are not successful.

Uniaxial strain-life. Used only for uniaxial stresses, and may be very non-conservative for biaxial stresses. Requires only uniaxial materials data.

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma_f'}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c \qquad (7.17)$$

Principal strain (or axial strain). Used for brittle metals. May be very non-conservative for ductile metals. Requires only uniaxial materials data.

$$\frac{\Delta\varepsilon_1}{2} = \frac{\sigma_f'}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c \qquad (7.19)$$

Maximum shear strain. Used for ductile metals, where it tends to be conservative. Requires only uniaxial materials data

$$\frac{\Delta\gamma_{max}}{2} = 1.3\frac{\sigma_f'}{E}(2N_f)^b + 1.5\,\varepsilon_f'(2N_f)^c \qquad (7.21)$$

von Mises strain. Not recommended.

$$\frac{\Delta\varepsilon_{eff}}{2} = \frac{\sigma_f'}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c \qquad (7.27)$$

McDiarmid. Should be used only for very high cycle fatigue. Requires extra materials data.

$$\frac{\tau}{t} + \frac{\sigma_n}{2\,\sigma_T} = 1 \qquad (7.22)$$

Brown-Miller. Gives realistic life estimates for ductile metals. May be non-conservative for brittle metals. Requires only uniaxial materials data.

$$\frac{\Delta\gamma_{max}}{2} + \frac{\Delta\varepsilon_n}{2} = 1.65\,\frac{\sigma_f'}{E}(2N_f)^b + 1.75\,\varepsilon_f'(2N_f)^c \qquad (7.23)$$

Socie-Bannantine. Should be used only for brittle metals.

$$\frac{\Delta\varepsilon_1}{2}\,\sigma_{1max} = \frac{\sigma_f'^2}{E}(2N_f)^{2b} + \varepsilon_f'\sigma_f'(2N_f)^{b+c} \qquad (7.28)$$

Fatemi-Socie. Proposed for ductile metals, but Brown-Miller is preferred. Requires additional materials data, although these parameters can be approximated from uniaxial data.

$$\frac{\Delta\gamma_{max}}{2}\left(1 + \frac{\sigma_n}{\sigma_y}\right) = \frac{\tau_f'}{G}(2N_f)^b + \gamma_f'(2N_f)^c \qquad (7.29)$$

Wang-Brown. Requires extra materials data, but does not seem to offer improved accuracy over Brown-Miller.

$$\frac{\Delta\gamma_{max}}{2} + S\varepsilon_n^* = C_1\,\frac{\sigma_f'}{E}(2N_f)^b + C_2\,\varepsilon_f'(2N_f)^c \qquad (7.30)$$

Dang Van. Requires extra materials data, but is very successful for infinite-life design.

Principal stress (or axial stress) with S-N curves can be a successful method for brittle metals in high-cycle fatigue.

In summary, there are three criteria which may be recommended.

Brown-Miller, with mean stress corrections, for ductile metals.

Principal (or axial) strain, with mean stress corrections, for brittle metals.

Dang Van for infinite life design.

7.11 References

7.1 Devlukia J, Davies J

Fatigue Analysis Of A Vehicle Structural Component Under Biaxial Loading

Biaxial Fatigue Conference, Sheffield University, Dec 1985

7.2 Bannantine J A, Socie D F

A Variable Amplitude Multiaxial Fatigue Life Prediction Method

Fatigue under Biaxial and Multiaxial Loading, Proc. Third International Conference on Biaxial/Multiaxial Fatigue, Stuttgart, 1989.

EISI Publication 10, MEP, London.

7.3 McDiarmid D L
A General Criterion For High Cycle Multiaxial Fatigue
Fatigue Fract Engng Mater Struct Vol 14, No 4, pp429-453, 1991

7.4 Kandil F A, Brown M W, Miller K J
Biaxial Low Cycle Fatigue Fracture Of 316 Stainless Steel At Elevated Temperatures
Book 280, The Metals Society, London, 1982

7.5 Brown M W, Miller K J
A Theory Of Fatigue Under Multiaxial Strain Conditions
Proc Inst Mech Eng 187, pp745-755 (1973)

7.6 Miller K J
Fatigue Under Complex Stress
Metal Science Aug-Sept 1977, pp482-488

7.7 Wang C, Brown M W
A Path-Independent Parameter For Fatigue Under Proportional And Non-Proportional Loading
Fatigue Fract Engng Mater Struct Vol 16 No12 pp1285-1298, 1993

7.8 Lipson C, Juvenal R C
Handbook Of Stress And Strength - Design And Material Application
published by MacMillan, 1963

7.9 Bannantine J A
A Variable Amplitude Multiaxial Fatigue Life Prediction Method
PhD Thesis, University of Illinois at Urbana-Champaign, 1989

7.10 Tipton S M, Fash J W
Multiaxial Fatigue Life Predictions For The SAE Specimen Using Strain Based Approaches.
Multiaxial Fatigue: Analysis and Experiments, SAE AE-14, 1989

7.11 Mroz Z.
On the Description of Anisotropic Work Hardening.
J.Mechanics of Physics and Solids, Vol 15 1967 pp 163-175

7.12 Chu C C.
The Analysis of MultiAxial Cyclic Problems with an Anisotropic Hardeninig Model.
Int J Solids and Structures, Vol 23 No 5 1987 pp 569-579

7.13 Lamba H S.
Non-proportional Cyclic Plasticity.
PhD Thesis, T and A M Report No 413, University of Illinois at Urbana-Champaign

7.14 Chu C C, Conle F A and Bonnen J F
Multiaxial Stress-Strain Modelling and Fatigue Life Prediction of SAE Axle Shafts.
American Society for Testing and Materials, ASTM STP 1191, 1993 pp 37-54

Reference 7.10 provides a very readable introduction to multiaxial fatigue. For further reading, the European Structural Integrity Society publishes proceedings from the International Conference on Biaxial/Multiaxial Fatigue which is held every two years.

For a comprehensive treatment of multiaxial fatigue, the following book is recommended

Socie D F, Marquis G B.
Multiaxial Fatigue
Society of Automotive Engineers, 2000. ISBN 0-7680-0453-5

8 Fatigue analysis from Finite Element models

8.1 Introduction

Finite element models are used to analyse engineering components. The results may include stresses, strains, temperatures, deflections and frequency response. These results can be used for fatigue analysis.

Many fatigue analyses use the stress results from a linear elastic FEA. A linear analysis assumes a linear relationship between load and response. An elastic analysis assumes that the material stress-strain behaviour is elastic, with no allowance for yielding. The reasons for using linear elastic FEA, and the fatigue analysis methods, are described in this chapter.

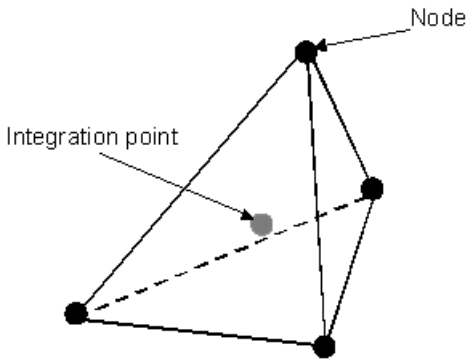
For some components it may be necessary to use a non-linear FE analysis, or an inelastic analysis. Some examples are discussed in this chapter.

8.2 Terminology of Finite Element analysis

There are many different element formulations, and no general description will apply to all elements. This chapter is not a finite element primer, but some explanation of the terms used in FE analysis is useful. The following information is a general guide.

Three basic types of element are used for stress analysis - shell elements, tetrahedral elements and hexagonal (or brick) elements. Tetrahedral and hexagonal elements are solid three-dimensional elements. Engineering components produced by casting, forging and machining may be modelled using solid elements. Shell elements are thin, and are used to model thin-walled structures fabricated from sheet and plate, for example. The different types of elements may be mixed in the same model. A fourth element, a beam element, does not give sufficiently detailed stress information to be used for fatigue analysis.

Figure 8.1 shows a tetrahedral element with a node at each corner. The separate elements in a finite element model will usually be linked together at the nodes. The stresses in the element may be calculated at one or more points inside the element, called *integration points*, or *Gauss points*. Nodal stresses are calculated by extrapolating the internal integration point stresses to the nodes of the element. The user may select to write integration point stresses and/or nodal stresses to the FEA results file.



Figue 8.1 4-node tetrahedral element

Surface stresses are required for the analysis of fatigue crack initiation from the surface of a component. There will always be nodes at the surface of the FE model and so nodal stresses should be used for the fatigue calculation rather than integration point stresses. It is usually possible to output strains as well as stresses from the FE analysis and the strains could also be used for fatigue analysis. The accuracy with which surface nodal stresses are extrapolated from integration point stresses may have a significant influence on the subsequent fatigue analysis.

Since elements are joined at nodes, each node may have several values of stress, extrapolated from the adjacent elements. It is normal for FE codes and post-processing software to average these extrapolated stresses, giving a single '*nodal average*' stress tensor at each node. In *Figure 8.2* each of the four brick elements would be used to calculate the stress tensor for the central node 'A' and the four values would be averaged to give the nodal averaged stress tensor.

The difference between the averaged and un-averaged stresses at a node is an indication of the quality of the finite element mesh, a large difference indicating an inadequate mesh. Although it is difficult to generalise, differences of more than perhaps 15% in the un-averaged stresses at a particular node could indicate an inadequate mesh.

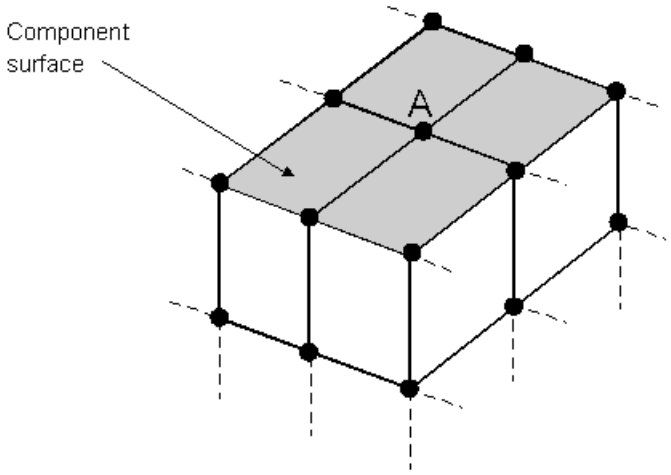


Figure 8.2 Node 'A' is common to four brick elements

When nodal stresses are displayed as contour plots, the graphics software may perform some averaging and interpolation to produce the contours. The user can control how much averaging is used. If 'third-party' plotting software is being used, the algorithms used to calculate the averaged stresses may be different from those used in the FEA software. The plotting software may also allow contours of surface stress to be plotted from integration point data. Again, third-party graphics software may use different algorithms from those used in the FEA software to calculate surface stresses.

Fatigue analysis from FEA models is a new subject, and many of the ground rules have yet to be established. However, as a general guide, nodal stresses should be used in preference to integration point stresses. There is some indication that if fatigue lives are calculated from un-averaged nodal stresses and then plotted with some averaging of the lives, the results can be more realistic than if the lives are calculated from averaged nodal stresses.

The nodal stresses for solid elements are calculated as a full stress tensor, consisting of three direct stresses and three shear stresses, usually orientated on the global axis system for the model. Standard equations can be used to resolve the stresses into the plane of the surface of the element, to produce the two in-plane direct stresses and the in-plane shear stress. The out-of-plane direct stress will also be calculated. This should be approximately zero on the surface of the component. A significant out-of-plane stress is an indication of inadequate meshing.

The stress tensor for a shell element will be the two in-plane direct stresses and the in-plane shear stress.

For linear elastic analysis the finite element model may produce one result for the stress tensor at each node. The terminology varies between FEA software vendors, but this set of stresses may be call a '*step*' or a '*data set*'. The stresses are calculated for the specified value of load.

For some analyses, such as a non-linear analysis or a thermal analysis, the finite element analysis may apply the load in a number of increments in order to calculate the final stresses. These intermediate stress results may be called the '*increments*' within the step. In many analyses, only the stresses for the final increment are of interest. In other instances, maximum stresses may be produced by intermediate increments, or the sequence of stresses calculated in the various increments in the step may produce fatigue cycles. In the FEA software the user can control whether each increment or only the final increment is written to the results file, and the fatigue analysis software may also allow the user to select all increments or only the final increment.

8.3 Analysing a linear elastic model with single applied load history

The finite element load case will consist of a linear elastic FEA solution for the stresses at each node, calculated for a single applied load – most conveniently a unit load. These results will be written to the FEA results file as a step.

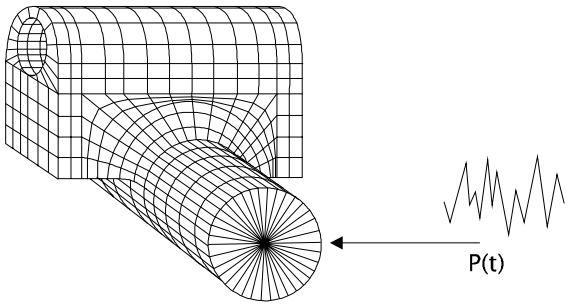


Figure 8.3 Single load history applied to the FE model

At each node, the elastically-calculated stress tensor is multiplied by the load history to give a time history of the stress tensor.

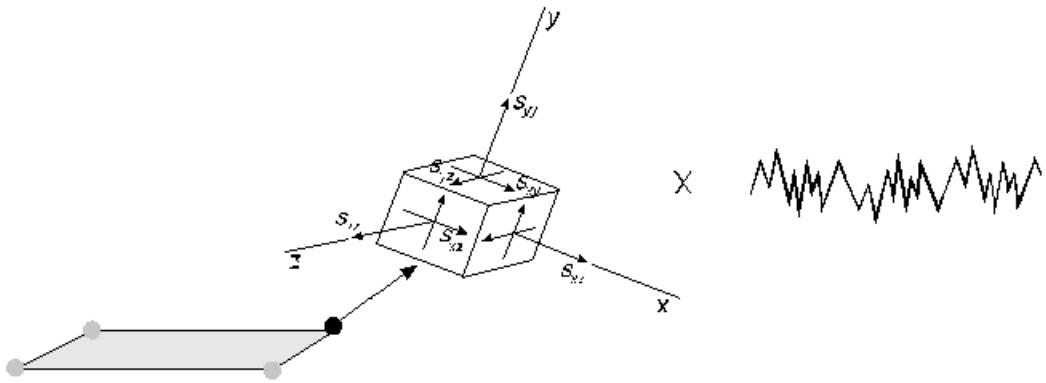


Figure 8.4 Multiplying the unit load stress tensor by the time history of loading

If $S_{i,j}$ is one component of the stress tensor from the unit load linear elastic FEA, and
$P(t)$ is the time history of the loading
then the time history of the stress component $S_{i,j}(t)$ is

$$S_{i,j}(t) = (S_{i,j}) \times P(t)$$

On the surface of the model, fatigue software will calculate the time histories of the in-plane principal stresses, and their directions. The elastic strains can be calculated from the stresses.

If the stress tensor represents uniaxial stress, the time history of the principal stress can be converted into elastic-plastic stress-strains using Neuber's or Glinka's rule.

The strain-time history can be used in a strain-life fatigue calculation (Section 2.10) and the associated stresses can be used to apply a mean stress correction (Section 2.6). This procedure is repeated for each node on the model.

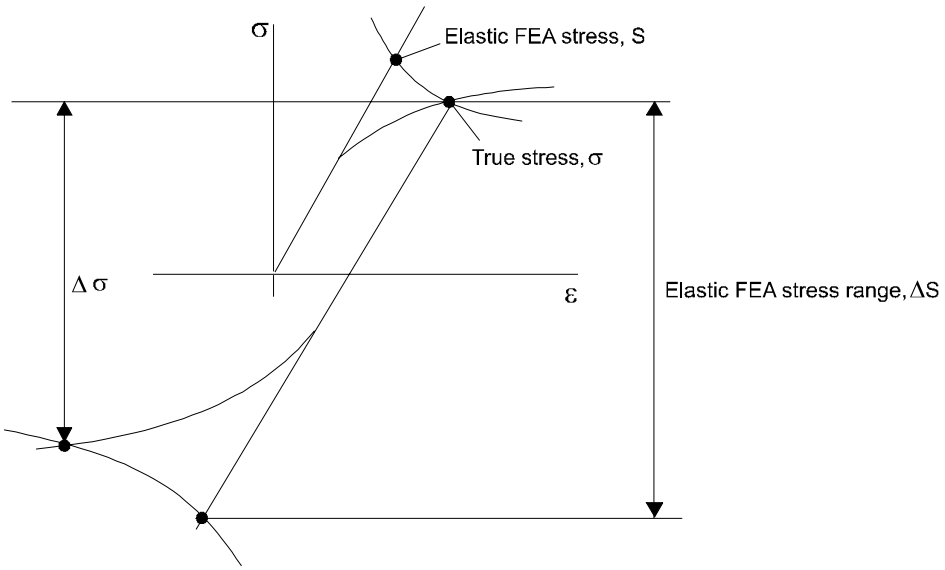


Figure 8.5 Neuber's rule for uniaxial stresses

If the nodal stresses are biaxial, Neuber's uniaxial elastic-plastic correction cannot be used. The are several proposed methods for calculating elastic-plastic stress/strain from elastic stress/strain for biaxial stress conditions when the direction of the principal stresses does not change during the loading history. Glinka's method (Ref 8.1) equates the total strain energy in the elastic and elastic-plastic conditions. It may be applied using Neuber's rule or the Glinka integration rule (Sections 2.91 and 2.92). Using Neuber's rule as an example, for elastically calculated principal stresses S_1 and S_2 and strains e_1 and e_2 the elastic-plastic principal stresses and strains are calculated using the relationship

$$\sigma_1 \varepsilon_1 + \sigma_2 \varepsilon_2 = S_1 e_1 + S_2 e_2$$

More complex equations are used for situations where the principal stresses change direction during fatigue loading.

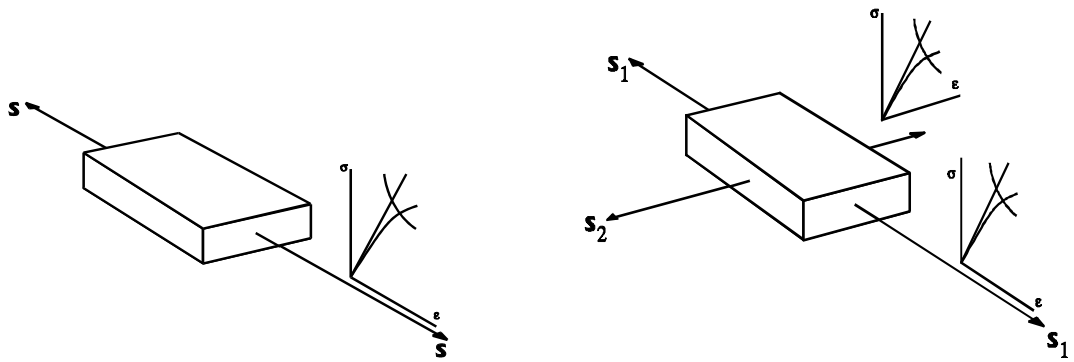


Figure 8.6 Uniaxial and biaxial Neuber's rule

On the surface of the component, the principal stress S_3 should be zero and so there is no strain energy in this direction.

The cyclic stress-strain curves are modified to allow for the effect of biaxial stresses, and the full set of simultaneous equations can be solved for each data point in the applied loading. *Figure 8.7* shows the effect of biaxial stress on the cyclic stress-strain curve. The calculated elastic-plastic strains and stresses can be analysed using the multiaxial fatigue algorithms from Chapter 7.

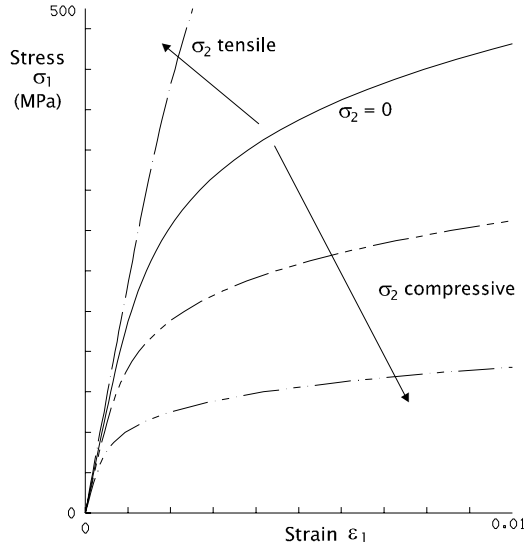


Figure 8.7 Calculated cyclic stress-strain curves for various biaxial stresses

Figure 8.8 shows the fatigue life contour plots for a simple shaft loaded in bending. Contour plots of fatigue life provide the clearest indication of fatigue lives and the location of crack initiation sites. Some graphics software allows the user to define a $\log_{10}$ scale to display the life contours. Graphics software may not be able to plot fatigue life contour plots on a linear scale because the contour algorithms cannot cope with the large range of numbers, which may encompass several orders of magnitude.

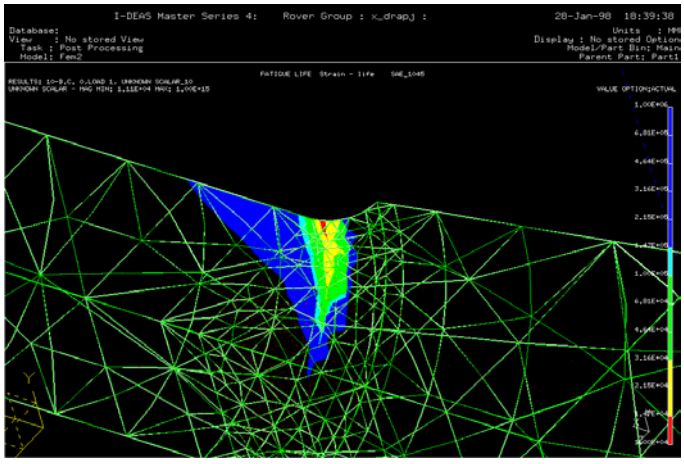


Figure 8.8 Fatigue life contours for a notched shaft in bending
(the calculated crack initiation site is shown in red)

Where the software does not allow the user to define a $\log_{10}$ scale, the fatigue analysis software may export the fatigue results as the $\log_{10}$ of fatigue lives and these may be plotted using a linear scale. Other types of output may be provided, and these are described in Section 8.7.

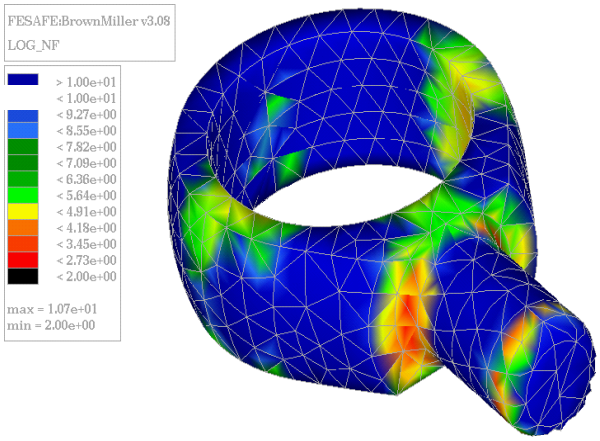


Figure 8.9 Fatigue life contours for an aircraft component plotted as $\log_{10}$ of fatigue life.

This section has described the process of calculating fatigue lives from a single linear elastic FEA solution. The calculation of each stress in the time history, and the conversion of the elastically calculated stresses to elastic-plastic stress-strain, has been performed by the fatigue analysis software. It would of course be possible to use an elasto-plastic FE analysis to calculate the stresses and strains for each point in the load history. This is not usually done because the processing time for long load histories would be prohibitive. However, the fatigue analysis is making two simplifications, and these should be clearly understood. Firstly, the elements in the FEA model have been put into equilibrium in terms of forces and deflections using elastically calculated stresses and strains. By allowing these stresses to yield in the fatigue software, this equilibrium is not maintained. Secondly, the Neuber and Glinka rules for calculating elastic-plastic stress-strain from elastic values only apply when yielding is localised and the bulk of the component is behaving elastically.

The effect of these simplifications is negligible for many components, and the majority of fatigue analyses are carried out from linear elastic models. However, where significant yielding occurs in large regions of a component it may be necessary to use an elastic-plastic FE analysis to calculate the stresses at each point in the loading. Caution is advised however. Chapter 2 introduced the concept of cycle closure and 'material memory' for stress histories. It is necessary to ensure that the plasticity models in the FEA software reproduce this behaviour. It is also necessary to use the cyclic materials properties described in Chapters 2 and 3, rather than the conventional static or monotonic properties which may be provided with the FEA software.

Contact pressure on the surface of the component may produce significant out-of-plane stresses. This may imply a much more complex fatigue analysis for triaxial stresses (as opposed to biaxial stresses). Fatigue algorithms for contact situations are not well defined. The contact condition may also mean that it is not valid to take a unit-load stress tensor and scale it by a load history, because the presence or absence of contact may be determined by the loading.

8.4 Analysing a linear elastic model with multiple applied loads

Components may have loads applied at several different points. As an example, the car suspension component *(Figure 8.10)* has three loads applied at the tyre contact patch. Each of these loads will have it's own time history.

The FE analysis is used to calculate the unit load stresses for each load applied separately. For the suspension component three separate stress solutions would be produced, for the vertical force applied on its own, the cornering force applied on its own, and the braking force applied on its own. These are three separate 'steps' .

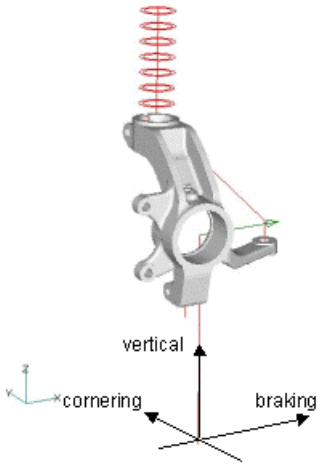


Figure 8.10 Suspension component with three applied load histories

Consider the component in *Figure 8.11*, with two applied load histories *P(t)* and *Q(t)*. The FEA results will have a set of stresses (a step) for a unit load P, and a set of stresses for a unit load Q.

If the elastically calculated stress tensor at a node is $S^P_{i,j}$ for unit load *P,* the stress tensor can be multiplied by the load-time history *P(t)* to produce the time history of the stress tensor, $S^P_{i,j}\ (t)$

i.e. $$S^P_{i,j}\ (t) = S^P_{i,j} \times P(t)$$

This is calculated for each of the six components of the unit load stress tensor, $S^P_{xx}, S^P_{yy}, S^P_{zz}, S^P_{xy}, S^P_{yz}, S^P_{xz}$ to produce six stress-time histories.

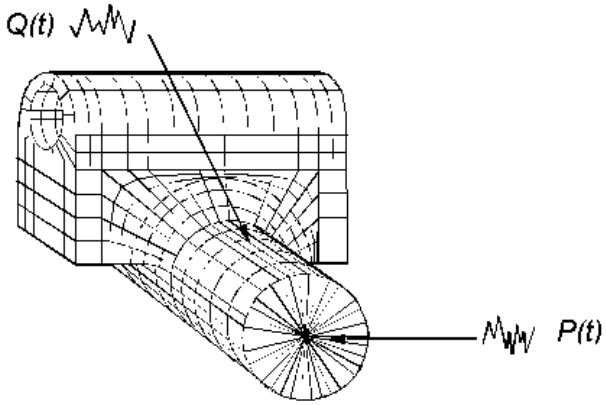


Figure 8.11 A component with two load histories

If the elastically calculated stress tensor at the same node for the unit load Q is $S^Q_{i,j}$, this stress tensor can be multiplied by the load-time history $Q(t)$ to produce the time history of the stress tensor, $S^Q_{i,j}\ (t)$

i.e. $$S^Q_{i,j}\ (t) = S^Q_{i,j} \times Q(t)$$

and again this is calculated for each of the six components of the stress tensor.

The stresses are then added together point by point to give the stress-time history for both load histories applied simultaneously

$$S_{i,j}(t) = S^{P}_{i,j}(t) + S^{Q}_{i,j}(t)$$

to produce six stress-time histories

$$S_{xx}(t) = S^{P}_{xx}(t) + S^{Q}_{xx}(t)$$

$$S_{yy}(t) = S^{P}_{yy}(t) + S^{Q}_{yy}(t) \quad \text{etc.}$$

The procedure is illustrated in Figure 8.12

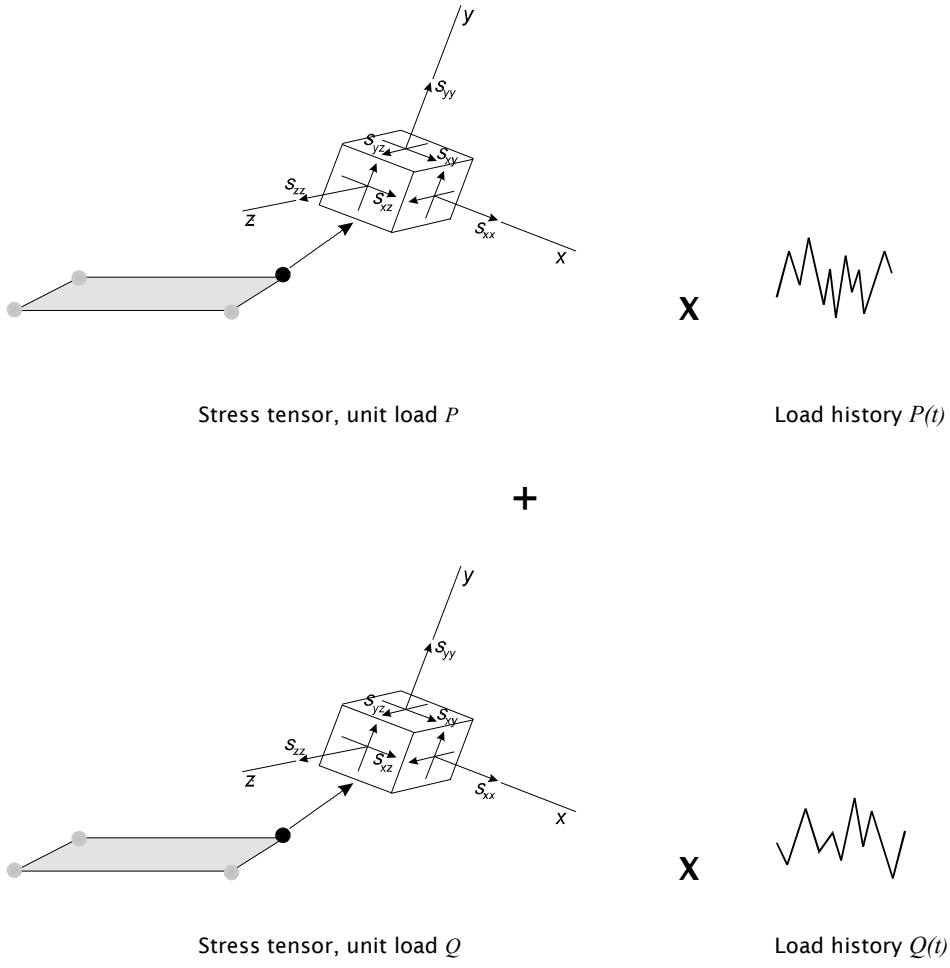


Figure 8.12 Superimposing stress tensors from two time histories

If the load histories are not in-phase, the time histories of the final stress tensor will not be in-phase, and so the direction of the principal stresses will not be constant.

The time history of elastic-plastic principal stresses and strains (magnitude and direction) can be calculated using the multiaxial Neuber's rules described in Section 8.3, and the fatigue life calculated. If the direction of principal stresses is not constant, a critical plane analysis must be carried out to determine the most damaged plane at each node (See section 7.5)

Figure 8.13 shows fatigue life contours for the suspension component in *Figure 8.10*. The analysis used three unit-load FE analyses, for braking, cornering and vertical forces at the tyre contact patch. In the fatigue analysis software each stress data set was multiplied by the corresponding load history and the stress results superimposed. Fatigue lives were calculated using a critical plane analysis.

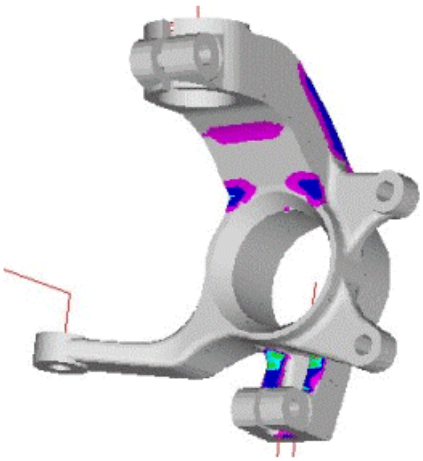


Figure 8.13 Fatigue life contours for a suspension component

Fatigue analysis software is capable of analysing components with more than one thousand load histories applied to a model. There is no in-built limit to the length of the histories.

8.5 Analysing a sequence of data sets

This capability is required because designers may model several discrete events in the duty cycle of a component, then chain the stress results in sequence to form stress-time histories. Transient FE analysis produces this type of data, and engine designers produce sequences of data sets to model the stresses produced by several revolutions of a crank-shaft. This type of analysis is not limited to linear FEA because the stresses are not scaled. An elastic or an elasto-plastic FEA could be used.

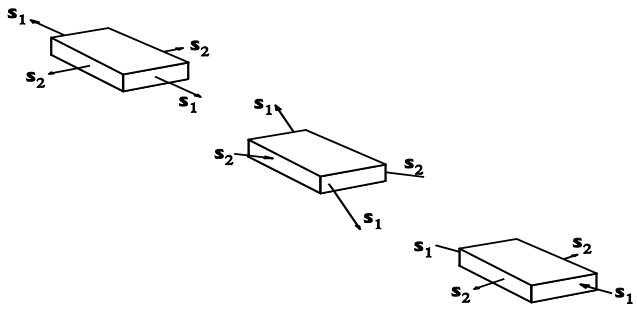


Figure 8.14 Principal stresses at a node from a sequence of three data sets

The elastic FEA stresses are converted to elastic-plastic stress/strain using a biaxial Neuber's rule allowing for non-proportional principal stresses. Fatigue lives are calculated using critical plane analysis. Sequences of several thousands of stress results can be analysed. The analysis of the crank-shaft shown in *Figure 8.15*, and many other analyses, has shown that contour plots of von Mises stresses do not indicate the fatigue crack initiation sites reliably. This is because von Mises stress is not an adequate parameter for fatigue analysis, and also because it may be difficult to select an appropriate condition for plotting when the loading is complex.

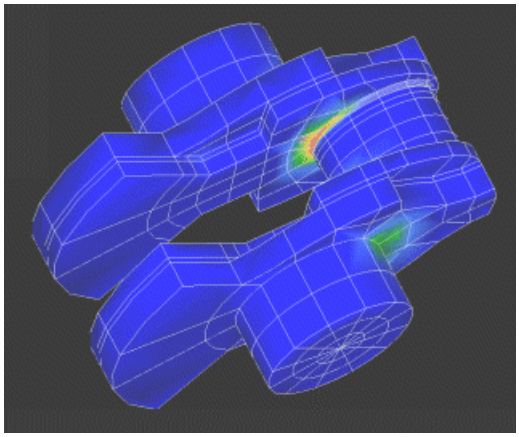


Figure 8.15 Fatigue life contours for a crank-shaft
(Boundary element model from BEASY, fatigue life contours plotted in HYPERMESH)

8.6 Other types of loading

The foregoing sections have described the main forms of fatigue analysis from finite element models. Modern software allows great flexibility in the definition of applied loads. For example, repeat counts can be applied to blocks of time histories, and different blocks of time histories (with different scale factors and repeat counts) can be specified. Sets of 'data set sequences' can be inserted in between these blocks. This could be used to create a simulation of a vehicle proving ground, for example.

If only one load is applied to a component, this may be defined by a Rainflow histogram of cycles. The fatigue analysis will then be based on the procedures described Section 2.11.

Finite element software provides powerful techniques for analysing components in the frequency domain (see Section 12). The techniques described in Section 12 can be used to calculate fatigue lives from frequency domain data.

8.7 Output

Fatigue lives, output as lives or $\log_{10}$ lives, can be displayed as contour plots. Where the designer can specify a design life for the component, the fatigue analysis software may calculate a 'fatigue strength reserve factor'. This is defined as the linear scale factor which, when applied to the elastically-calculated nodal stresses, gives a calculated life which is equal to the design life. The fatigue life is calculated by applying the scale factor to the elastically calculated nodal stresses, and recalculating the elastic-plastic stresses and strains. A contour plot of these scale factors, calculated at each node, shows how much the component is over-strength or under-strength in terms of the FEA stresses, for the specified design life.

The 'reserve factor' calculation may be combined with statistical information on variability in materials properties and loading to estimate the probability of failure at the design life. Calculating the probability of failure (or of survival) at several specified lives allows an estimate of the 'warranty claim' curve to be made *(Figure 8.16)*

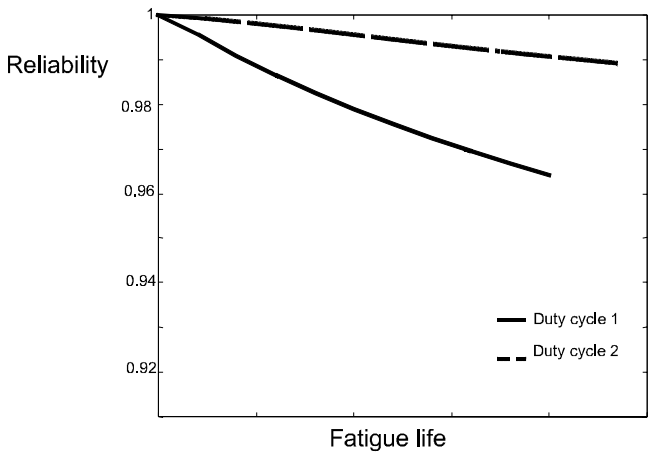


Figure 8.16 Effect of two different service duties on component survivability

Additional information may include the orientation of the critical plane, information on the amount of biaxiality in the stresses, the extent to which the principal stresses are changing orientation, and histograms of stress or strain cycles and fatigue damage.

8.8 Speed considerations

Finite element models may contain many thousands of elements, and the applied load histories may contain millions of data points. The processing time may then be unacceptably long.

8.8.1 Time history pre-processing

A multi-channel peak/valley program may be used to process the time histories to remove data points that do not form cycles. A hysteresis gate may be also be used to remove small fluctuations from the signal. Peak/valley extraction for a single time-history is described in Section 4.

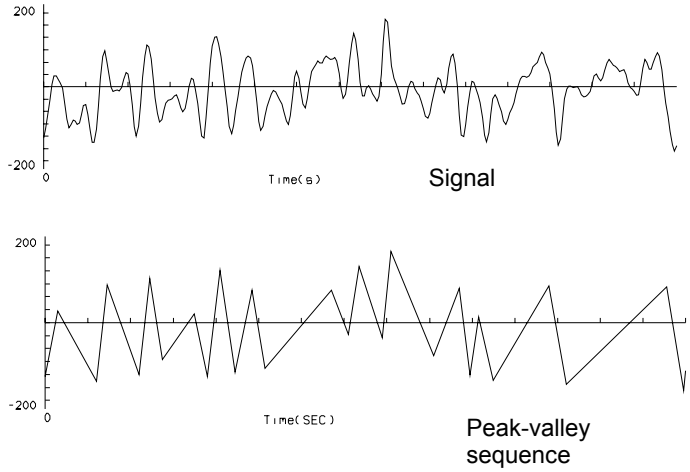


Figure 8.16 Single channel peak/valley extraction with small fluctuations omitted

For a multi-channel load history, the peak/valley extraction must maintain the phase relationship between the data points in the channels.

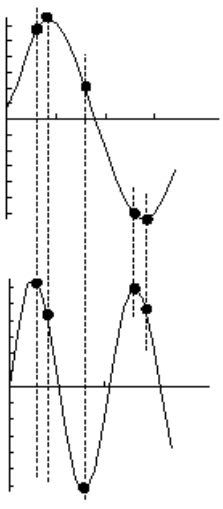


Figure 8.17 Multi-channel peak/valley extraction

When a peak/valley is found on one channel, the data points at the same time instant on the other channels are also saved. A hysteresis gate may be included, so that peaks and valleys formed from small fluctuations are not saved.

An example of the results of this pre-processing is shown in *Figure 8.18*. Three signals with random phase relationships were peak/valley picked with a hysteresis gate of 100. *Figure 8.18* shows the cycle density diagrams produced by Rainflow counting each signal. Most of the small cycles are removed. However, the data points that are retained to maintain the phase relationship between the channels can themselves form small cycles, so some cycles are retained, even though their amplitudes are less than the gate value.

This procedure introduces an approximation into the subsequent fatigue analysis, because it assumes that peak/valleys in the calculated 6-stress tensor at a node will be formed only by the data points produced by the peak/valley extraction. In fact, it is possible that the process of superimposing the stress tensors could produce peak/valleys from other data points in the signal. However, depending on the type of loading, the process may be be justified by the significant increase in processing speed.

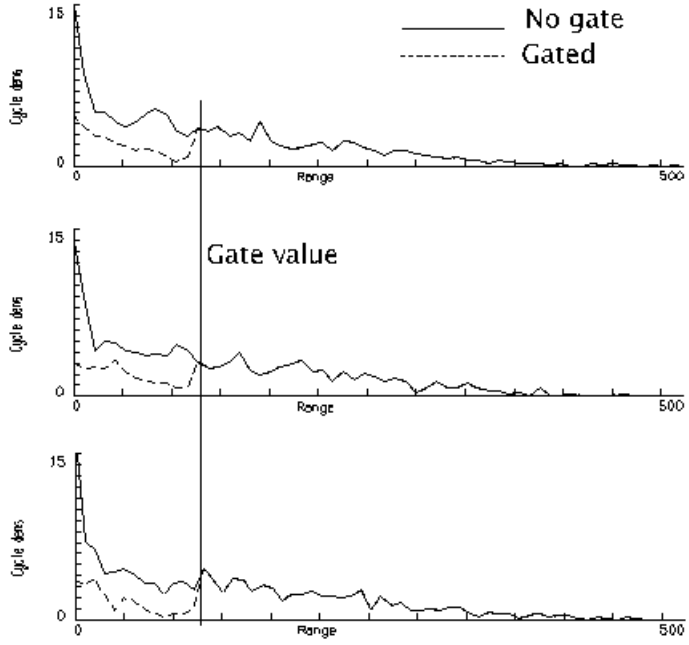


Figure 8.18 Cycle density diagrams for three channels of load history without gating and with a gate

8.8.2 Assessing multiaxiality

Uniaxial stresses will exist at many locations in an FE model. Analysis time can be reduced if these locations are analysed using the uniaxial Neuber's rule and a uniaxial fatigue life equation. Also, some locations, which have multiaxial stresses, may show little variation in the direction of the principal stresses, therefore a critical plane analysis will not be necessary at these nodes, even though the stresses are multiaxial. Fatigue analysis software may make the decision automatically.

Biaxiality ratios calculated from the variation in principal stress direction are computationally intensive (Ref. 8.2). Chu, Conle and Hubner (Ref. 8.3) have proposed a 3-dimensional cross-plot of the time histories of the two in-plane direct stresses and the in-plane shear stress for a surface node *(Figure 8.19)*. A regression analysis of this diagram gives a quantitative measure of the amount of biaxiality. For biaxially stressed nodes it also provides a description of the amount of variation in the direction of principal stress, and hence indicates when a critical plane analysis is necessary. For biaxially stressed nodes with little variation in principal stress direction, the best-fit line through the data gives the orientation of the critical plane.

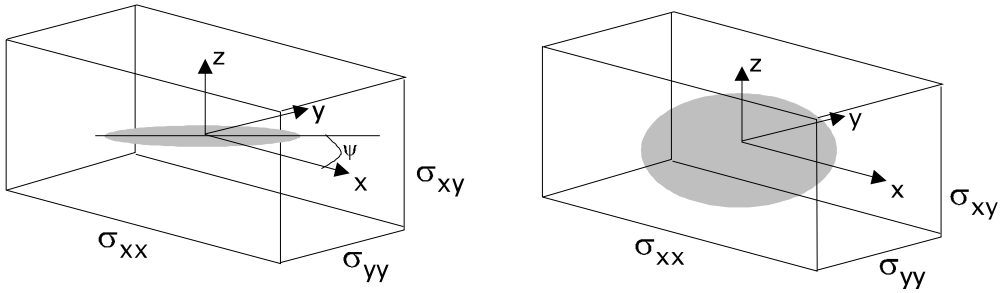


Figure 8.19 Distributions of stress history

Referring to *Figure 8.19*, in the left-hand plot the shaded area contains the cross-plot points, and shows little variation in the orientation of the stresses. The angle ψ between the regression line and the x-axis defines the orientation of the critical plane. In the right-hand diagram the cross-plot of the three stresses is much more diffuse, indicating a wide variation in principal stress orientation. A critical plane analysis would be necessary to identify the most damaged plane.

In the example presented in (Ref. 8.3), some 170 elements were fatigue-critical, and 11% of these elements required a full critical plane analysis. The saving in computation time was very significant.

8.9 Accuracy of results

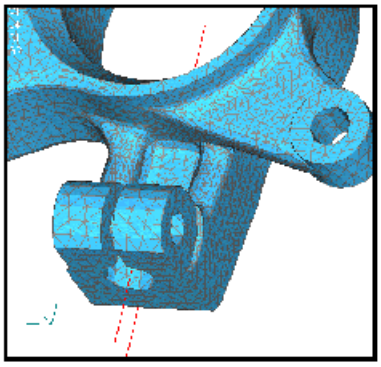
The subject of fatigue analysis from finite element models is relatively new. However, the initial experience with existing analysis techniques is encouraging. It is being shown that fatigue hotspots can be identified reliably in components subjected to multiaxial loading with random phase relationships. Calculated fatigue lives also agree well with test lives.

The fatigue analysis requires accurate stress information, and inadequate meshing will not give accurate stresses. A sensitivity analysis, where fatigue lives are calculated for the same model with two different mesh densities, is recommended. It was shown in Section 7 that biaxial stresses can occur in fillet radii even when a single load is applied. An inadequate mesh density will not define biaxial stresses accurately.

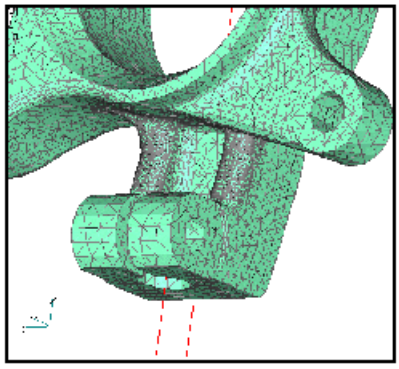
Figure 8.20 shows the effect of an initial 'design' mesh and a final fully-featured mesh (from Ref. 8.4).

8.10 Elastic-plastic FEA

In Section 8.5 it was mentioned that the analysis of a sequence of FEA stresses need not be limited to elastic FEA, and could be applied to results from elastic-plastic FEA. In this case both the stresses and strains are required, and no ‘Neuber’s rule’ correction will be applied to this data. However, when using elastic-plastic FEA for fatigue applications, it is important to select an appropriate hardening model. As described in Section 7, a kinematic hardening model is usually used for fatigue analysis, and a kinematic hardening model must also be selected for the elastic-plastic FEA. Even if this is selected for the FEA analysis, there is some doubt that FEA suites correctly simulate the stress-strain sequences produced by complex fatigue loading.



Initial mesh - typically 35 000 elements, 65 000 nodes.



Fully featured mesh - typically 60 000 elements, 100 000 nodes.

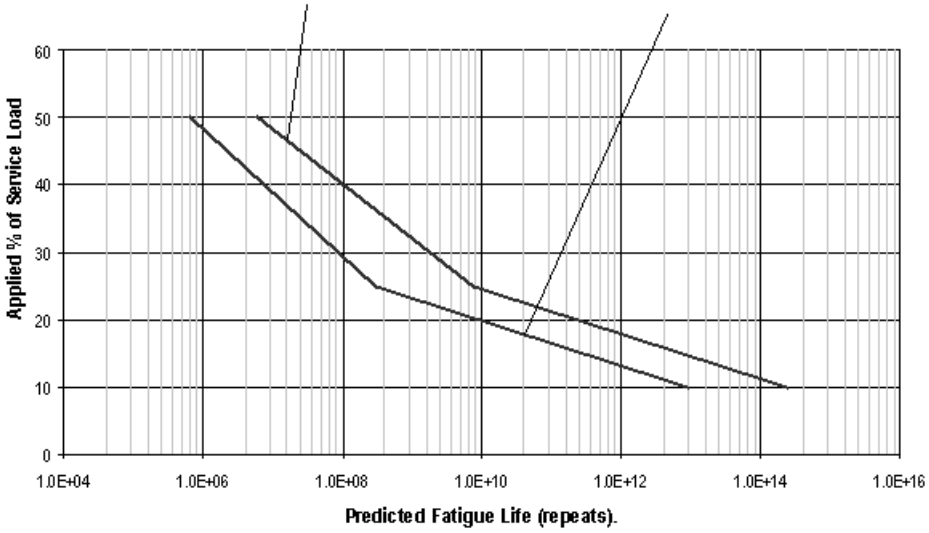


Figure 8.20 Effect of mesh density on calculated fatigue life

8.11 Summary

Major advances have been made in fatigue analysis of finite element models, and correlation between prediction and test results can now be quite acceptable, both for identifying hotspots and in the actual fatigue lives themselves (Ref. 8.5). There is much development to be done, particularly in the field of cyclic plasticity models for out-of-phase biaxial stresses, and in the effect of different types of element on the calculated fatigue lives. Further work will refine many of the algorithms used, and to increase processing speed.

8.12 References

8.1 Moftakhar A, Buczynski A, Glinka G

Calculation of Elasto-Plastic Strains and Stresses in Notches Under Multiaxial Loading.

8.2 Heyes P, Dakin J, St John C

The Assessment and Use of Linear Elastic FE Stress Analysis for Durability Calculations

SAE Paper 951101, 1995

8.3 Chu C, Conle F A, Hubner A

An Integrated Uniaxial and Multiaxial Fatigue Life Prediction Method

VDI Berichte NR 1283, 1996

8.4 Colquhoun C, Draper J

Fatigue Analysis of an FEA Model of a Suspension Component, and Comparison with Experimental Data

Proc. NAFEMS Conference 'Fatigue Analysis from Finite Element Models', Wiesbaden, November 2000.

8.5 Devlukia J, Draper J, McDiarmid D

Integrated Software for Durability Design on Metallic Materials - Final Report to the Department of Trade and Industry, 1997

Rover Group Ltd, Safe Technology Limited and City University London.

9 Using statistics in fatigue

9.1 Introduction

Although strain-controlled fatigue tests on smooth polished specimens of steel and aluminium exhibit little scatter in the test results, fatigue tests on real components do show scatter. This may be caused by variations in material properties, machining tolerances, defects, surface finish and other variations in manufacturing.

At long endurance the scatter is usually expressed as a variation in fatigue strength at a specified endurance. For low cycle fatigue the scatter may be expressed as a variation in endurance for a specified stress or strain amplitude.

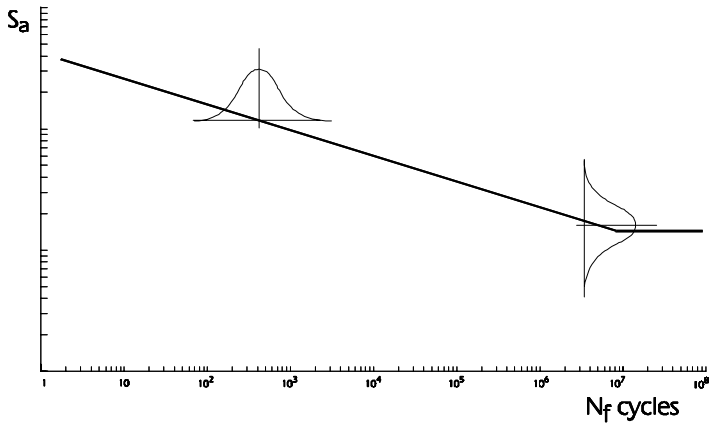


Figure 9.1 Scatter on an S-N curve

The data are usually plotted on $\log_{10}$-$\log_{10}$ axes. Taking the low endurance test results, the $\log_{10}$ mean life of a number of specimens tested at the same stress amplitude is the average value of the $\log_{10}$ values of the test endurance. The mean endurance, *m*, is given by

$$log_{10}\, m = \frac{1}{n}\sum_{i=1}^{i=n} log_{10}(N_f)_i \qquad (9.1)$$

where there are *n* test results and $(N_f)_i$ is the endurance in cycles of the *i* th specimen.

The extent to which the results are dispersed from the mean value is defined by the standard deviation of the test results. The standard deviation *s* of a set of test data is obtained from

$$log_{10}(s^2) = \frac{1}{n}\sum_{i=1}^{i=n}\left[log_{10}(N_f)_i - log_{10}(m)\right]^2 \qquad (9.2)$$

The parameter s^2 is the variance of the data.

Similar equations apply for the strain-life curves used in local strain fatigue analysis.

9.2 Gaussian distribution

9.2.1 Probability of failure

The scatter in the data about the mean value is expressed by a probability distribution, which shows the probability that a specimen will have a test life less than a specified endurance, greater than a specified endurance, or between two specified values of endurance.

One probability distribution which has been proposed for fatigue test results is the Gaussian or normal distribution. Its probability density is given by the equation

$$p(\log_{10}(N_f)) = \frac{1}{\sqrt{2\pi}\,\log_{10} s} \exp\left(- \frac{(\log_{10} N_f - \log_{10} m)^2}{2\,\log_{10} s^2} \right) \qquad (9.3)$$

The distribution shows that most test results will have a value close to the mean life, because in general the various factors which determine the fatigue life will balance each other. Occasionally most of the factors will act together to produce a very good or a very poor test result.

Because the distribution is calculated from the $\log_{10}$ values of endurance, it is sometimes called the *lognormal* distribution.

The Gaussian distribution is a two-parameter distribution, requiring only the mean and the standard deviation to define its shape. Once these have been calculated, or estimated, the probability of any test result can be determined. The probability of fatigue failure below a certain life is obtained from the area in the lower tail of the probability distribution *(Figure 9.3)*. The probability of failure between two specified lives can be obtained by integrating the probability density distribution between these lives.

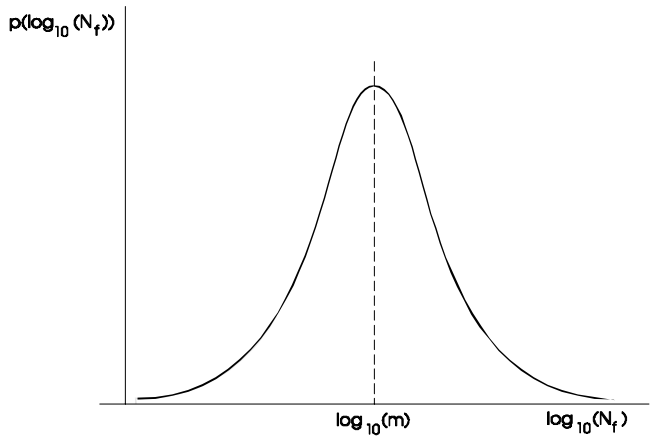


Figure 9.2 Gaussian probability

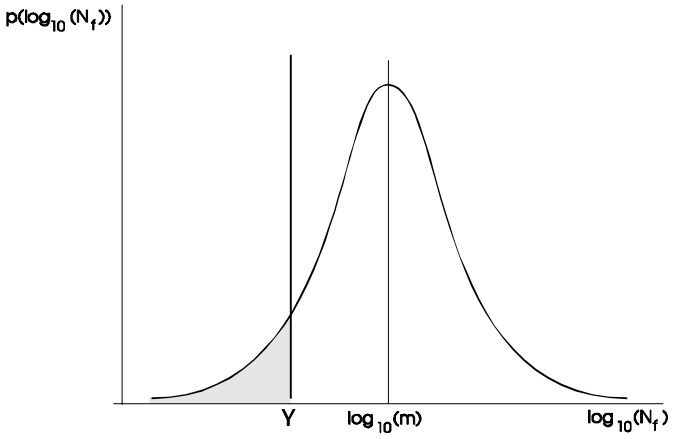


Figure 9.3 Probability of a failure before a specified life

The following table shows the probability of failure at a life shorter than the calculated life, for various numbers of standard deviations below the mean life.

No of SD's	probability of failure
0	50 % (mean life)
1	15.9 %
2	2.2 %
3	0.13 %

Because the Gaussian distribution applies to many naturally occurring phenomena, it has been widely analysed, and published tables can be used to obtain the probabilities shown in *Figure 9.3* (Ref. 9.1).

Standard Gaussian probability paper can be used to obtain the mean and standard deviation of a set of test results, as shown in example 9.1.

Example 9.1

The following endurances were obtained from constant amplitude tests on nine identical components. The specimens are numbered in order of failure.

Order number k	1	2	3	4	5	6	7	8	9
Endurance N_f x 10^5 cycles	2.0	2.4	2.5	2.8	2.9	3.2	3.4	3.8	4.2
$\log_{10} N_f$	5.30	5.38	5.40	5.44	5.46	5.50	5.53	5.58	5.62

The probability of failure for each specimen is defined by

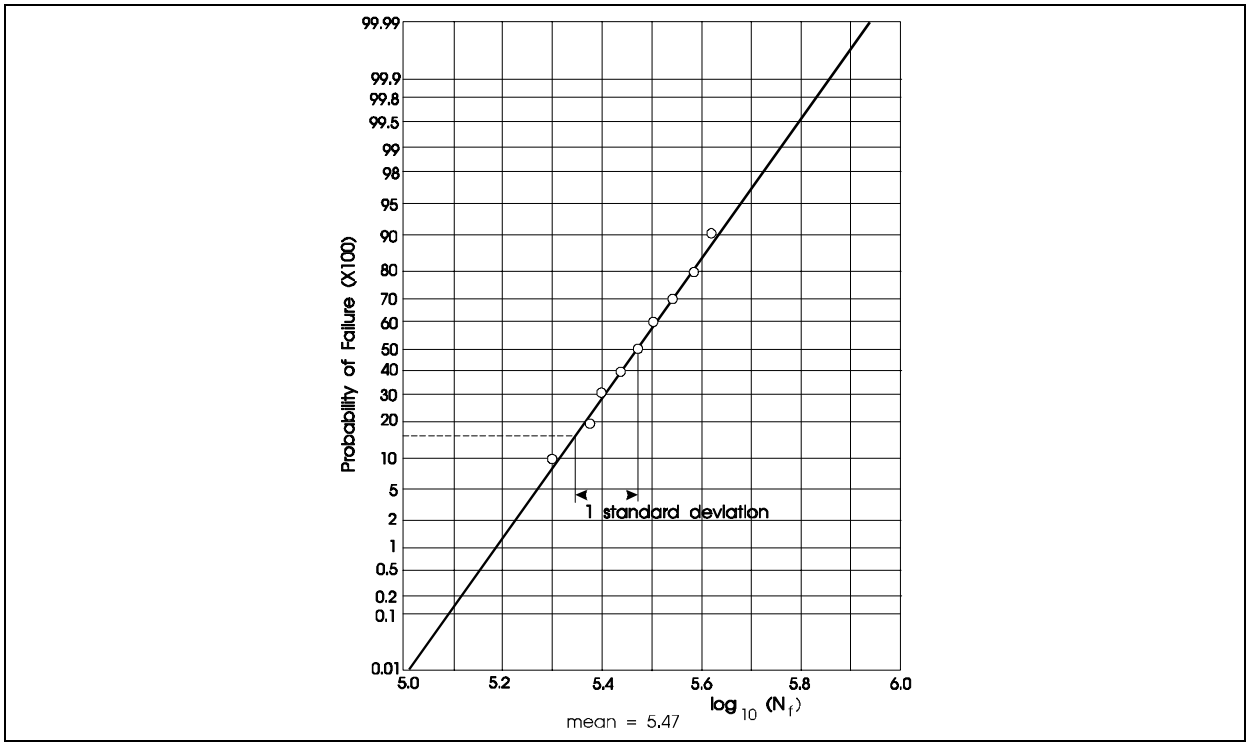
$$p(x) = \frac{k}{n+1}$$

where k is the order number and n is the number of components tested, so that for example, specimen 1 has a probability of failure of 0.1, specimen 2 a probability of 0.2, etc.

The data points are plotted in the figure. The mean life is the 50% probability life, so

$\log_{10}$ (mean life) = 5.47, and the mean life = 2.95 x 10^5 cycles.

One standard deviation is equivalent to a 15.9% probability of failure, and from the graph the log standard deviation s = 0.12



(Table corrected 4.12.01)

9.2.2 Estimates for the population

The data in example 9.1 was a small set of test results. The calculated values of mean and standard deviation apply to the set of results, or sample, not to the population of components from which the sample was taken. To estimate the statistical properties of the population, the following equations can be used (Ref. 9.1).

If μ_e is the estimate of the mean of the population, $\mu_e = m$

where m is the mean obtained from the test results

If σ_e is the estimate of the standard deviation of the population,

$$\sigma_e = \frac{s}{c}\left(\frac{n}{n-1}\right)^{0.5} \qquad (9.4)$$

where n is the number of specimens in the sample, and c is a correction factor for the bias in the estimate of σ_e

The equation can be re-written as σ_e = Ks, and values of K can be obtained from (Ref. 9.1). Example values are:

n	**5**	**10**	**20**	**50**	**100**
K	1.19	1.07	1.03	1.01	1.0

9.2.3 Confidence limits

As the values of μ_e and σ_e are *estimates* for a population made from test results on a number of specimens, it is necessary to define a level of confidence for the population estimates. For the true mean of the population μ

$$\mu = m \pm \left(\frac{t_{(\alpha)}}{\sqrt{n}}\right)\sigma_e$$

where $t_{(\alpha)}$ is the number of standard deviations between the sample mean and the mean of the population at a confidence level of α.

At a confidence level α of 90%, the following values apply (Ref. 9.1).

n	**5**	**10**	**20**	**50**	**>100**
$\left(\frac{t_{(\alpha)}}{\sqrt{n}}\right)$	0.95	0.65	0.38	0.25	0.17

This means that, for example, if a sample of 10 specimens was tested, there is a 90% chance that the true mean of the population will be within $\pm 0.65\sigma_e$ of the sample mean.

For the true standard deviation of the population, σ,

$$\left(\frac{\sigma}{\sigma_e}\right)_{q_2} > \left(\frac{\sigma}{\sigma_e}\right) > \left(\frac{\sigma}{\sigma_e}\right)_{q_1} \qquad (9.5)$$

where $q_1 = \frac{100 + \alpha}{2}$ and $q_2 = \frac{100 - \alpha}{2}$

Again for a 90% confidence level, $q_1 = 95$, $q_2 = 5$ and

n	**5**	**10**	**20**	**50**	**100**
$\left(\frac{\sigma}{\sigma_e}\right)_{q_2}$	2.4	1.64	1.35	1.2	1.15
$\left(\frac{\sigma}{\sigma_e}\right)_{q_1}$	0.65	0.73	0.8	0.86	0.9

For the same sample of 10 specimens, the table shows that there is a 90% chance that the standard deviation of the population will lie between 0.73 and 1.64 times the standard deviation calculated.

When using these relationships with fatigue test data, the $\log_{10}$ values of endurance, or stress/strain amplitude, must be used, as a lognormal distribution is assumed. (Ref. 9.1) provides values for a wide range of confidence levels.

9.3 The Weibull distribution

In fatigue it is necessary to predict the probability of a premature failure. For safety-critical components the requirement may be for a very low probability of premature failure, perhaps 1 in 1000 or less. The Gaussian probability distribution has not been proven for fatigue applications at such low probabilities - more than 3 standard deviations below the mean life. It must be in error for very low probabilities of failure because the probability distribution never falls to zero, implying that there is a finite, though very small, chance that a component will fail in fatigue on the first load application. In fact, it takes a number of load cycles for the development of crack initiation along slip-planes.

The Weibull distribution is a three-parameter distribution, with the third parameter allowing for changes in shape. *Figure 9.4* shows examples of the Weibulll distribution, and it can be seen that including the shape parameter allows the 'short-life' tail of the distribution to be modified.

If *R(t)* is the cumulative probability of survival, i.e. the ratio of survivors, after a time *t*, to the total initial population, and *F(t)* is the cumulative probability of failure, that is 1-*R(t)*, after a time *t*, then the probability density distribution of failure, *f(t)*, is given by the rate of change of *F(t)*, i.e.

$$f(t) = \frac{d}{dt}(F(t)) = -\frac{d}{dt}(R(t))$$

f(t) defines the number of failures per unit time (or cycle) experienced by the remaining population of survivors, which is itself changing with time as the number of failures increases.

To derive the risk of failure for an individual component, or hazard function *Z(t)*, the total failure rate in the current population must be divided by the current population. This current population equals the proportion still surviving, R(t).

Then $$Z(t) = \frac{f(t)}{R(t)} = -\frac{\frac{d}{dt}(R(t))}{R(t)} \qquad (9.6)$$

This equation relates the risk of failure of an individual component to the observed frequency of failures in the total population. In fatigue, the hazard function *Z(t)* increases with time as the progressive application of loading cycles contributes to the processes of slip and crack propagation.

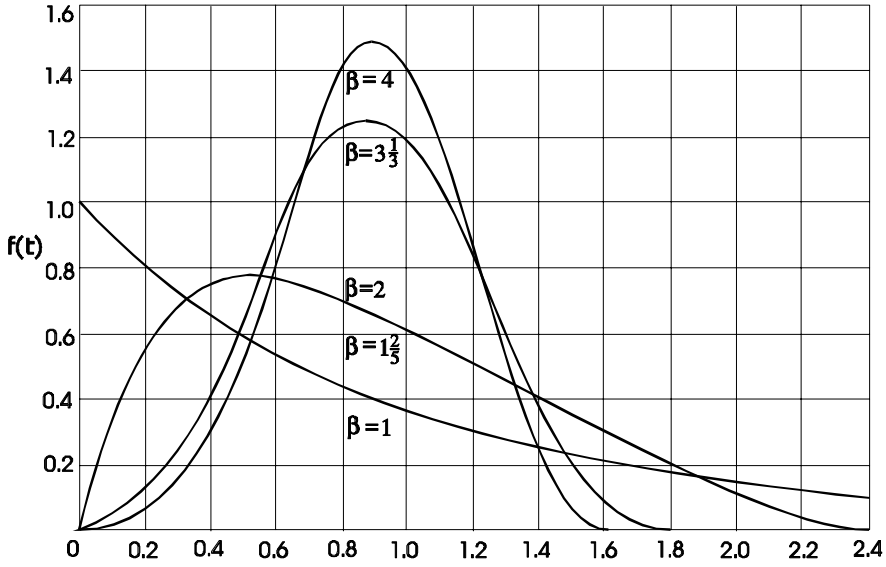


Figure 9.4 Weibull probability distributions

Developing equation (9.6) by integrating it,

$$\int_o^t Z(t).dt = \int_o^t \frac{-dR(t)}{dt} \cdot \frac{1}{R(t)} \cdot dt$$

$$= \int_0^t \frac{dF(t)}{dt} \cdot \frac{dt}{(1-F(t))}$$

$$= \Big(-\log(1-F(t))\Big)_{F(o)}^{F(t)}$$

$$= \Big(-\log R(t)\Big)_o^t = -\log R(t)$$

and so

$$R(t) = exp\left(- \int_{o}^{t} Z(t)\, dt \right) \qquad (9.7)$$

This expresses the link between the hazard function *Z(t)* and the resulting cumulative probability of survival and therefore the cumulative probability of failure. Assuming a form for *Z(t)* then determines the distribution of failures.

The distribution proposed by Weibull is of the form

$$R(t) = exp\left(-\left(\frac{(t - \gamma)}{\eta} \right)^{\beta} \right) \qquad (9.8)$$

where t = time at which the survival probability is estimated

γ = time at which F(t) = 0

β = a factor determining the shape of the distribution

η = a scaling factor

The influence of the shape factor β can be seen in Figure 9.4.

Note that when β = 3.3 the distribution has a Gaussian-like shape.

The introduction of the parameter γ means that instead of assuming that the possibility of failure starts when *t* is zero, a period of zero probability of failure is allowed. Data to derive *Z(t)* is usually contained in observations of early failures, and the most common procedure is to use this information to estimate the three parameters. As there are no simple formulae, most methods rely on plotting cumulative failure probabilities.

Graph paper to the Weibull distribution is available. The formulae used to transform the axes is derived as follows.

Equation (9.8) can be re-arranged as:-

$$\frac{1}{1-F(t)} = exp\left(\frac{t - \gamma}{\eta} \right)^{\beta}$$

Taking (natural) logs twice then gives

$$log_e\, log_e \frac{1}{1-F(t)} = \beta\ log_e (t - \gamma) - \beta\, log_e\ \eta \qquad (9.9)$$

Weibull probability paper is based on this equation. Weibull plots link the slope with β, and position with the parameter η which is the scaling factor (similar in its effect to population mean). The remaining factor is γ, the point in time at which failures start. Various suggestions have been made for estimating γ (Ref. 9.2), but in practice trial and error methods are usually used to derive a value which produces a straight line on the Weibull paper.

Because the difference between the Weibull and Gaussian distributions is small for failure rates above about 1 in 500, Weibull plotting is rarely useful for laboratory test results (fatigue tests on more than 500 identical specimens are not common). However, the Weibull distribution is used for monitoring early service failures. The number in the sample is then the total number of components in service. (If test data plotted on Gaussian probability paper shows a pronounced curve, this may be an indication that a Weibull distribution may be more applicable.)

The factor η is usually referred to as the "characteristic life" and defined as the life at which the cumulative failure probability is 63%. This figure is chosen by putting $(t-\gamma)/\eta$ equal to unity, so that whatever the value of β, the right hand side of equation (9.8) gives $^1/_e$.

9.4 Failure probability in design

Once the distribution of failures has been defined this information could be used to choose a combination of design and allowable duty to achieve a specified failure rate. Taking as an example a welded joint, the scatter in test results has been defined, and a typical value of standard deviation of $\log_{10}$ endurance is 1.7 (Ref. 9.3). Using the lognormal (Gaussian) distribution, this implies that if all the components are withdrawn from service at $\frac{1}{3.4}$ = 29% of the mean life (2 standard deviations below the mean) there is a 2.3% chance of a component failing before withdrawal. This is the basis of 'safe-life' design, and shows how uneconomic it can be, as at least half the components could have been left in service for more than three times the 'safe' service life without failure.

To lower the probability of premature failure from 2.3% to (say) 0.13%, (3 standard deviations below the mean) the safe life would have to be reduced to $\frac{1}{5.1}$ = 19% of the mean life. The lognormal distribution starts to become invalid at such low probabilities of failure, and indeed a large number of test results would be required to support any statistical prediction at such low probabilities. (In practice, many components designed on 'safe-life' principals are left in service well beyond their calculated safe life, and are inspected at regular intervals to ensure that cracks are not initiating or propagating to dangerous lengths. The components are then treated as 'damage tolerant' or 'fail-safe' by means of regular inspection.)

9.4.1 The P-S-N curve

The S-N curve in *Figure 9.1* approximates to a 50% probability of failure curve. Similar S-N curves can be drawn for 10% and 1% of failures, giving a family of P-S-N (probability-stress-life) curves. A horizontal section along a constant-stress line gives the cumulative distribution of lives at constant stress, and a vertical section at constant life gives the distribution of stress which would produce this life in components of differing initial strength.

These values refer only to test results obtained from the set of specimens. The procedures described in sections 9.2.2 and 9.2.3 must be used to produce P-S-N curves for the population of components, with appropriate confidence limits

9.5 References

9.1 ESDU

Data Sheets

9.2 Bloomer N T and Roylance T F

A Large-Scale Fatigue Test Of Aluminium Specimens

Aeronautical Quarterly, Vol XVI, November 1965, pp307 - 322.

9.3 British Standards Institution

The Design and Specification of Steel and Concrete Bridges

British Standard BS5400 part 10:1980

See also

Rompas-Smith J H

Mechanical Survival: The Use Of Reliability Data

McGraw-Hill, London 1973.

American Society for Testing and Materials

A Guide for Fatigue Testing and the Statistical Analysis of Fatigue Data

ASTM STP 91-A, 1963.

Davies O L
The Design and Analysis of Industrial Experiments
Oliver and Boyd, London, 1956.

10 Crack propagation

10.1 Introduction

Fatigue design for many components is a process of design to prevent the initiation of cracks. For many components this represents the total design life, either because a crack will propagate so rapidly that the life to crack initiation represents almost the total life, or because product liability legislation precludes the release into service of components which will develop cracks during service even though the component may still achieve its design life.

This design philosophy cannot be applied to all components. For example, aircraft structures can experience accidental damage in flight, and crack growth calculations must demonstrate that the structure is airworthy after impact damage. Other components may develop cracks in service as a result of design errors, and it may then be necessary to calculate safe inspection intervals to prevent the cracks causing catastrophic fracture.

The study of predominantly brittle fracture from short cracks and defects is the subject of *Fracture Mechanics*.

10.2 The purpose of fracture mechanics

A structure may develop a fatigue crack under repeated service loading. As the crack length increases, it causes a greater stress concentration, and so the rate of crack propagation increases. The strength of the structure is reduced as the crack length increases. Eventually, the strength of the structure will be insufficient to withstand the **highest** service loads, and failure is possible if the highest service load occurs. As the crack continues to grow, the structural strength will be reduced until the structure cannot withstand the **normal** service loads, and so fracture under normal loading becomes very probable.

The designer must be able to calculate

-the residual strength as a function of crack size
-the crack size that can be allowed at the expected service load (the critical size)
-how long it takes the crack to grow from a certain size to a critical size
-the size of initial flaws that can be tolerated in a new component
-the interval between inspections for cracks.

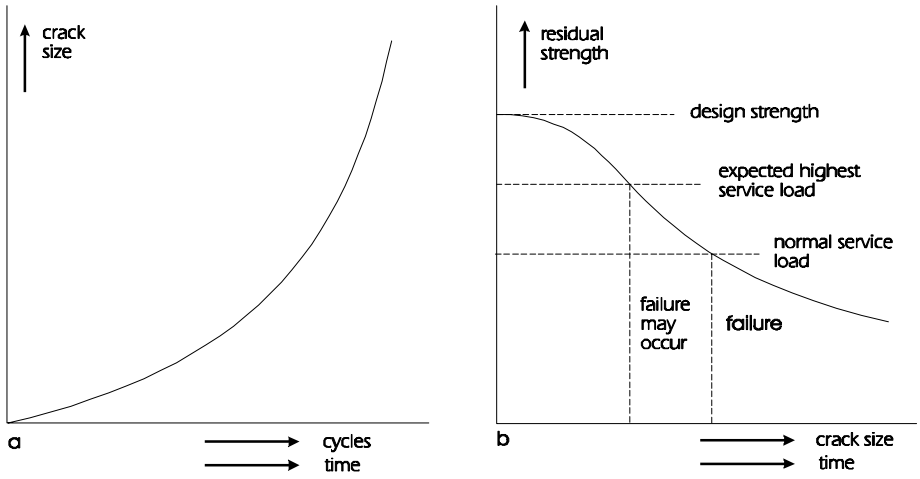


Figure 10.1 Relationship between crack length and failure load

Fracture mechanics tries to provide the tools with which to answer these questions. The subject includes the materials science studies of fracture processes on an atomic scale, the growth of cracks, the analysis of crack tip stresses and the behaviour of cracks in these stress fields, the provision of materials

properties by testing (and the development of test techniques), and finally the engineering application of these techniques to the analysis of real structures.

10.3 Stresses at the crack tip

A crack can be stressed in three different modes.

Mode I is the normal or opening mode, Mode II is the tearing mode caused by in-plane shear, and Mode III is the tearing mode caused by out-of-plane shear.

Most problems can be addressed by superimposing one or more of these modes, but Mode I is by far the most important for practical analysis.

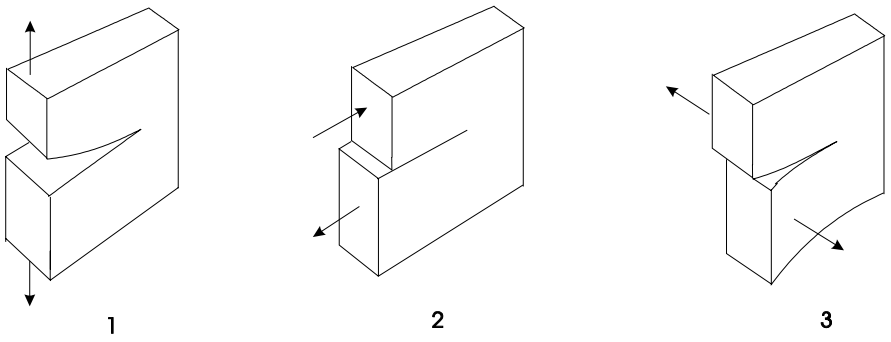


Figure 10.2 The three crack-opening modes

For an arbitrary through-thickness crack of size *a*, in a body of arbitrary shape and loaded by arbitrary Mode I loading, the stress distribution near the crack tip is :

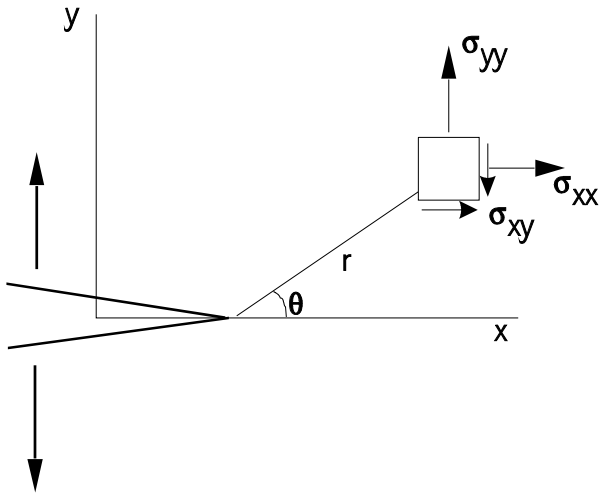


Figure 10.3 Generalised stresses at the crack tip

It would be expected that the stresses reduce with increasing *r*, and will be a function of the angle θ. The general expression for the stress field is

$$\sigma_{ij} = \frac{K_I}{\sqrt{2\pi r}} f_{ij}(\theta) \qquad (10.1)$$

where σ_{ij} are the stresses acting on an element *dxdy* at a distance *r* from the crack tip at an angle θ from the crack plane, and $f_{ij}(\theta)$ are known functions of θ. The factor K_I is as yet unknown, because the loading and the shape of the body are not yet defined.

As it appears that the entire stress field close to the crack tip can be described if K_I is known, K_I is an important parameter, called the *stress intensity factor*. The subscript I denotes mode I loading.

Equation 10.1 can be applied to an infinite plate subjected to uniform tension, with a central crack of length *2a*.

The standard convention is to let the length of the crack equal *2a* if the crack has two tips and equal *a* if the crack has one tip.

As the left hand side of equation 10.1 has unit of stress, and as f*(θ)* has no units, then $\frac{K_I}{\sqrt{2\pi r}}$ must have units of stress, and so K_I must have units of "stress' multiplied by 'square root of length'". As the only stress is the remote stress, and as the only length is the crack length '*a*', then

$$K_I = \beta\sigma\sqrt{a} \qquad (10.2)$$

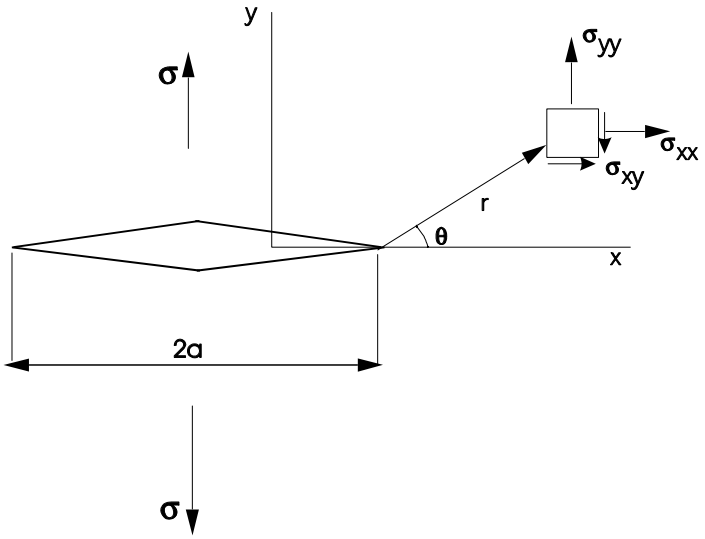


Figure 10.4 Generalised stresses for a crack in an infinite plate

In fact $\beta = \sqrt{\pi}$ for a centre crack in an infinite plate, but it is the convention to refer all stress intensity factors to the infinite plate case (i.e. make the constant equal to 1 for this case). Therefore K_I is given by

$$K_I = \beta\sigma\sqrt{\pi a} \qquad (10.3)$$

The factor β is a function of the crack dimension '*a*' and the shape of the component.

Values of β have been tabulated in the references.

The stress σ in these equations is always the remote stress, not the stress in the cracked section. The fact that the average stress in the net section will be higher is accounted for in the tabulated values of β.

Combining equations 10.1 and 10.3

$$\sigma_{ij} = \beta\sigma\sqrt{\frac{a}{2r}}\, f_{rj}(\theta) \qquad (10.4)$$

For mode I loading, the function $f_{ij}(\theta)$ can be added to give

$$\sigma_x = \beta\sigma\sqrt{\frac{a}{2r}} . \cos\frac{\theta}{2}\left(1 - \sin\frac{\theta}{2} . \sin\frac{3\theta}{2}\right)$$

$$\sigma_y = \beta\sigma\sqrt{\frac{a}{2r}} . \cos\frac{\theta}{2}\left(1 + \sin\frac{\theta}{2} . \sin\frac{3\theta}{2}\right) \qquad (10.5)$$

$$\sigma_z = \beta\sigma\sqrt{\frac{a}{2r}} . \sin\frac{\theta}{2} . \cos\frac{\theta}{2} . \cos\frac{3\theta}{2}$$

This single set of functions $f(\theta)$ applies to all mode I cases.

Equations 10.4 provides the stresses close to the crack tip only. It is in fact a truncation of a series

$$\sigma_{ij} = \beta\sigma\sqrt{\frac{a}{2r}} + \text{2nd term} + \text{3rd term} + \ldots\ldots \qquad (10.6)$$

For small values of r, the first term tends to infinity (see later) but the remaining terms tend to zero, and so can be neglected.

Using only the first term implies that the stresses tend to zero at large values of r. In fact they tend to σ , the remote stress. Hence the need for the 2nd and subsequent terms at large values of r.

10.4 Fracture toughness

Equation (10.1)

$$\sigma_{ij} = \frac{K_I}{\sqrt{2\pi r}} f_{ij}(\theta)$$

applies to all different cracked bodies, the above formula still applies. If the load is increased until fracture occurs, then the fracture stress from the above equation can be substituted into the equation

$$K_I = \beta\sigma\sqrt{\pi a}$$

The value of K_I at fracture is called the fracture toughness.

Fracture occurs when the value of K_I equals the fracture toughness value. It is therefore a similitude parameter. Fracture occurs at a particular value of K_I whatever the shape of the body or the size of the crack.

10.5 Plasticity

Equation 10.1 is an elastic equation and implies that the crack tip stresses become infinite at the crack tip, i.e. $r = 0$. This cannot be true, and in fact plastic deformation occurs which keeps the crack tip stress finite.

A first estimate of the size of the plastic zone can be obtained from diagrams such as *Figure 10.5* by determining the distance from the crack tip (r_p^*) over which the stresses exceed the yield stress for the material.

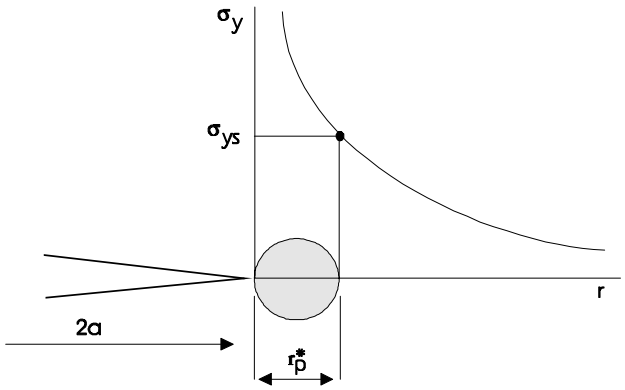


Figure 10.5 Estimate of crack tip plastic zone size

Substituting $\sigma_y = \sigma_{ys}$ into equation 10.1 gives

$$\sigma_y = \frac{K_I}{\sqrt{2\pi r_p^*}}$$

so that

$$r_p^* = \frac{K_I^2}{\sqrt{2\pi\sigma_{ys}^2}}$$

In fact the size of the plastic zone is larger *(Figure 10.6)*

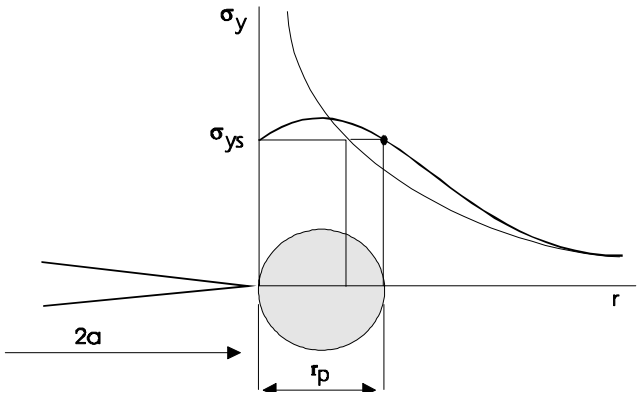


Figure 10.6 More accurate estimate of crack tip plastic zone size

However, the approximate value r_{p^*} has the advantage that it can be obtained directly from the stress intensity factor and the yield stress. K_I still represents a similitude parameter, provided the size of the plastic zone is still determined by K_I. In other words, plasticity must be restricted to an area close to the crack tip - the area where the 2nd and subsequent terms in equation 10.6 can be neglected.

10.6 Multiaxial stresses at the crack tip

If a bar of material is subjected to a tensile stress, it will strain in the longitudinal direction, and also transverse strains will be developed - the bar will become thinner. The transverse strain is given by

$$\varepsilon_z = \frac{\sigma_{zz}}{E} - \nu\frac{\sigma_{xx}}{E} - \nu\frac{\sigma_{yy}}{E} \qquad (10.6)$$

which means its magnitude is about 1/3rd of the longitudinal strain, as it is usual to take $\nu = 0.3$. For plastic deformation, the transverse strain is larger, as a value of $\nu = 0.5$ is usually used for fully plastic conditions.

At the crack tip the local stresses are very high. A large transverse strain should develop, particularly in the plastic zone. However, the large bulk of material surrounding the plastic zone is at a much lower stress, and so will contract to a much less extent. This restrains the crack tip zone from contracting, and so the transverse strains will not develop. Therefore the material must be subjected to a transverse stress.

Clearly these transverse stresses will only develop if the material is sufficiently thick. In a thin sheet, the out-of-plane stresses are zero (the condition of plane stress).

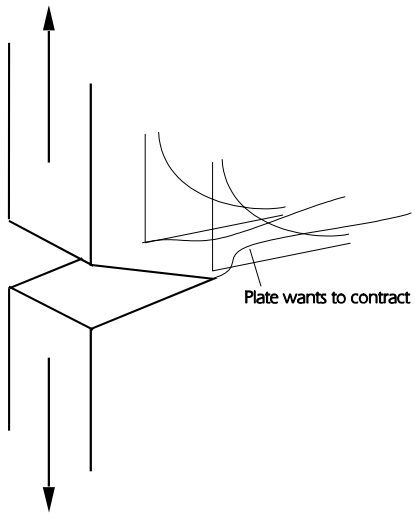


Figure 10.7 Out-of-plane stresses at the crack tip

The ratio between the plate thickness and the diameter of the plastic zone will determine whether plane strain conditions exist. If the thickness of the plate is B, and the diameter of the plastic zone is

$$r_p^* = \frac{C K_I^2}{\sigma_{ys}^2}$$

then if the ratio $\frac{\beta}{(K_I/\sigma_{ys})^2}$ exceeds a certain value, say Q, then *plane strain* conditions exist.

If the ratio $\frac{\beta}{(K_I/\sigma_{ys})^2}$ is less than (say) q, then *plane stress* conditions exist.

The value for Q has been determined experimentally to approximately 2.5, so plane strain conditions exist if

$$\beta > 2.5\left(\frac{K_I}{\sigma_{ys}}\right)^2$$

Because of the different stresses in the through-thickness direction, the behaviour of cracks is different in plane strain and non-plane strain conditions. The values of stress intensity at which fracture occurs

(the fracture toughness) is also different. This means that the toughness depends on the thickness of the material, unless the thickness is sufficiently great for plane strain conditions to exist.

The toughness in plane strain conditions is called the plane strain fracture toughness. It is the critical value of the stress intensity at which the crack extends to fracture, and is denoted (for mode I) by K_{IC}

10.7 Stress corrosion cracking

In a corrosive environment, fracture may occur at a lower value of $K_{IC,}$ particularly if the load on the component is held for a period of time.

The value of K_{IC} for which fracture will not occur however long the load is held, is the threshold value shown in Fig 10.8, and is denoted by K_{ISCC}.

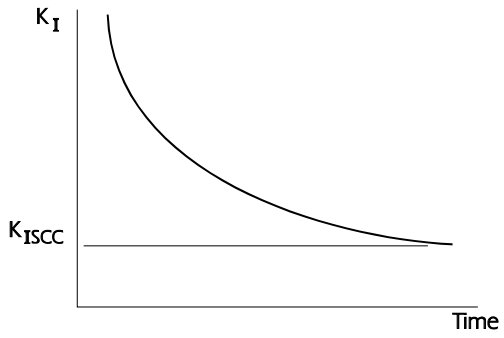


Figure 10.8 Definition of K_{ISCC}

10.8 Crack propagation

The preceding sections have concentrated on the calculation of final fracture. However, it is possible that the applied stress on a crack may cause the crack to extend a small amount, without causing fracture.

If the applied load varies between zero and some positive value (constant amplitude cycling) the stress intensity K varies over a range ΔK. The rate of crack propagation da/dN must depend on the range of the stress intensity ΔK so that

$$\frac{da}{dN} = f(\Delta K) = f(2 S_a \sqrt{\pi a}) \qquad (10.7)$$

where S_a is the stress amplitude.

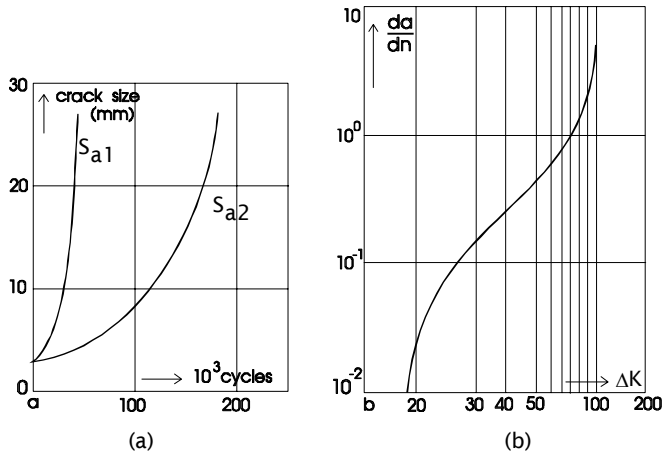


Figure 10.9 Crack propagation diagrams

The results of the tests in Fig 10.9(a), for two different stress amplitudes, can be plotted on the basis of $\frac{da}{dN}$ at the instantaneous value of ΔK. In *Figure 10(b)* the data has been plotted on a log-log basis.

The fact that both sets of data fall on the same curve shows that equation (10.7) applies for small cracks at high stress, or large cracks at low stress. The two cracks have the same growth rate providing the values of ΔK are the same. Crack growth rate is a function of ΔK together with two material constants. The central part of the curve in *Figure 10(b)* is the approximately straight line and for this central region, equation (10.7) can be re-written as

$$\frac{da}{dN} = C(\Delta K)^n \qquad (10.8)$$

Equation 10.3, that $K_I = \beta\sigma\sqrt{\pi a}$, allows values of ΔK to be calculated for any combination of crack geometry and remote stress, and these two equations are the fundamental crack growth equations.

Figure 10.10 shows an example of the variation of β with crack length for a crack in a circular shaft.

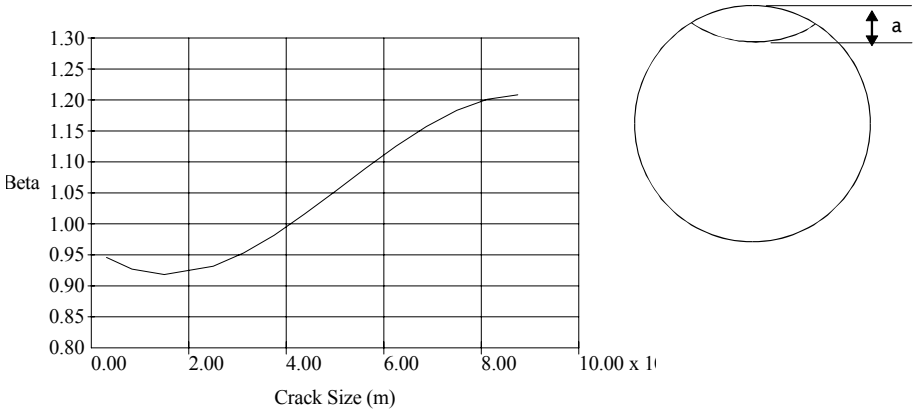


Figure 10.10 Example of variation of β with crack length

For constant amplitude sinusoidal stress cycles, with stress range $\Delta\sigma$, the crack growth calculation is:

For an initial crack size *a*, the value of β can be obtained from the graph.

The parameter $\Delta K = \beta\Delta\sigma\sqrt{\pi a}$ can be calculated.

The value of $\frac{da}{dN}$ can be calculated from the crack growth relationship

$$\frac{da}{dN} = C(\Delta K)^n$$

The number of cycles, N , for the crack to grow (say) 1mm, can be calculated.

For the new crack length, a new value of β can be obtained, and the crack growth calculation repeated. This incremental calculation is repeated until the required maximum crack size is achieved, or until the value of fracture toughness K_{IC} is reached.

For variable amplitude loading, the loading can be Rainflow cycle counted to produce a cycle histogram. The crack growth for one repeat of the histogram (one 'block') can be calculated, and the number of blocks required to propagate the crack to the next chosen length can be obtained. The value of b is adjusted as before, and the calculation repeated. This method works well if the block lengths are small,

producing many repeats for each crack increment. An example of a crack growth curve is shown in *Figure 10.11*.

Two factors complicate this simple calculation - crack retardation, and the effect on crack growth of the compressive part of a stress cycle.

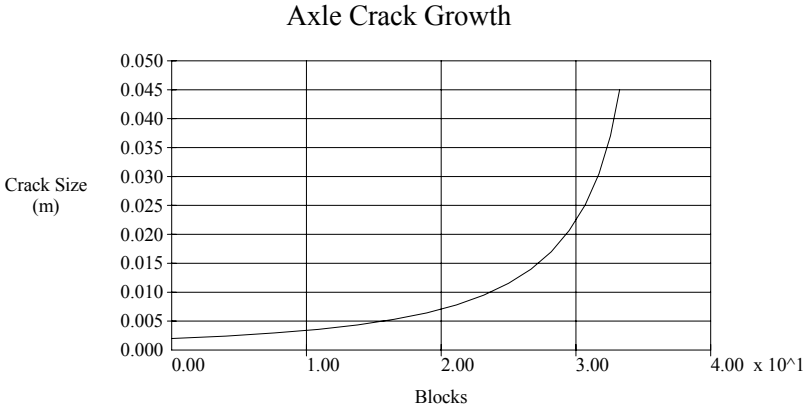


Figure 10.11 Example of a calculated crack growth curve

Crack retardation can occur as a result of a large tensile stress. After a large tensile stress, the crack propagation rate is reduced, (and may cease altogether if the stress is large enough). The extent of retardation increases with increasing overload stress, and is a function of the plastic zone size at the crack tip. A compressive overload can cause crack growth acceleration.

Several models have been proposed to allow crack retardation effects to be included in crack growth calculations. As an example, the Wheeler model modifies the Paris equation

$$\frac{da}{dN} = C(\Delta K)^n$$

by the addition of an empirical 'retardation parameter' C_i to give $\frac{da}{dN} = C_i C(\Delta K)^n$

C_i is a function of crack length, plastic zone size, and an empirical shape factor.

There is considerable doubt on the application of crack retardation models, as most models predict that maximum retardation occurs immediately after an overload, whereas experimental results show that maximum retardation occurs after the crack has propagated about one-quarter of the way through the plastic zone caused by the overload. Also, some experimental results show that crack growth can accelerate after large tensile stresses. It may be that the retardation model must be chosen with regard to the type of load spectrum being considered. The application of retardation models also precludes 'block loading' calculations, as each cycle must be considered separately. This requires a measured signal, or a cycle-by-cycle randomisation procedure, and the calculation takes a large amount of computer time. The inclusion of a crack retardation model may also mean that crack propagation will be faster than calculated if the component is subjected to a less severe load history than that used for the calculation. Block loading without crack retardation should generally produce a conservative life estimate for tensile signals.

10.9 Reference

10.1 David Broek

Elementary Engineering Fracture Mechanics

David Broek

Martinus Nijhoff Publishers 1987

Also by David Broek, and equally valuable, is

The Practical use of Fracture Mechanics

Kluwer Academic Publishers 1989

11 Fatigue of welded steel joints

11.1 Introduction

The fatigue strength of welded joints in structural steels is defined by British Standards. A major re-analysis of welded joint data took place in the 1970's and the results were incorporated into the design standard for welded steel bridges (BS5400:1980, Part 10) and later into the fatigue standard BS 7608 (Ref. 11.1).

The major features of BS 7608 are that, for the range of conditions covered,

different welds are placed in groups having similar fatigue strengths

each group has a characteristic stress-life curve

the fatigue analysis is based on nominal engineering stress

fatigue strength is not sensitive to changes in mean stresses in the applied loading

fatigue strength is not sensitive to material tensile strength

the user may define the probability of failure

The importance of fatigue calculations for welded joints can be seen from the following diagram, which compares the fatigue strength of a typical welded detail with those of plain and notched steel plate.

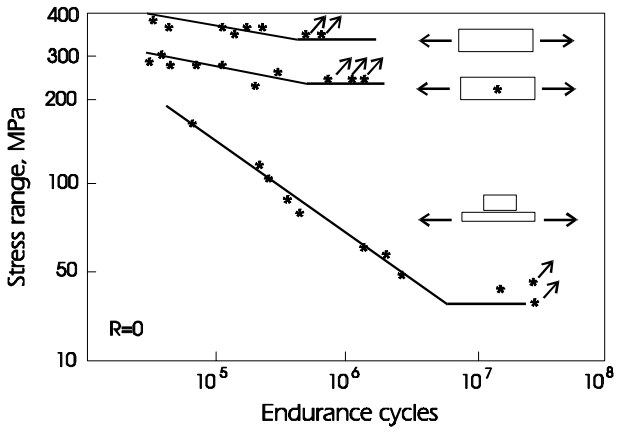


Figure 11.1 Comparison of the fatigue strength of a welded steel joint with smooth and notched specimen data

11.2 Weld classification

In BS 7608 structural details are divided into three broad groups according to the anticipated location of fatigue crack initiation. These groups are

a) **Away from all welding.**

This covers plain plate, geometric stress concentrations such as holes and cutouts, and bolted and riveted joints.

b) **Welded details on the surface of a member.**

This includes longitudinal fillet welds (for example, between the webs and flanges of build-up beams), and welded-on reinforcing plates on the surface or edge of a member.

c) **Welded details at ends of members.**

This includes transverse butt welds and transverse fillet welds joining members end to end in load-carrying joints.

These three groups are subdivided into different types of weld. For example, transverse butt welds are subdivided depending on whether they have been ground flush or left as welded, whether the weld joins plates of equal width and/or thickness, and so on. Each type of weld is then assigned a 'classification', which defines the S-N curve to be used in its fatigue assessment.

Class B has the highest fatigue strength, class C is lower, followed by D, E, F, F2 and G, with class W for load-carrying weld metal. The wide variation in fatigue strength can be seen from the following table, which shows the allowable stress range at 2 x 10^6 cycles for some typical details (based on Ref. 11.2)

plain steel, mill scale	class B	150	N/mm^2
plain steel, flame cut	class C	124	N/mm^2
longitudinal butt weld	class C	124	N/mm^2
longitudinal butt weld	class E	80	N/mm^2
welded cover plate	class G	50	N/mm^2

Certain types of weld are excluded from the classification system. For example, partial penetration transverse butt welds are not classified because of the wide variation in fatigue strength which can occur. In such cases the user is recommended to test a number of specimens at the same stress range, and use the results to select the most appropriate S-N curve from the Standard.

11.3 Fatigue life curves

BS 7608 defines a number of S-N curves for welded joints, as shown in *Figure 11.2*. Each curve is defined on the basis of nominal stress range vs life in cycles, and corresponds to one of the weld classification (B, C, D, etc.).

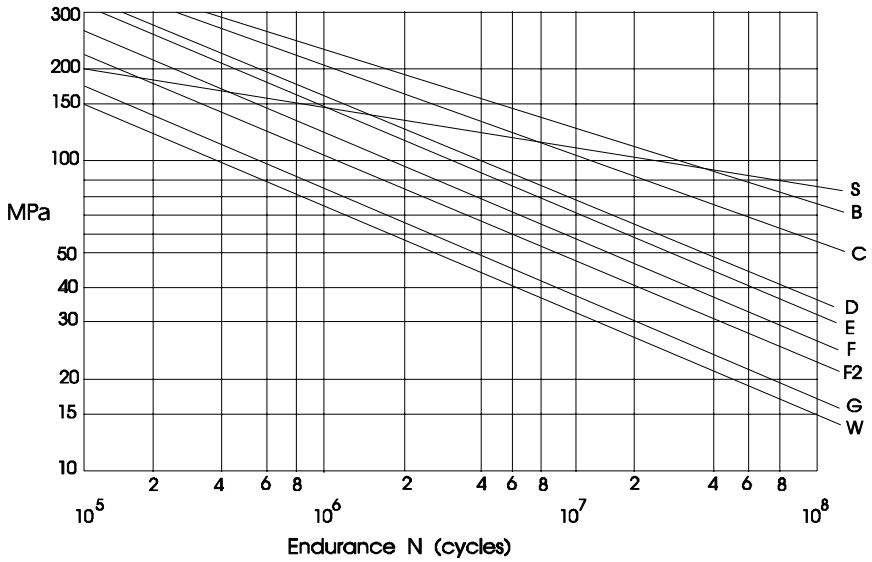


Figure 11.2 S-N curves for welded steel joints

The curves have a constant slope between 10^5 and 10^7 cycles, where the stress-life relationship is defined by the equation (for the mean life):

$$N = \frac{k_0}{S^m} \qquad (11.1)$$

where N is the endurance in cycles

S is the nominal stress range

k_o is the constant for a particular weld classification

m is the slope of the S-N curve on log - log axes

For most curves, m has a value of 3, from the Paris crack growth law.

The curve between 10^5 and 10^7 is defined from experimental test data. The curves were extended for longer lives using theoretical calculation. The life to crack initiation for welded joints is a small part of the total life, as most welded joints contain cracks or crack-like defects produced during manufacture. The life is therefore dominated by the propagation of these cracks. Although the defect may initially be small and therefore not effected by small cycles, the larger cycles present in the applied loading may propagate the defect, and as the defect size increases it will be propagated by smaller cycles. The concept of an endurance limit therefore is not appropriate. The result is that if all the cycles fall below the stress level for 10^7 cycles, the signal can be considered non-damaging. If larger cycles exist, all cycles must be considered, and the S-N curve is extended indefinitely, with the value of m increased to (m + 2).

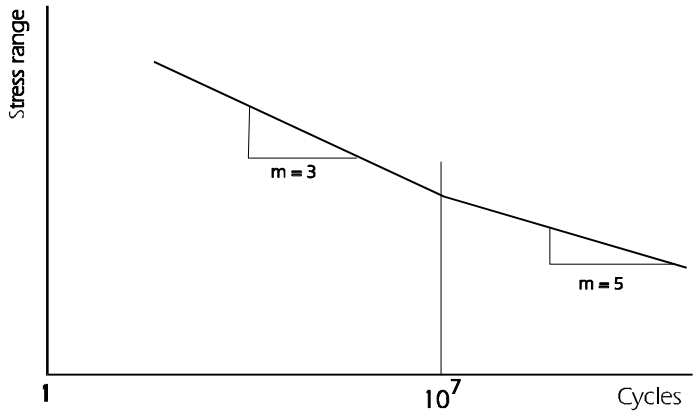


Figure 11.3 Extension of the S-N curve for endurances greater than 10^7 cycles

For very large stress ranges, the curve is extended back at the slope of m until static strength limitations apply.

11.4 Calculation of applied stress

The combination of stress concentrations and pre-existing defects in welds produces a very complex condition, which is not amenable to analytical treatment. The S-N curves defined in 11.2 have been derived on the basis that the stress range will be calculated 'as if the weld did not exist', i.e. the nominal stress calculated by engineering stress analysis. The stress concentration caused by the weld geometry may be ignored, as its effect has been included in the S-N curve.

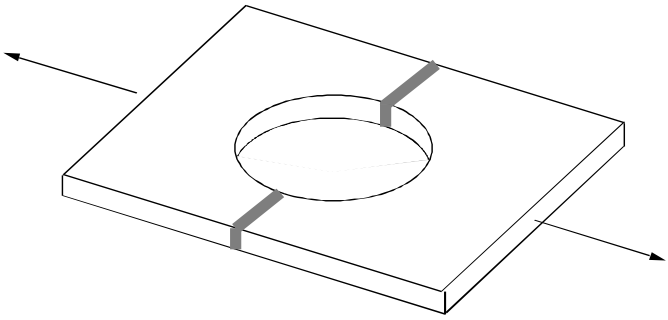


Figure 11.4 An example where an additional stress concentration would be applied

11.5 Effect of mean stress

The data used for BS 7608 showed that the fatigue strength of welded joints is relatively insensitive to changes of mean stress in the signal. There are two reasons for this.

Firstly, from Section 11.1, the fatigue strength reduction caused by a welded joint is very high, and it was shown in section 5 that mean stress effects are less pronounced at high stress concentrations.

Secondly, the welding process produces high tensile residual stresses in the weld toe, which approach the yield stress of the parent material. A cycle at nominally zero mean stress is actually positioned, in 'local stress-strain' terms, as shown in *Figure 11.5* - in effect, all cycles have mean stresses close to the yield stress of the parent plate, whatever the apparent nominal mean stress.

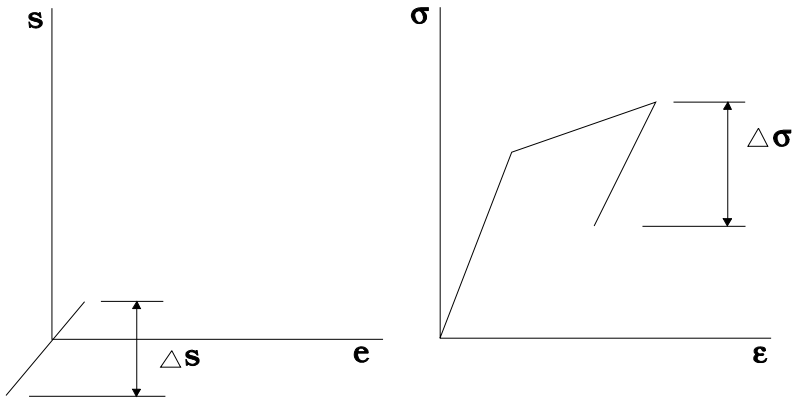


Figure 11.5 Effect of residual stress on apparent mean stress

An important consequence is that compressive cyclic loading is as damaging in fatigue terms as tensile loading. It has been shown experimentally that, in beams loaded in bending, fatigue cracks can initiate in the compressive flange of the beam and propagate across and through the flange. It is not possible therefore, for a designer to 'hide' poor structural details in areas of nominally compressive cyclic stress.

11.6 Effect of material UTS

For most structural steels, the fatigue strength of welded joints is not sensitive to the static strength of the parent plate. Again, there are two main reasons. Firstly, the welding process reduces the effect of any heat treatment or work hardening, which may have been introduced in order to achieve the static properties of the material. Secondly, as was shown in chapter 2, high stress concentrations in themselves reduce any relationship between static strength and fatigue strength. BS 7608 is generally taken to be applicable to welds in structural steels in general.

11.7 Calculation of probability of failure

Welded joints exhibit a significant degree of scatter in their fatigue strength, for nominally similar joints. This is inherent in the welding process, where defects such as blowholes, porosity, slag inclusions, and

weld stop-starts, together with geometric variations such as weld toe radii, are randomly distributed. The resulting variation in fatigue strength can be characterised by a Gaussian distribution. This distribution is defined by its mean value, and by the standard deviation which determines the 'width' of the distribution.

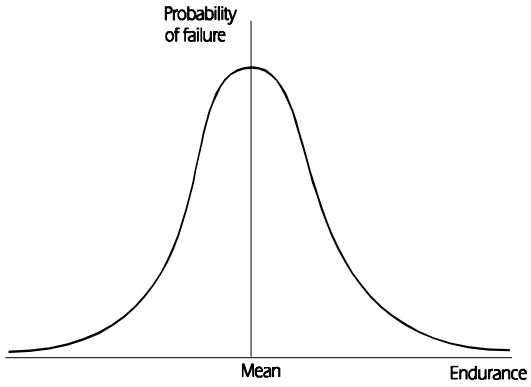


Figure 11.6 Gaussian distribution of endurance

The data used to derive the fatigue rules in BS 7608 were sufficient in quantity for an estimate to be made of the variation in fatigue strength for each class of weld. Equation (11.1), which defines the mean life of the welded detail, can be extend as follows

$$N = \frac{k_o \Delta^d}{S^m} \qquad (11.2)$$

where N is the endurance in cycles
S is the nominal stress range
k_o is the constant for a particular weld classification
m is the slope of the S-N curve on log - log axes
Δ is a function of the standard deviation of the fatigue life,
and d is the number of standard deviations from the mean.

The variable d is termed the 'design criterion', and is expressed as a number of standard deviations from the mean life. Its relationship to probability of failure is shown below. Note that in BS7608 the value of *d* is assumed to be positive if it is below the mean line. However, much commercially available fatigue analysis software uses the convention that negative values of *d* refer to lives below the mean.

d	**probability of failure**
0	50 % (mean life)
1	15.9 %
2	2.2 %
3	0.13 %

A value of d = 2 (for example) implies that 2.2% of components will 'fail' at lives shorter than that calculated from BS 7608. The criterion for failure is not precisely defined. It is generally taken to be a 'large' crack.

11.8 Practical fatigue calculations

BS 7608 recommends Rainflow cycle counting as the method of analysing complex signals. If these signals are obtained from strain measurements then it is important to ensure that truly nominal strains are measured. The stress disturbance caused by welded details can extend far beyond the detail itself, depending on the geometry of the detail. A typical example is shown below. Clearly it is important to ensure that strain gauges are located away from this disturbed area if unduly conservative life calculations are to be avoided.

By their very nature, complex welded structures may produce multiaxial stress states. BS 7608 suggests that if the variation in direction in principal stress does not exceed 45^{o}, the principal stress can be used. BS 7608 gives a full explanation of the treatment of multiaxial stress conditions, and design stresses in general.

Recently, engineers have used critical plane analysis to determine the most damaged orientation, rather than use the treatment of multiaxial stresses specified in the standard.

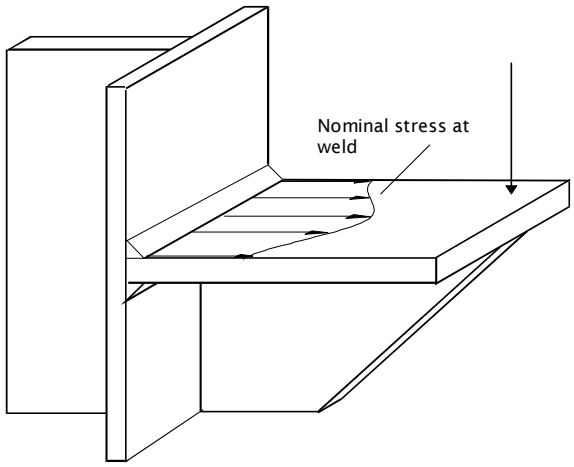


Figure 11.7 Stress concentration adjacent to a welded detail

11.9 Non-welded details

BS 7608 has been generalised to include non-welded structural steel, in plain plate and with various stress concentrations, and non-welded joints. Appropriate detail classifications are provided. Although one can see the benefits of covering all conditions in a single standard, it is apparent that two major assumptions in BS 7608, namely that mean stress is not significant, and that the same S-N curves apply to all structural steels, are major simplifications when applied to plain materials. In these instances the use of local stress-strain analysis is to be recommended.

11.10 References and further reading

11.1 BS 7608

Fatigue design and assessment of steel structures, British Standards Institution, 1993

11.2 Improving the fatigue strength of welded joints.

The Welding Institute, 1983

See also:

GURNEY T R.

Fatigue of Welded Structures.

2nd ed. Cambridge University Press, 1979.

And papers by Gurney T R and Maddox S J, and publications by TWI, Cambridge, England.

11B Appendix to Chapter 11: Equivalent structural stress method (Verity™)

Design codes such as BS7608 relate fatigue life to the nominal stresses at the weld toe, and use an S-N curve that represents the fatigue strength of a particular type of weld or group of welds.

The nominal stress approach is quite useful for simple design calculations, but is difficult to apply when using strain gauges or finite element models. Strain gauges must be placed at a distance from the weld toe in order to obtain stresses undisturbed by the presence of the welded joint. In a finite element model the life must also be calculated some distance from the weld toe. The choice of this distance can be quite subjective.

The choice of S-N curve depends on the assessment of the weld geometry and its relationship to a particular type of weld. This can also be subjective and prone to error.

The Verity™ method was developed by the Battelle Memorial Institute as a means of addressing these difficulties. It was developed for use with finite element models, but can also be extended for use with strain gauges.

The essential features of the method are

1. The stresses at the weld toe are separated into three contributors – membrane stress, bending stress and a self-equilibrating stress.
2. These stresses are calculated from the nodal forces for nodes at the weld toe.
3. The structural stress is calculated from the nodal forces.
4. The structural stress is adjusted for other effects, including thickness, to produce an equivalent structural stress
5. The fatigue strength of all welds can be described by a single curve of stress range versus endurance (an S-N curve), where the stress is the equivalent structural stress. One S-N curve can be used for welded joints in steel, and a separate curve for welded joints in aluminium alloys.

The benefits are

1. The equivalent structural stress is calculated at the weld toe, not at some distance from it.
2. A single S-N curve can be used, so there is no need to 'classify' the welded joint.

11.1 Stresses at the weld toe

The local stress distribution at the weld toe consists of an axial stress σ_x normal to the weld toe, and a shear stress τ_y.

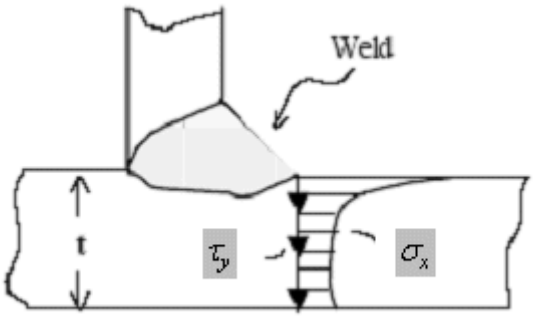


Figure 1 Stress distribution at the weld toe

The stress distribution σ_x is separated into a membrane stress σ_m and a bending stress σ_b. The difference between the actual stress distribution and the sum of the membrane and bending stresses is called the self-equilibrating stress.

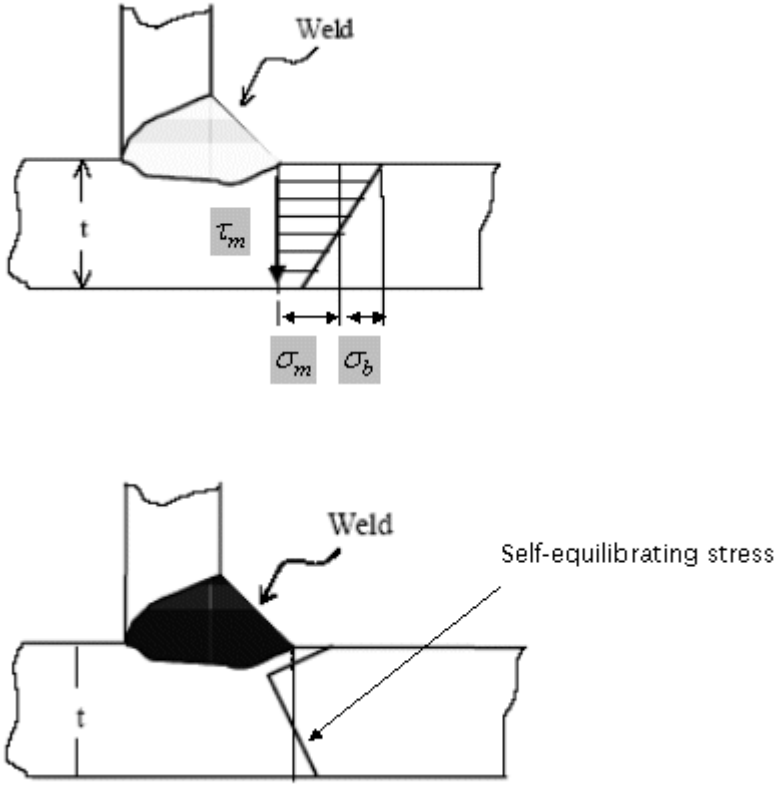


Figure 2 Membrane, bending and self-equilibrating stresses

11.2 Structural stress from nodal forces

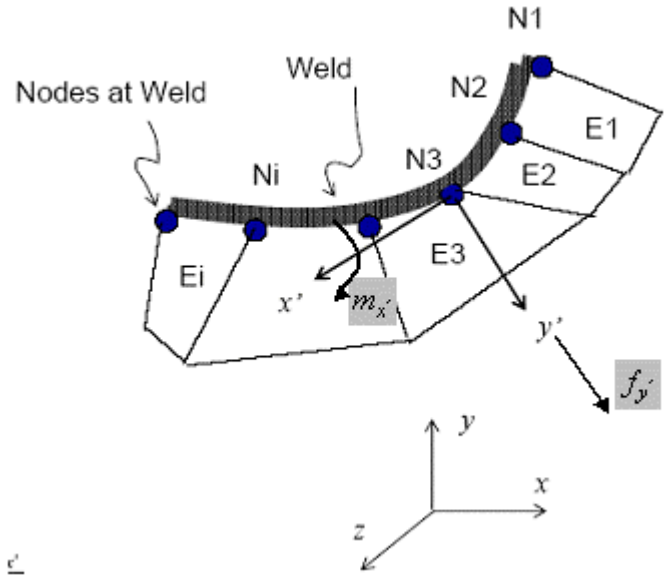


Figure 3 Force notation

The axial force $f_{y'}$ and the bending moment $m_{x'}$ are calculated for each node and include the effects of the forces on adjacent nodes.

The structural stress is the sum of the bending and membrane stresses, calculated from the axial force $f_{y'}$ and the bending moment $m_{x'}$.

$$\text{Structural stress } \sigma_s = \sigma_m + \sigma_b = \frac{f_{y'}}{t} + \frac{6m_{x'}}{t^2}$$

This parameter correlates the fatigue strengths of steel welds quite well. Figure 3 shows the results for a wide range of welds in steels with thickness from 1.5mm to 104 mm, and includes butt, fillet, cruciform and other joint configurations, and also includes spot welds.

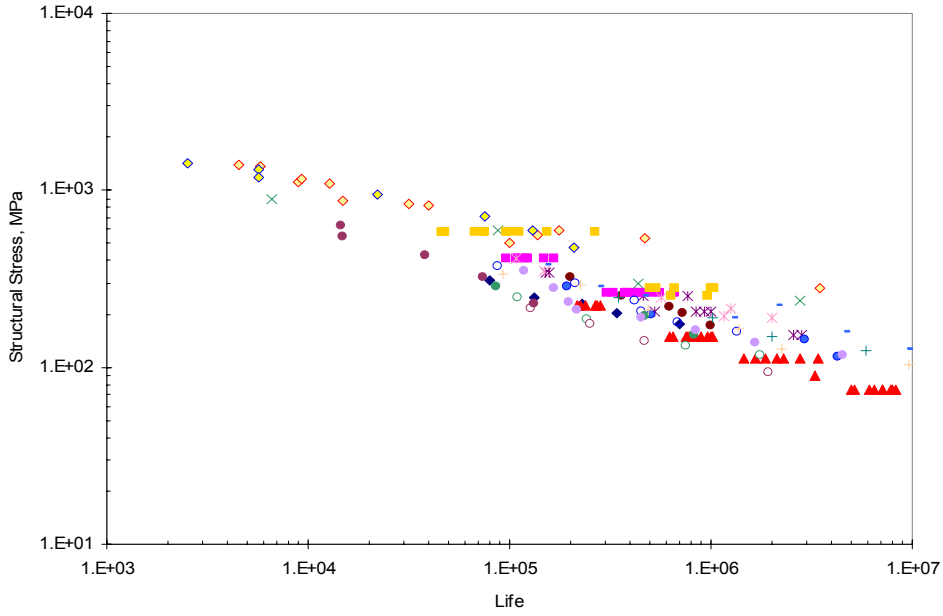


Figure 4 Welded joint test results correlated using structural stress range (life in cycles)

11.3 Equivalent structural stress

The scatter shown in Figure 4 can be reduced by the introduction of two more parameters. The first parameter is a measure of how bending-dominated the structural stress is. The parameter r is the ratio of the bending and structural stress :

$$r = \frac{\sigma_b}{\sigma_s}$$

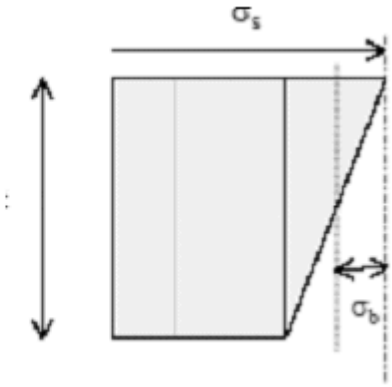


Figure 5 Load parameter r

The second parameter is a plate thickness correction.

The equivalent structural stress is then given by

$$\Delta S_{eq} = \frac{\Delta \sigma_s}{t^{\frac{2-m}{2m}} . I(r)^{1/m}}$$

where t is the thickness in mm, and m is the exponent in the Paris crack growth equation, and is shown to be 3.6 for welded steel and aluminium joints.

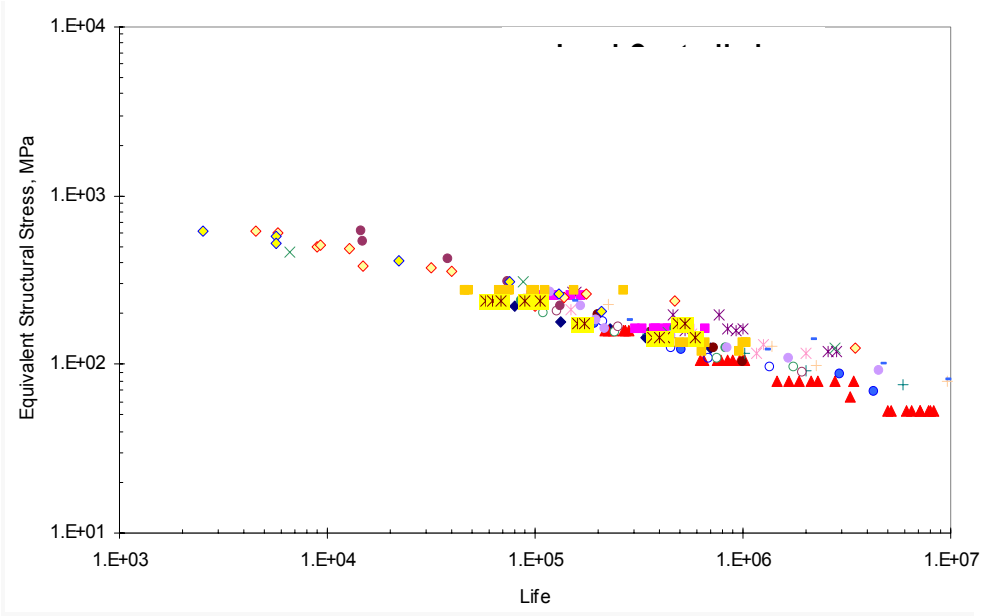


Figure 6 Welded joint test results correlated using equivalent structural stress range

11.4 Effect of finite element mesh and choice of element

The Verity™ method has been shown to be insensitive to mesh density and element type. In essence this is because nodal forces are insensitive to these parameters.

Figure 7 shows the calculated stress concentration (SCF) at the weld toe for different element sizes and element types.

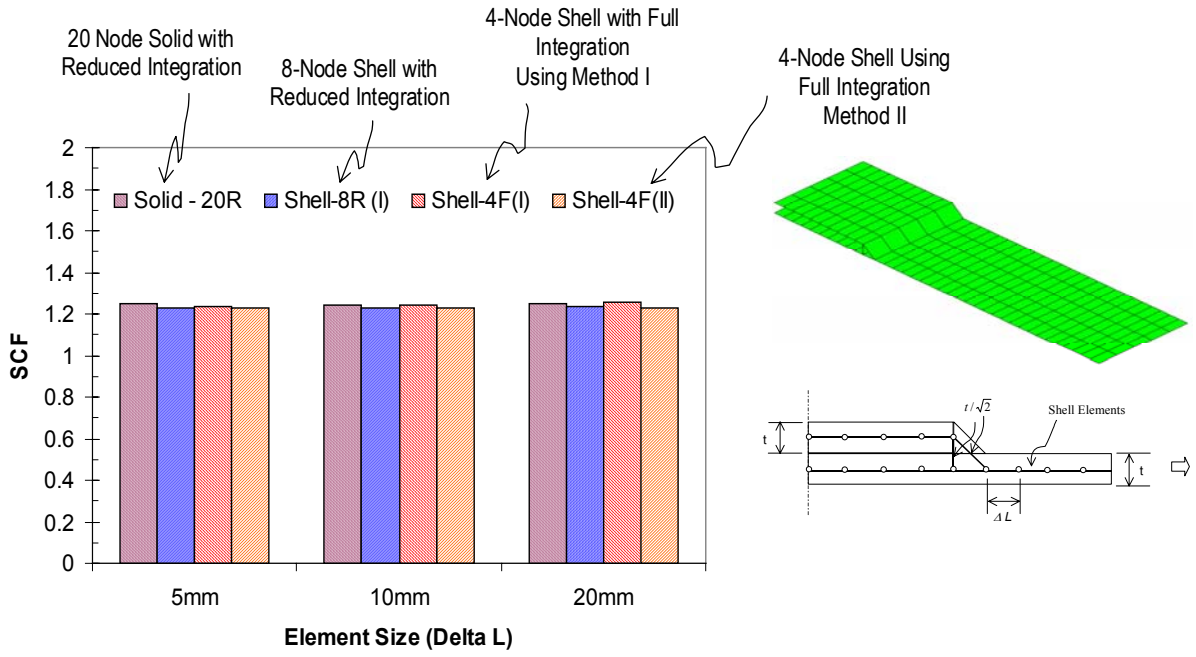


Figure 7 Effect of element size and element type
on the stress concentration (SCF) calculated at the weld toe

11.5 References and further reading

Dong, P.
"A Structural Stress Definition and Numerical Implementation for Fatigue Analysis of Welded Joints,"
International Journal of Fatigue, 23, pp. 865-876, 2001.

Dong, P., Hong, J.K., Cao, Z.
"A Mesh-Insensitive Structural Stress Procedure for Fatigue Evaluation of Welded Structures"
International Institute of Welding, IIW Doc. XIII-1902-01/XV-1089-01, July, 2001.

Dong, P., Hong, J.K., Cao, Z.
"Structural Stress Based Master S-N Curve for Welded Joints,"
International Institute of Welding, IIW Doc. XIII-1930-02/XV-1119-02, June, 2002.

Dong, P. and Hong, J.K.
"Analysis of Hot Spot Stress and Alternative Structural Stress Methods",
Paper No. OMAE2003-37315, Proceedings of the 22nd OMAE International Conference, June 8-13, Cancun, Mexico.

Dong, P.
"A Robust Structural Stress Method for Fatigue Evaluation of Ship Structures,"
Paper No. OMAE2003-37313, Proceedings of the 22nd OMAE International Conference, June 8-13, Cancun, Mexico (also in press, ASME Transaction: Journal of Offshore Mechanics and Arctic Engineering, 2004).

Dong, P. Hong, J.K., Osage, D. Prager, M.
"Assessment of ASME's FSRF Rules for Vessel and Piping Welds Using a New Structural Stress Method,"
Journal of International Institute of Welding (IIW) : Welding In the World, Vol. 47, No. 1, 2003, pp. 31-43.

Dong, P., Hong, J.K., Osage, D., and Prager, M.
"Master S-N Curve Approach for Fatigue Evaluation of Welded Components,"
WRC Bulletin, No. 474, 2002, Welding Research Council, New York, New York.

Dong, P., Hong, J.K., and Cao, Z.
"Stresses and Stress Intensities at Notches: 'Anomalous Crack Growth' Revisited",
International Journal of Fatigue, 25 (2003), pp. 811-825.

Dong, P. and Hong, J.K.
"The Master S-N Curve Approach to Fatigue of Vessel and Piping Welds,"
Welding in the World, Vol. 48, No. 1/2, 2003, pp. 28-36.

Dong, P., Hong, J.K., Potukutchi, R., and Agrawal, H.
"Master S-N Curve Method for Fatigue Evaluation of Aluminum MIG and Laser Welds,"
SAE Paper No. 2004-01-1235, March 2004.

Dong, P, and Hong, J.K.
"The Master S-N Curve Approach to Fatigue Evaluation of Offshore and Marine Structures,"
Paper No. OMAE2004-51324, Proceedings of the 23rd OMAE International Conference, Vancouver, Canada, June 2004.

Healy, B.
"A Case Study Comparison of Surface Extrapolation and Battelle Structural Stress Methodologies,"
Paper No. OMAE2004-51228, Proceedings of the 23rd OMAE International Conference, Vancouver, Canada, June 2004.

Dong, P., and Hong, J.K.
"An Effective Structural Stress Parameter for Evaluation of Multi-Axial Fatigue,"
International Institute of Welding (IIW) Document No. IIW-XIII-2034-04/IIW-XV-1173-04, Osaka, Japan, July 2004.

Dong, P.
"The Mesh-Insensitive Structural Stress and Master S-N Curve Method for Ship Structures,"
Paper No. OMAE-FPSO'04-0021, Proceedings of OMAE Specialty Conference on Integrity of Floating Production, Storage & Offloading (FPSO) Systems, Aug. 30-Sept. 2, 2004, Houston

Healy, B.
"Hot Spot Stress Analysis of a Side Shell Connection Using Surface Extrapolation and the Battelle Structural Stress Method,"
Paper No. OMAE-FPSO'04-0022, Proceedings of OMAE Specialty Conference on Integrity of Floating Production, Storage & Offloading (FPSO) Systems, Aug. 30-Sept. 2, 2004, Houston

Kang, et al.
"Testing and Analysis of Fatigue Behavior in Edge Details,"
Paper No. OMAE-FPSO'04-0025, Proceedings of OMAE Specialty Conference on Integrity of Floating Production, Storage & Offloading (FPSO) Systems, Aug. 30-Sept. 2, 2004, Houston

TX Hong, J.K. and Dong, P.
"Hot Spot Stress and Structural Stress Analysis of FPSO Fatigue Details,"
Paper No. OMAE-FPSO'04-0023, Proceedings of OMAE Specialty Conference on Integrity of Floating Production, Storage & Offloading (FPSO) Systems, Aug. 30-Sept. 2, 2004, Houston

Hyuba, H. and Dong, P.
"Equilibrium-Equilibrium Structural Stress Approach to Fatigue Evaluation of a Rectangular Hollow Section Joint,"
International Journal of Fatigue, 27, pp. 85-94

Dong, P., Hong, J.K., and De Jesus, M.
"Analysis of Recent Vessel Fatigue Data Using the Structural Stress Method in ASME Div. 2 Rewrite,"
to appear in Proceedings of ASME PVP 2005 Conference, July, 2005, Denver, CO

12 Fatigue analysis from the PSD

12.1 Introduction

The power spectral density diagram, or PSD, contains a description of the frequency content of a signal. The signal in *Figure 12.1* is constructed by adding together sine waves of various amplitudes and frequencies. The result is a close approximation to a square wave. Any signal that repeats at regular intervals can be constructed by adding together sine and cosine waves of different amplitudes, phases and frequencies. By implication, any measured signal can be considered as one repeat of a signal that would have been repeated if measurement had been longer.

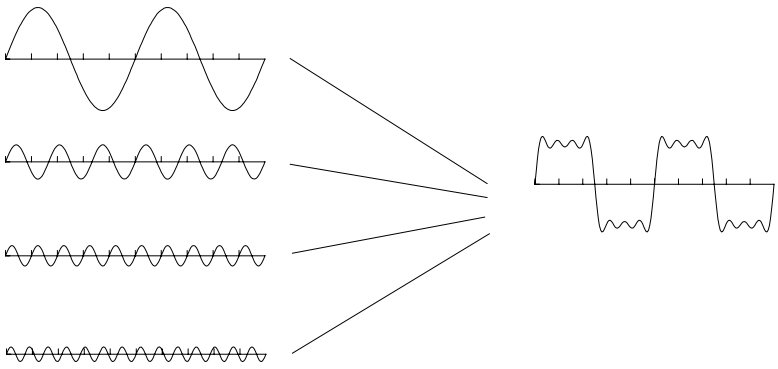


Figure 12.1 Signals are the summation of sine waves

Frequency domain analysis is essentially the reverse of this signal construction process, in that it takes a given signal and determines the amplitudes, phases and frequencies present in the signal.

The result can be expressed as a diagram which shows the amplitudes present at a number of different frequencies. In fact the result is usually plotted as the square of the rms amplitude - the variance - and it is usual to scale the diagram by dividing the vertical axis by the interval between the frequencies, to produce a density diagram. This is the power spectral density diagram, or PSD. It is sometimes referred to as auto-spectral density, or ASD.

For the synthesised square wave, the PSD would be as shown in *Figure 12.2*.

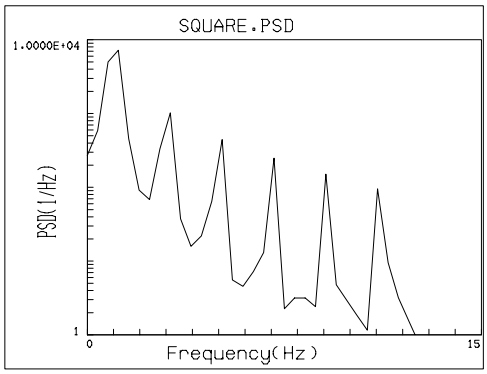


Figure 12.2 PSD of a synthesised square wave

Some of the properties of a signal y(t) are:

MAXIMUM VALUE The maximum value in the signal y_{max}

MINIMUM VALUE The minimum value in the signal y_{min}

MEAN VALUE The mean or average value in the signal $\overline{y}$

For a signal *y(t)* the mean value is

$$\overline{y} = \frac{1}{T}\int_0^T y(t).\, dt$$

MEAN SQUARE VALUE The mean square value is the average value of $y^2(t)$

$$\text{mean square} = \overline{y^2} = \frac{1}{T}\int_0^T y^2(t).\, dt$$

ROOT MEAN SQUARE (rms) The rms is the square root of the mean square value.

$$rms = \sqrt{\overline{y^2}}$$

STANDARD DEVIATION The standard deviation σ is defined by

$$\sigma^2 = \frac{1}{T}\int_0^T (y(t) - \overline{y})^2.\, dt$$

where σ^2 is the VARIANCE of the signal.

The standard deviation is the square root of the average value of the square of (the signal minus the mean value of the signal), and shows the extent to which the signal fluctuates about the mean value.

For digitised signals -

the *mean value* is the average value of the samples in the signal,
the *mean squared value* is the average value of the square of each sample,
the *variance* is the average value of the square of each term
(sample value minus signal mean value),
the *standard deviation* is the square root of the variance.

A PSD may be analysed to provide a number of parameters which describe the properties of the signal. For example, integrating the PSD provides the variance of the signal, and the square root of the variance gives the root mean square (rms) of the signal.

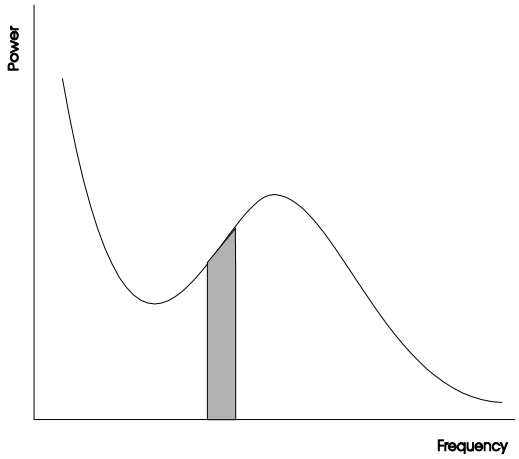


Figure 12.3 Area under the psd is the variance of the signal

Integrating the PSD between any two frequencies and taking the square root gives the rms of the signal between these two frequencies. Taking moments of area of the PSD about the zero Hz axis yields further information.

Figure 12.4 defines the term $m_n = \int Ax^n$

where m_n is the *n*th moment of the PSD about the zero Hz axis.

The rms is therefore obtained from the *zero*th moment, ie.

$$\text{rms} \quad \sigma = (m_0)^{\frac{1}{2}} \qquad (12.1)$$

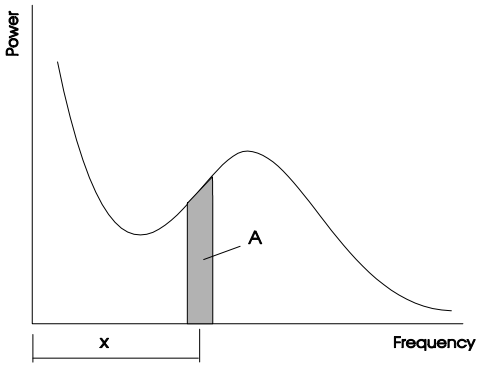


Figure 12.4 Definition of moments of the PSD

The number of positive slope zero crossings per unit time is the number of times the signal crosses zero with a positive slope, per unit time:

$$\lambda_0 = \left(\frac{m_2}{m_0}\right)^{\frac{1}{2}} \qquad (12.2)$$

The number of peaks per unit time is

$$\mu = \left(\frac{m_4}{m_2}\right)^{\frac{1}{2}} \qquad (12.3)$$

The irregularity factor of a signal is defined as the number of positive slope zero crossings divided by the number of peaks. It is calculated from the PSD using above equations.

Clearly, although the PSD is obtained by a frequency analysis, it can be analysed to provide information about the amplitudes present in the signal. As fatigue damage is determined by these amplitudes, the implication is that the PSD contains information, or can be analysed to reveal information, that is relevant to fatigue life calculation.

Many naturally occurring signals have an important property in that their probability density closely approximates to a Gaussian distribution. In a probability density diagram the area between any two signal values shows the probability that a data point in the signal will lie between these two signal values.

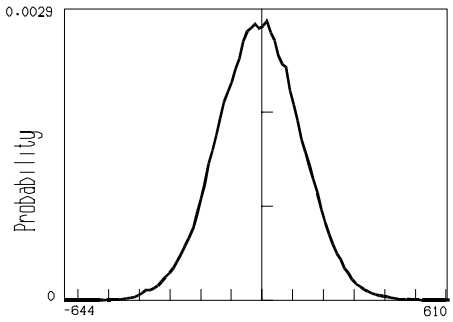


Figure 12.5 Probability density distribution

The probability density distribution for a Gaussian signal is defined by

$$p(y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(y-m)^2/2\sigma^2}$$

where m is the mean value of the signal

σ is the standard deviation

and y is a value in the signal

Examples of Gaussian processes are wind speed, which influences the loading on wind turbine generators and aircraft components, and wave height, which influences the loading on off-shore structures.

12.2 Why fatigue analysis from PSD's?

Finite element analysis is very efficient in taking the PSD of applied loads and determining the PSD of the resulting stresses at various points in the structure. A method of taking the PSD of stress and calculating fatigue lives therefore has attractions. In fact, much of the early work on fatigue analysis from PSD's was carried out by NASA in order to determine the fatigue damage caused by vibration and buffeting of space vehicles.

12.3 Early methods of fatigue analysis from the PSD

The simplest PSD contains just a single frequency. As all the frequencies fall within a narrow band, this is a 'narrow band' signal. If it is obtained from a Gaussian signal, then the signal will be as shown in *Figure 12.6*.

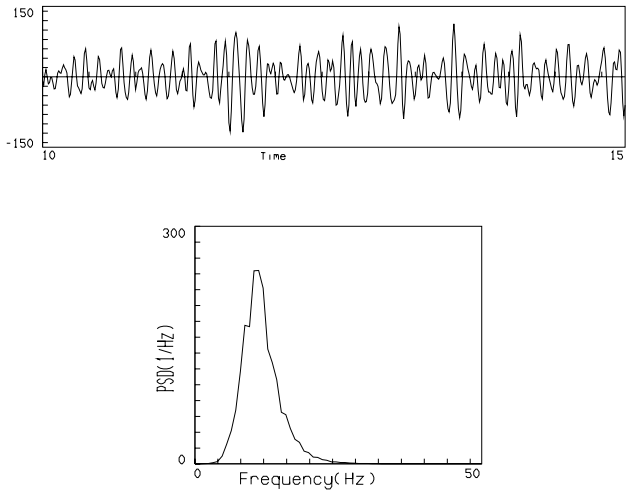


Figure 12.6 Narrow band signal and its PSD

For this signal, each positive slope zero crossing is followed by a peak. From equations 12.2 and 12.3, the irregularity factor is therefore equal to 1.0. As the data points in the signal are defined by the Gaussian distribution, the distribution of peaks can also be calculated. From *Figure 12.6* it seems that each peak is followed by an approximately equal and opposite valley. The ranges in the signal, which would be obtained from a simple range counting analysis (chapter 4), can therefore be determined. In early attempts to calculate fatigue damage from PSD's it was assumed that these ranges were approximately equal to Rainflow cycles in magnitude and number of occurrences - in other words the Rainflow range histogram was approximately equal to the histogram of simple ranges. Hence the fatigue damage could be estimated from this histogram using an S-N curve or strain-life curve.

For the special case of a narrow band signal, this method is quite accurate. However, real signals are not usually narrow band. *Figure 12.7* shows a broad band signal and its PSD.

Clearly in this case each peak is not preceded by a positive slope zero crossing, nor is each peak followed by an equal and opposite valley. The method described for a narrow band signal will not be accurate for this broad band signal.

From equations 12.2 and 12.3 the ratio of the number of zero crossings to the number of peaks is given by the irregularity factor. It was proposed that this could be used as a correction factor. If the signal was first assumed to be narrow band, its fatigue damage could be calculated, and the result corrected using the irregularity factor, which can be calculated from the moments of the PSD. In order to obtain the correction factor, it was necessary to assume a slope for the S-N curve to be used. This method was widely investigated, and a number of different correction factors were proposed. In general, the results were not of sufficient accuracy or general applicability, because the correction factor was fixed to a particular slope of S-N curve, and because the method assumes that Rainflow ranges can be approximated by simple ranges. As the method gives a fatigue life but no intermediate information, it was difficult to be satisfied that when a good life prediction was obtained, this was not an accidental instance of various errors cancelling each other.

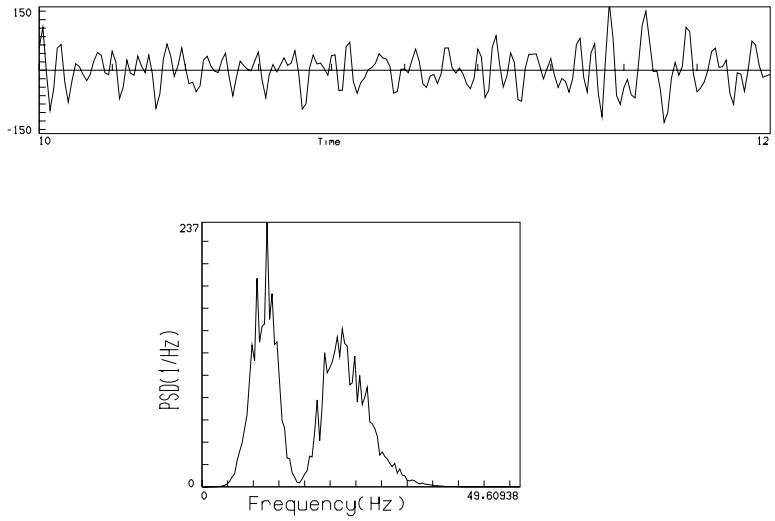


Figure 12.7 A broad band signal and its PSD

12.4 Calculation of rainflow ranges from a PSD

A significant advance in fatigue analysis from PSD's was made by Sherratt and Dirlik (Refs. 12.1, 12.2). Using a large number of idealised PSD's, and signal generation software, they derived an empirical expression for both the distribution of simple ranges and the distribution of Rainflow ranges. This was a major advance, because for the first time the method could be assessed for accuracy by comparing cycle distributions - it was not necessary to make assumptions about S-N curve slopes. The empirical relationship was later validated analytically by Bishop and Sherratt (Ref. 12.3)

In outline the method adopted by Dirlik proposed that the distribution of Rainflow ranges was a sum of three distributions, an exponential function and two Rayleigh functions. It was proposed that the probability density distribution of Rainflow cycle ranges was given by

$$P_{rr}(z) = c_1\left(\frac{1}{\tau}\right)e^{\frac{-z}{\tau}} + c_2\left(\frac{z}{\alpha^2}\right)e^{\frac{-z^2}{2a^2}} + c_3\, z\, e^{\frac{-z^2}{2}}$$

where $c1 + c2 + c3 = 1$

$$c_1 = \frac{2(x_m - \gamma^2)}{1 + \nu^2}$$

$$\alpha = \frac{\gamma - x_m - c_1^2}{1 - \nu - c_1 + c_2}$$

$$c_2 = \frac{1 - \gamma - c_1 + c_1^2}{1 - \alpha}$$

$$\tau = \frac{1.25}{c_1}(\gamma - c_3 - c_2)$$

z is a normalised parameter for the cycle range, and is obtained by dividing the cycle range x by 2σ.

The constants in these equations are

σ = the rms of the signal (or the standard deviation if the signal has a non-zero mean value)

$$\gamma = \frac{\lambda_0}{\mu}$$

λ_0 = the number of positive slope zero crossings (equation 12.2)

μ = the number of peaks per unit time (equation 12.3)

and $$x_m = \left(\frac{m_1}{m_0}\right)\left(\frac{m_2}{m_4}\right)^{\frac{1}{2}}$$

This method has been tested against measured signals from wind turbines (Ref.. 12.4). Measured strain gauge signals from the blade of an experimental wind turbine were analysed using cycle-by-cycle fatigue analysis. The PSD's of the signals were calculated, and the Sherrat-Dirlik equation used to generate cycle histograms. The fatigue lives calculated from the signals and from the PSD-generated cycle histograms agreed to within 10% even when the measured signals were quite non-Gaussian. The predicted distributions of Rainflow cycles and fatigue damage also agreed closely with those obtained from the signals.

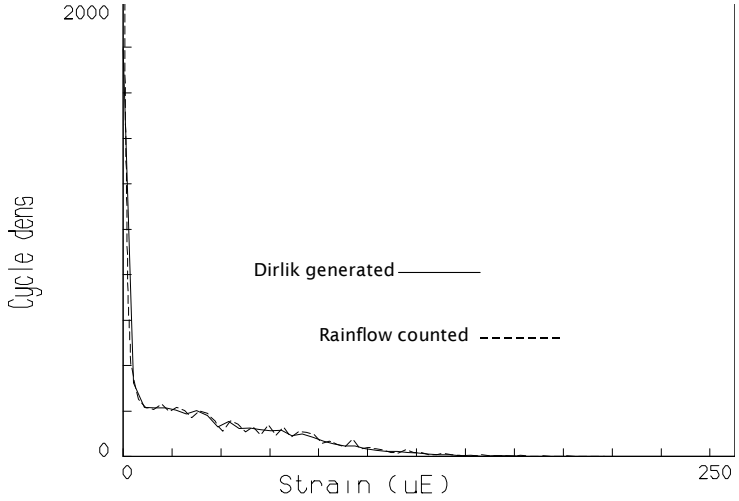


Figure 12.8 Comparison of cycle densities calculated from a signal and from its PSD.

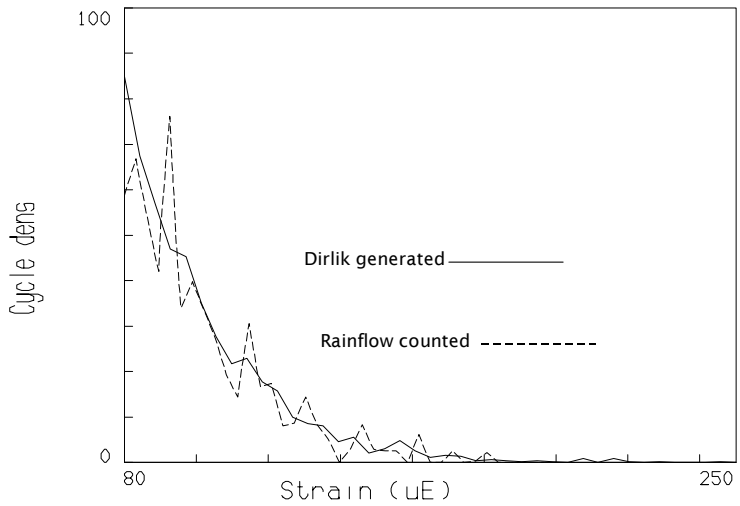


Figure 12.9 Comparison of cycle densities - largest cycles

It has also be argued that a fatigue damage calculation from a PSD may be superior in estimating the damage contribution from infrequent events, which in a measured signal may only be captured by chance.

The Sherrat-Dirlik method has been incorporated into a standard for the fatigue assessment of off-shore structures. The algorithms are being extended (Ref. 12.5) to signals which are non-Gaussian.

12.5 References

12.1 Dirlik T

Application of Computers in Fatigue

PhD Thesis, University of Warwick, England, 1985

12.2 Sherratt F

Current Applications of Frequency Domain Fatigue Life Estimation

Proc EIS 95 'Product Optimisation for Integrity' , Engineering Integrity Society, Sheffield 1995

Editors Cawte E R, Draper J, Trigwell N

12.3 Bishop NWM, Sherrat F

A Theoretical Solution for the Prediction of Rainflow Ranges from Power Spectral Density Data

Fat and Fract of Eng Mats and Structures, 13, pp311-326, 1990

12.4 Hoskin RE, Warren JG, Draper J

Prediction of Fatigue Damage in Wind Turbines

Proc 11th International Wind Energy Conference, British Wind Energy Association, 1989

12.5 Bishop NWM, Sherrat F

Fatigue Life Prediction from Power Spectral Density Data, Part 2: recent developments

Environmental Engineering, Vol 2, No 2, June 1989

13 Fatigue test signals and cumulative damage

13.1 Introduction

Modern servo-hydraulic test systems allow realistic fatigue testing of components under representative service loads. It is widely accepted that tests using recorded service loads, or using service loads obtained by computer-regeneration, provide the most accurate method of determining component life. However, this type of testing is time-consuming and expensive, and there are continuing pressures on engineers to accelerate and simplify test procedures. This chapter considers various methods of fatigue testing, and methods of accelerating fatigue tests.

13.2 Constant amplitude tests

The simplest fatigue tests use constant amplitude sinusoidal loading and it has been estimated that more than 99% of all fatigue tests use constant amplitude loading. This is unfortunate, because the results of constant amplitude tests can be very misleading.

Constant amplitude tests may not identify the failure locations. *Figure 13.1* shows the failure locations obtained from service load tests on a fabricated panel from an aircraft.

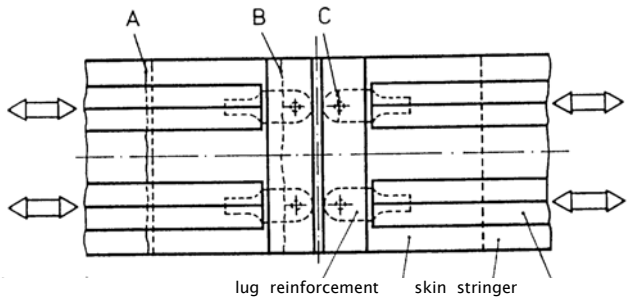


Figure 13.1. Failure locations in an aircraft wing joint

The service load tests showed that there were three potential failure locations (A, B and C). Constant amplitude tests on the same panel reproduced only failure location A. In general, constant amplitude tests cannot be relied on to identify the fatigue-critical areas which require re-design or which require regular inspection in service. Constant amplitude tests may also identify different failure locations at different test amplitudes, as shown by the tests on splined shafts *(Figure 13.2)*.

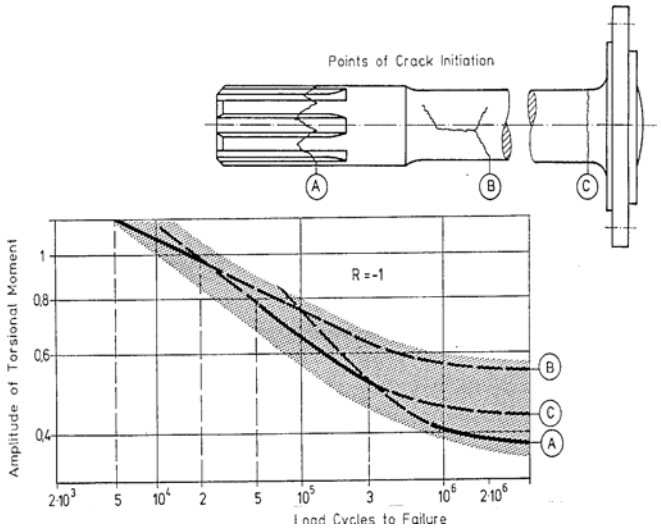


Figure 13.2. Failure locations produced at different test amplitudes

Constant amplitude tests may not identify the best material or manufacturing process.

Tests on riveted lap joints in an aluminium alloy sheet material ranked three different rivet materials (A, B, and C) using constant amplitude and flight-by-flight tests. The results are shown in Table 1.

	Constant Amplitude	Flight by Flight
Best	C	A
	A	B
Worst	B	C

Table 13.1 Ranking of design data using constant amplitude and variable amplitude tests

Tests to compare different materials can be similarly misleading. *Figure 13.3* shows S-N curves for two different materials. If a large amplitude signal is used, material B will be shown to be superior to material A. If a small amplitude signal is used, material A will be shown to be superior. In addition, in this example the small amplitude signal is below the constant amplitude endurance limit, which may also give misleading results. Re-scaling real service loading signals to accelerate a test may give similarly misleading comparisons.

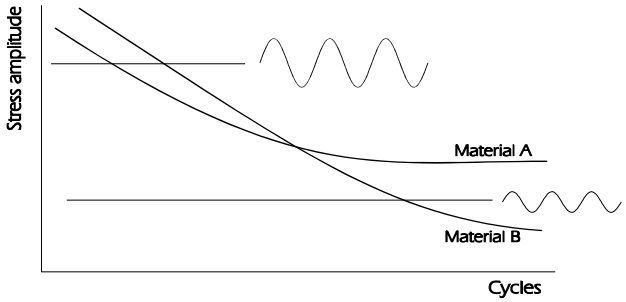


Figure 13.3 Constant amplitude tests on to two materials

Constant amplitude tests may give misleading fatigue lives.

Few components experience constant amplitude loading in service - real service loading contains many different amplitudes and different mean stresses. Cumulative damage rules and mean stress correction methods must be used to estimate service lives from constant amplitude tests, and these introduce errors and uncertainty. Some of the uncertainties in using cumulative damage rules (Miner's rule) are described in this chapter.

Constant amplitude tests increase the scatter in test results.

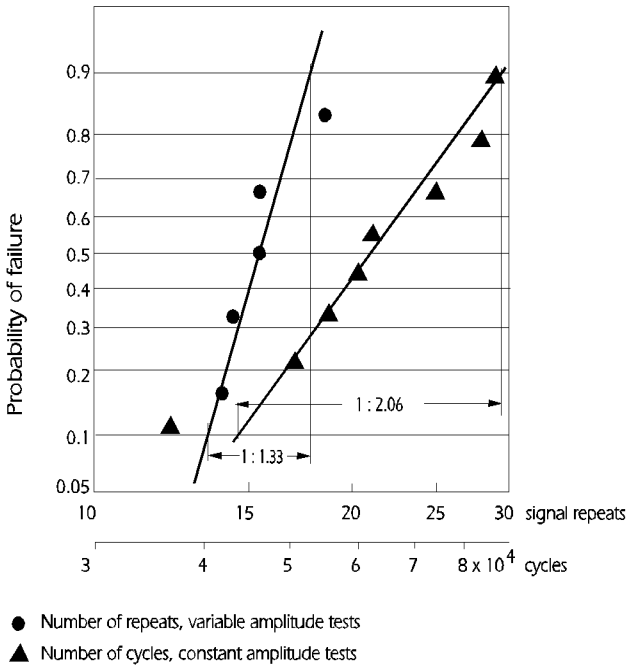


Figure 13.4 Scatter in test lives for constant amplitude and variable amplitude loading

Figure 13.4 compares the scatter in the fatigue life obtained from tests on nominally identical components tested under constant amplitude loading and under representative service loading. The horizontal axis covers a factor three on life.

The standard deviation of the test lives under constant amplitude loading is much greater. This means that whilst a single constant amplitude test may be quicker and cheaper, it will be necessary to test many more specimens to obtain a valid estimate of the mean fatigue life. In addition, safety factors estimated from the scatter on constant amplitude test lives might be excessively conservative.

13.3 Block loading test programs

Service loads can be Rainflow cycle counted to produce histograms of load cycles. These cycles can be grouped into blocks, each with a different constant amplitude, to form a fatigue test signal, and the sequence of blocks can be repeated until failure. The blocks of cycles can be grouped in many different ways, or the blocks could be applied in random order. *Figure 13.5* shows the effect on fatigue life of different types of block loading. The lives have been normalised so that a test using the original variable amplitude signal has a life of 1.0.

Most of the block loading patterns produce lives which are significantly longer than the lives obtained from tests using service signals. In addition, if the cycle amplitudes in the block loading patterns are re-scaled to accelerate the tests then the errors shown in *Figure 13.3* will be added to these errors.

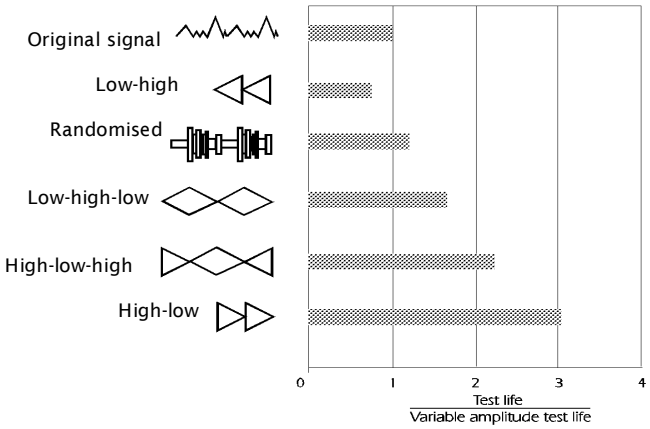


Figure 13.5 Test lives produced by different block loading programmes

The notation is

Low-high	small amplitude cycles applied before larger amplitude cycles
High-low	large amplitude cycles applied before smaller amplitude cycles
Randomised	blocks of cycle amplitudes applied in a random order, etc.

13.4 Editing signals to remove small cycles

Measured service signals contain many cycles of different ranges. A computer-based fatigue analysis may show that most of these cycles contribute no fatigue damage, and could therefore be excluded from a test command signal without effecting the test result. Removing small cycles from a test signal is called 'gating', and can be carried out using computer software *(Figure 13.6)*

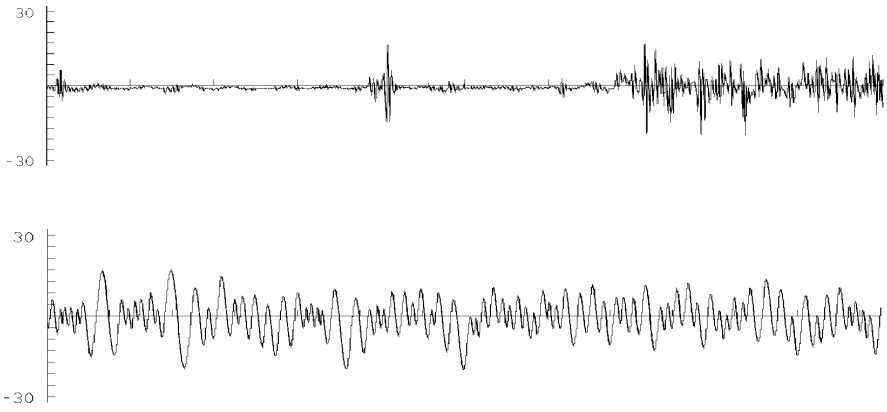


Figure 13.6 Original and gated test signals

Figure 13.7 shows a Rainflow cycle histogram from a measured service signal, and the corresponding computer estimate of the damage histogram.

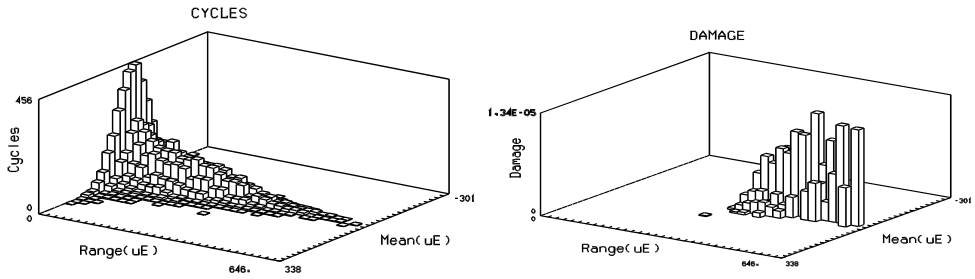


Figure 13.7 Calculated cycle and damage histograms

This type of histogram can be converted to a range-only diagram, instead of a range-mean histogram, as explained in Section 4. In *Figure 13.8* the analysis results have been plotted as cycle-density and damage-density versus cycle range. The area under the curve between any two cycle-ranges shows the total number of cycles between these ranges, or the total calculated damage between these ranges.

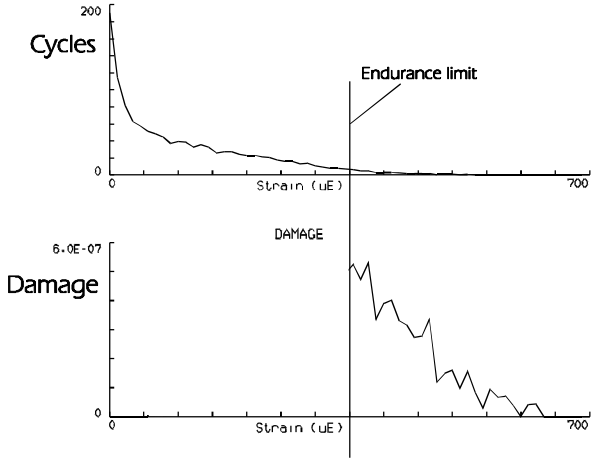


Figure 13.8 Cycle and damage range-only density diagrams

These calculations were based on constant amplitude S-N curves or strain-life curves which showed an endurance limit. For strain-life calculations on conventional steels, an endurance limit of 2 x 10^7 reversals is sometimes assumed, and for S-N curves for steels it has been accepted that the endurance limit will be at approximately 10^7 cycles.

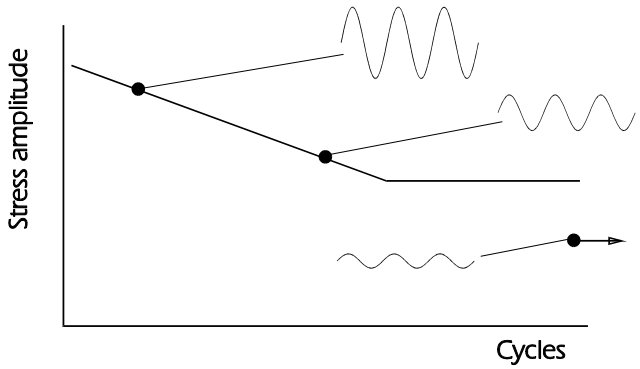


Figure 13.9 S-N curve showing an endurance limit

S-N curves and strain-life curves are obtained by testing a number of specimens, each under constant amplitude loading. For these tests it is clear that many materials do exhibit a clear endurance limit. However, there is increasing evidence that the endurance limit is a result of constant amplitude testing - it is the *constant amplitude endurance limit*. Under complex service load histories the endurance limit may be much less evident, or may not exist at all.

13.4.1 Welded joints using BS5400

In BS7608 (Ref. 13.1), the basic fatigue damage curves for welded joints take the form shown in *Figure 13.10*. However, BS7608 contains an important addition, in that it specifies that if all the cycles in the service history are smaller than the endurance limit, then the service history may be assumed to be non-damaging and the weld will have an 'infinite' life: but if there are cycles in the service history which are larger than the endurance limit, then the constant amplitude S-N curve must be modified as shown in *Figure 13.11*.

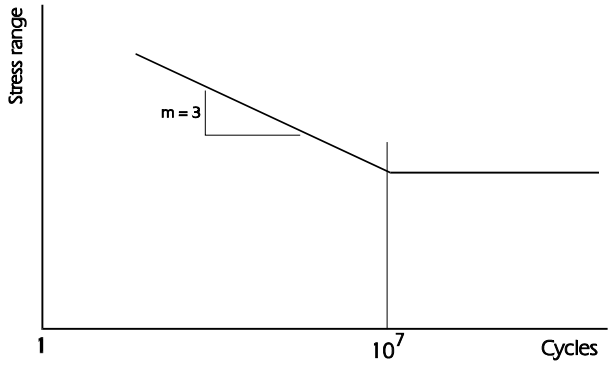


Figure 13.10 S-N curve for a welded joint

The modification removes the endurance limit and substitutes a modified S-N curve which has no endurance limit at all. All cycles, however small, must be assumed to cause fatigue damage. An example of the effect of this modified S-N curve is shown in *Figure 13.12* for a measured strain history.

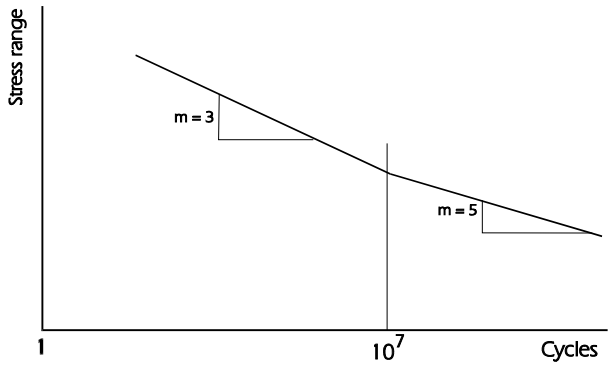


Figure 13.11 S-N curve modified for analysis of variable amplitude signals

The cycle and damage density diagrams are plotted against cycle range, and show that for this signal about half the total damage is caused by cycles which are below the constant amplitude endurance limit. This is illustrated in *Figure 13.13*, which shows the integrated damage density diagram.

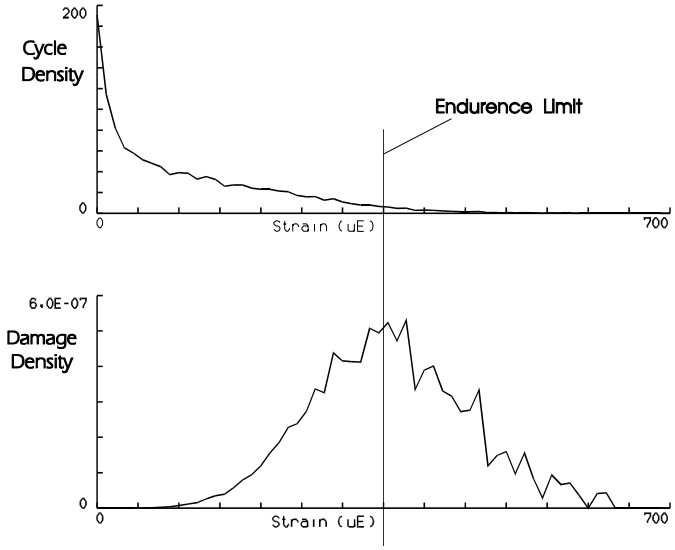


Figure 13.12 Cycle and damage density diagrams (BS5400 data)

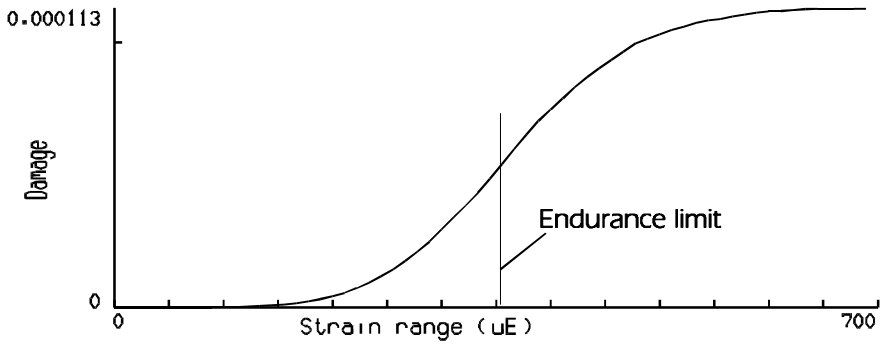


Figure 13.13 Integrated damage density diagram from Figure 13.12

The vertical axis in *Figure 13.13* is the total damage produced by cycles smaller than any chosen strain range. The diagram shows that in this particular case approximately 50% of the total damage is caused by cycles smaller than the endurance limit. Omitting test cycles below the endurance limit would therefore produce about twice the life compared with a test using the complete signal. Other signals combined with other S-N curves would produce greater or less proportions of damage from the cycles below the endurance limit.

The physical explanation proposed for the modified S-N curve in BS7608 is that the welding process produces a crack-like defect in the weld or at the weld toe. Early in the life of the weld, the defect is too small to be propagated by the small cycles. However, the larger cycles will cause the defect to propagate and eventually its size becomes sufficient to allow the small cycles to contribute to the crack propagation process - the small cycles then become damaging.

13.4.2 Effect of small cycles on non-welded components

The effect of small cycles on smooth specimens and machined components has been investigated by a number of researchers (Ref. 13.3), using aluminium alloy and steel specimens. The test program used by Topper et al (Ref. 13.4) consisted of a number of small cycles into which a larger cycle was inserted at regular intervals. The small cycles were below the constant amplitude endurance limit.

If the small cycles are non-damaging, then the fatigue life of the specimen would be determined by the large cycles, and the same test life would be obtained however many small cycles were included in the test program. This proved not to be the case, and the test lives reduced as more small cycles were included.

Effective S-N curves for analysis of variable amplitude signals can be produced by testing components using a large cycle followed by varying numbers of small cycles. The damage caused by the large cycle

can be calculated from a constant amplitude S-N curve, and so the damage caused by the small cycles can be deduced.

$$D_{SC} = D_{total} - D_{LC} \qquad (13.1)$$

where D_{total} is the total damage

D_{LC} is the damage from the large cycle

and D_{SC} is the damage calculated for the small cycles

Depending on the number of small cycles used, a family of S-N curves can be developed, the most severe being for the case where a larger cycle is followed by only a few small cycles (Ref. 13.4).

Figure 13.15 shows equivalent S-N curves (for the small cycles) for aluminium and steel specimens. The parameter η is the number of small cycles inserted between each large cycle *(Figure 13.14)*.

Although the small cycles are below the constant amplitude endurance limit, the test lives are reduced dramatically by their insertion into the command signal. The 'large' cycles in these tests were similar in magnitude to the monotonic yield stress.

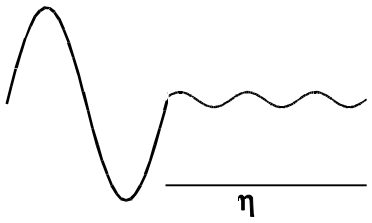
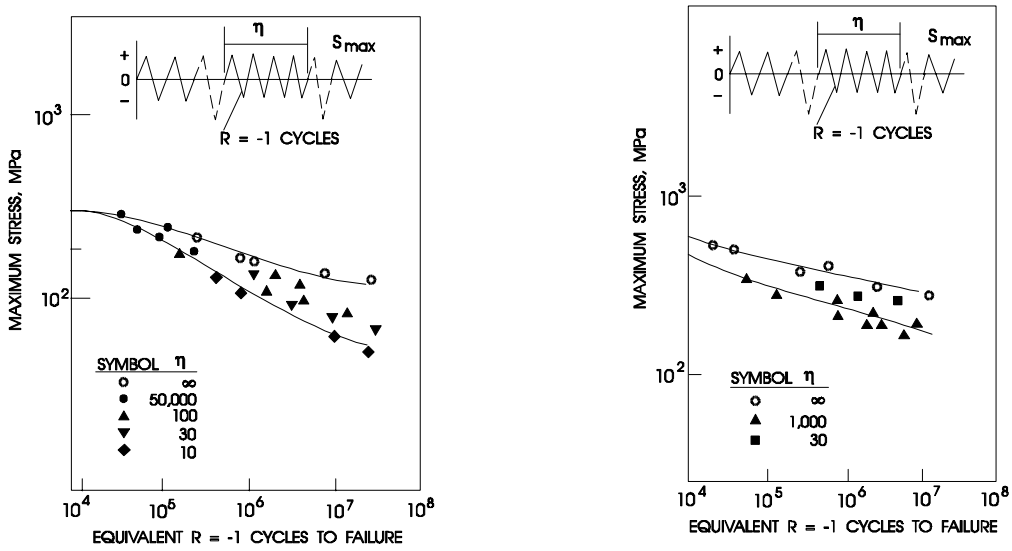


Figure 13.14 Parameter η is the number of small cycles between each large cycle

The 'small cycle' damage is illustrated in *Figure 13.16*, for different values of η. It can be seen that following a large cycle (applied at point O), the first few small cycles are the most damaging, and then as more small cycles are applied these further cycles cause less damage. Eventually, the damage caused by additional small cycles reduces to a constant value, or to zero if the cycles are below the constant amplitude endurance limit.

The physical explanation proposed for this effect is that the larger cycles reduce the force required to open a micro-crack. This allows the small cycles to propagate the crack. Once the crack has propagated outside the affected area, cycles below the constant amplitude endurance limit cannot propagate the crack, and cycles which are small but not below the constant amplitude endurance limit return to their original damage contribution.



Aluminium 2024-T351 Steel SAE1045

Figure 13.15 Effective damage curves for aluminium alloy and steel specimens

It is quite evident from these and other tests that small cycles, when mixed with large cycles, are very damaging. Omitting them from a test signal could cause serious errors in the test lives by producing much longer lives than would be obtained from the complete signal. From the examples above, test lives with signals which have been edited to remove all cycles below the constant amplitude endurance limit could be non-conservative by factors of 10 or more.

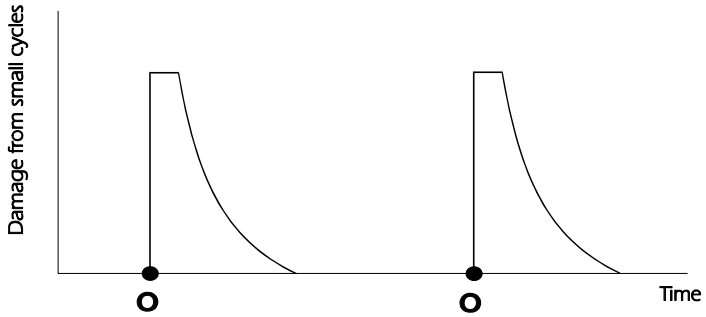


Figure 13.16 Damage for the small cycles plotted against time

This result does not just apply to fatigue testing. A computer-based analysis would also show that the small cycles were non-damaging, so calculated lives would be non-conservative by similar factors. The error clearly results from the use of the constant amplitude S-N curve or strain-life curve with its associated endurance limit.

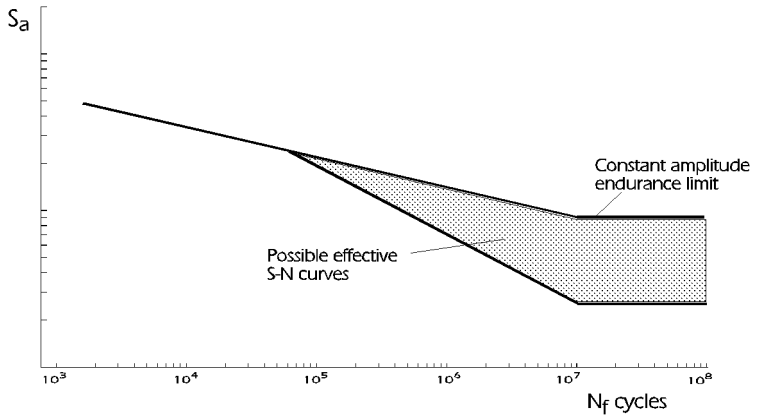


Figure 13.17. possible S-N curves for analysis of variable amplitude signals

From *Figure 13.15*, the average damage for each small cycle is greatest when only a few small cycles follow each larger cycle. In many service signals there are long sections where smaller cycles are intermingled with larger cycles, and something close to this 'worst-case' situation may apply. The endurance limit for these effective S-N curves can be as little as 25% of the constant amplitude endurance limit *(Figure 13.17)* (Ref. 13.4)

A method of estimating the damage curve for analysis of variable amplitude signals has been proposed (Ref. 13.4), and it may be advisable in the future to produce generic corrections to constant amplitude S-N curves for different industries.

There is evidence from tests on welded and machined components (to be published) that narrow band signals give Miner's summations close to 1.0, and that Miner's summation reduces as the signal becomes broader band. In the absence of other information, an approximate 'rule of thumb' is that only cycles with amplitudes less than 25% of the constant amplitude endurance limit should generally be considered for removal. It may be possible to relax this 'rule' if a signal contains long sections where all cycles are well below the endurance limit, as only the first few cycles will contribute significant damage. However, it must be remembered that these cycles could contribute to other factors which effect the fatigue life, such as the development of fretting debris. Also, signal gating which resulted in a test signal which was almost constant amplitude *(Figure 13.18)* could produce many of the errors associated with constant amplitude testing.

The assessment of non-damaging cycles is carried out on the basis of stress or strain cycles at the critical locations in the component, whereas the signal which is being gated is usually a load command signal. The gate level calculated in terms of stress or strain must be translated into an equivalent load value, and this must be carried out for each potential failure location. The same cautionary note applies if mean stress effects are included in assessing which cycles are non-damaging, as the mean stress in a fatigue cycle at a notch may bear little relation to the mean load in the applied load cycle.

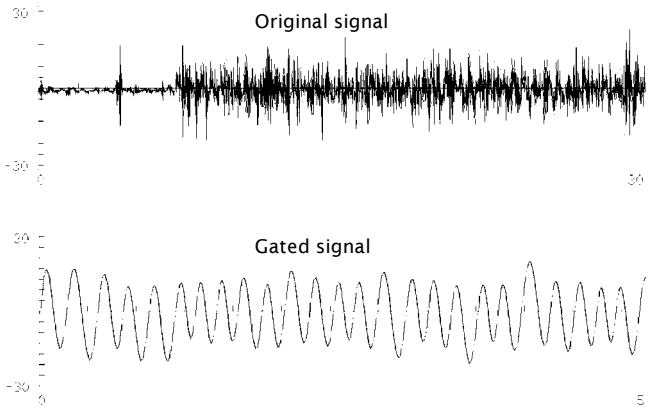


Figure 13.18. Excessive gating produces an almost constant amplitude signal

13.5 Effect of larger cycles

Very large cycles in a signal are usually rare, but their effect on the residual stresses at notches can be significant. A large tensile strain can leave beneficial compressive residual stresses at a notch, which can extend the fatigue life, and the converse is true for large compressive strains *(Figure 13.19)*. Large loads in a test signal should be applied with caution, and it has been suggested that they should be omitted from the test command signal until the design life has been achieved on the test, and then applied.

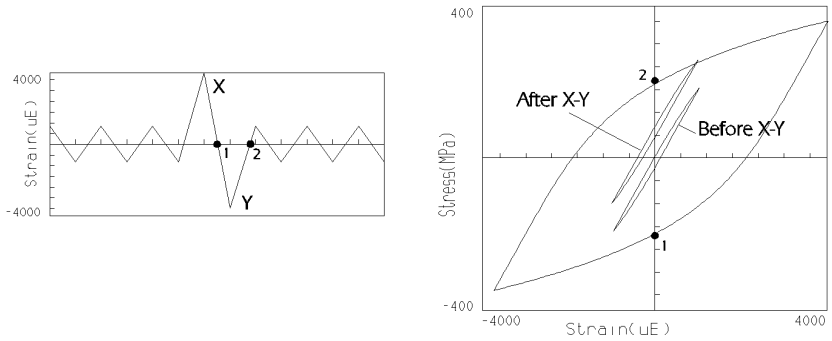


Figure 13.19. Effect of a large cycle on mean stress at a notch

13.6 Generic test signals

Real measured signals may not be available during the early stages of a design. However, it is clear from the preceding sections that fatigue tests to develop basic design data, and to compare candidate materials and manufacturing processes, must be carried out using test signals which represent typical service loading. For this reason, generic test signals have been developed for different industries. The first of these, TWIST, was completed in 1973 for transport aircraft wing loading, and other signals have been developed for fighter aircraft, helicopters, aircraft turbine engines, offshore structures, steel mills and wind turbines. Development of CARLOS, the car loading sequence, started in 1985, and has produced sets of signals for car wheel suspension forces and for power train loading. (Table 13.2)

The development of these signals revealed much about the required properties of fatigue test signals. They must contain the correct mixture of large and small cycles, and as small cycles below the constant amplitude endurance limit contribute fatigue damage, they must be retained in the signal. The signal must not repeat too often before failure of the test specimen, otherwise the effect will be that of a constant amplitude signal. Long signals must be used to achieve these requirements, and the more recent generic test signals contain hundreds of thousands of cycles.

TEST SIGNAL	PURPOSE	COMPLETED
TWIST	Transport wing standard	1973
GAUSS	Gaussian sequences	1975
FALSTAFF	Fighter aircraft	1976
HELIX, FELIX	Helicopter loading	1984
COLD TURBISTAN	Cold compressor disk	1987
HOT TURBISTAN	Hot compressor and turbine disk loading	Started in 1986
WASH	Offshore structures	Started in 1984
WAWESTA	Steel-mill drive	Started in 1986
WISPER	Wind turbine	Started in 1985
ENSTAFF	Environmental FALSTAFF	Started in 1983
CARLOS	Car components	Started in 1987
MINI-TWIST	Shortened TWIST	1979

SHORT FALSTAFF	Shortened FALSTAFF	1980
MINI-HELIX	Shortened HELIX and FELIX	1985
MINI-FELIX		

Table 13.20 Generic test signals

One of the earliest attempts to develop standard test sequences was made by the SAE during the 1970's (Ref. 13.5). Three signals from automotive components were produced, and these are the familiar 'transmission', 'suspension' and 'bracket' signals. They were used to assess the local strain fatigue algorithms which were then being developed. This was a very far-sighted project, but it was constrained by the available data-recording and computer technology. In particular, the signals were very short (only a few thousand data points), and in part this was because they were transferred between different makes of computer by typing in the data points on a keyboard. To make the signals short, small cycles were omitted, and so the signals repeated many times during tests. Modern analysis shows that, for the widely used 'transmission' history, much of the fatigue damage is produced by a narrow range of cycle amplitudes *(Figure 13.20)* so the signal is almost a constant amplitude signal. Subsequent testing has shown that reinstatement of the small cycles shortens the test life significantly.

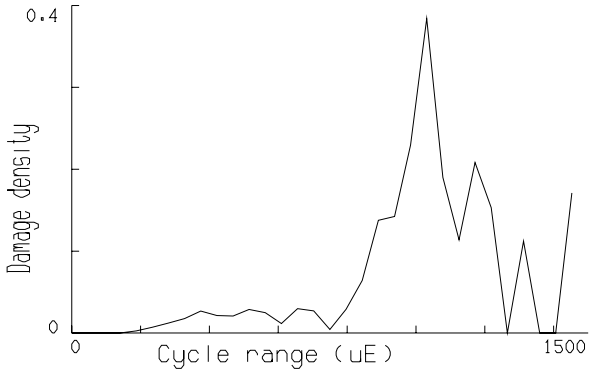


Figure 13.21 Damage distribution for the SAE transmission history

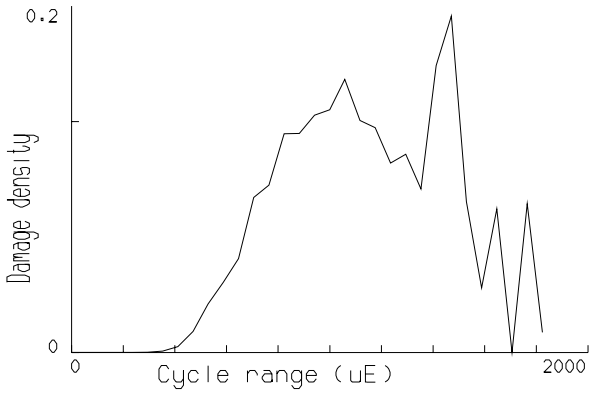


Figure 13.22 Damage distribution for the 'Carlos Vertical' suspension history

The CARLOS uniaxial sequence contains some quarter-million peak/valleys, and cycles as small as 15% of the maximum value are retained. The fatigue damage is produced over a wide range of cycle amplitudes *(Figure 13.22)*.

Generic test signals are designed not only to avoid the pitfalls of constant amplitude or block loading signals, but also to provide test data under conditions which are representative of a particular industry.

They are intended for comparison of different materials, design details, production techniques and methods of improving fatigue strength. They may also be used to compare different fatigue life prediction techniques, to investigate scatter in fatigue lives, and to provide an indication of allowable stresses. They have the added advantage that fatigue test results from different organisations are more directly comparable.

13.7 Development of in-company test signals

Appreciating that test signals must closely reproduce customer usage, many companies are developing their own test command signals. The usual procedure is to measure various aspects of the service duty, which for a road vehicle may involve service measurement on motorways, city roads, good and bad B-roads, etc. Rainflow cycle distributions of the measured forces may be combined in different proportions to reproduce the cycle distributions measured from real customer usage, perhaps using damage distributions or damage-oriented weighting functions to provide additional matching criteria.

Once the desired mixture of road surfaces has been obtained, a test signal can be generated, either from the original measured signals, or from the Rainflow cycle distributions using randomisation techniques. If the signals are generated from Rainflow distributions, it is important that they are generated from the original constituent parts (motorways, city roads, etc.) and not from the final combined cycle distribution. Although computer-based analysis may show the same fatigue damage distributions for both cases, the signal regenerated from the combined cycle distribution will have a different sequence (the potholes may appear in the middle of the motorway) and this could produce quite different test results in terms of fatigue lives, design comparisons and identification of critical locations.

13.8 Summary

Constant amplitude tests may appear to be a low-cost test method, but they give greater scatter than variable amplitude tests, may not identify possible failure locations, and may rank materials and manufacturing processes in the wrong order of merit.

Variable amplitude tests can be cheaper because they give reduced scatter, and do identify failure locations and give the correct ranking of materials and processes, provided that the test signals are representative of service usage.

Block loading patterns derived from service signals can give optimistic fatigue lives.

Small cycles, below the constant amplitude endurance limit, can contribute to fatigue damage when they are mixed with larger cycles. Signal gating software must be used with caution. In the absence of more specific information, cycles with amplitudes smaller than 25% of the constant amplitude endurance limit can probably be removed. However, gating which results in an essentially constant amplitude signal could give the potential errors inherent in constant amplitude testing.

Narrow band signals give Miner's summations close to 1.0. As signals become broader band, the apparent Miner's summation becomes much less than 1.0, because there is a greater mix of small and larger cycles.

Large cycles can leave beneficial residual compressive stresses at notches. Cycles with amplitudes greater than 150% of the constant amplitude endurance limit should be applied with caution.

Scaling up the amplitude of test signals to accelerate the test can produce non-representative failures, incorrect residual stresses at notches, and incorrect ranking of materials and processes.

13.9 References

13.1 British Standard BS7608

Fatigue design and assessment of steel structures, British Standards Institution, 1993

13.2 Marquis G.B

High Cycle Spectrum Fatigue of Box Beam Components

'Product Optimisation for Integrity '.

Proceedings of the Third International Conference of the Engineering Integrity Society,

Editors E R Cawte, J M Draper and N Trigwell. Sheffield Hallam University, Sheffield UK, April 1995

13.3 Conle A and Topper T.H

Overstrain Effects During Variable Amplitude Service History Testing

International Journal of Fatigue, Vol 2, No.3, pp130-136, 1980

13.4 DuQuesnay D.L, Pompetzki M.A, Topper T.H
Fatigue Life Prediction for Variable Amplitude Strain Histories
SAE Paper 930400, Society of Automotive Engineers

13.5 SAE test program described in SAE publication AE-6

14 Practical fatigue analysis

14.1 Introduction

Measured service histories, usually from strain gauges, are the basis of modern fatigue analysis for component assessment, prototype development and fatigue failure investigation. This chapter considers the location of strain gauges, sampling signals, strategies for data acquisition and analysis, and the interpretation of fatigue analysis results.

14.2 Locating strain gauges

Strain gauges may be located at the root of a notch to measure local strains, or at some position away from the notch to measure nominal strains unaffected by the local stress/strain distribution in the region of the notch.

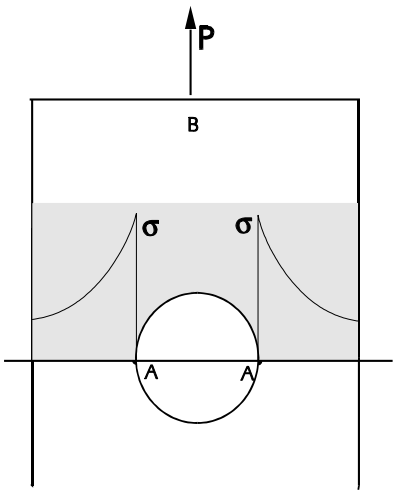


Figure 14.1 Location of strain gauges

Stress concentrations produce very complex distortions to the stress/strain field. It is important therefore not to place strain gauges in locations where the measured strains would be neither true local strains at the notch, nor truly nominal strains unaffected by the notch. For a plate with a hole, for example, strain gauges should be located at A or B in *Figure 14.1*. At location A, the measured strains will be very close to local values. B is sufficiently remote from the effects of the hole to provide true nominal strains and allow the application of stress concentration factors obtained from published sources. The locations shown shaded will provide strains which are neither nominal nor local, and the derivation of a stress concentration factor to convert the measured strains into local strains would be very difficult.

When measuring nominal strains some distance from a notch, attention should be paid the possibility of nominal strain gradients - for example changes in bending moment - which may require geometric corrections to the measured strains before applying a stress concentration factor.

For welded joints, design codes such as BS7608 require nominal stresses. The effect of the stress concentration caused by the weld is allowed for in the published S-N curves. Care must be taken to ensure that strain gauges are not too close to the weld toe if excessively conservative results are to be avoided. The Welding Institute (TWI) has published papers showing how far stress concentration effects extend from a weld toe, and this distance can be of a similar magnitude to the thickness of the welded plates *(Figure 14.2)*.

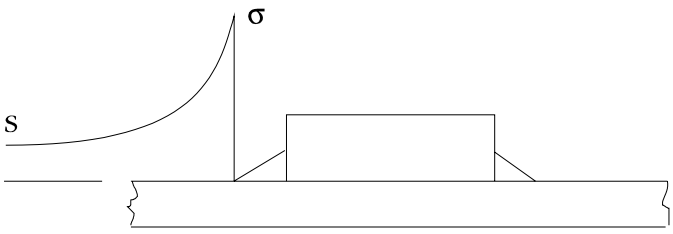


Figure 14.2 Stress concentration near a weld toe

For local strain fatigue analysis, the choice between nominal and local strain gauges may be difficult. If local strains are measured, the possible uncertainty in the calculation of the stress concentration is avoided, as is the additional uncertainty inherent in Neuber's Rule. However, the measured strains will include plasticity, and so the strain histories will be specific to the material *on which they were measured*. This is of great importance if measured strains are to be used to assess alternative materials. Using Neuber's rule for illustration, *Figure 14.3* shows the local strains which would be calculated for two materials from the same measured nominal strain, e.

The nominal strain, e, is on the elastic part of the stress-strain curve, and so the same strain would have been measured on both materials. For the purposes of this example, it will be assumed that the nominal strain was measured on a component made from SAE1005. Using Neuber's rule, the point C is calculated from the product K_t.e. A Neuber hyperbola through point C intersects the stress-strain curve for SAE1005 at point A, so this is the local stress-strain that would have been measured in this material. The same hyperbola intersects the stress-strain curve for MANTEN steel at point B, so this represents the local stress-strain that would have been measured on a component made from this material, for the same nominal strain e.

However, if local strains measured on the SAE1005 material are applied to MANTEN as local strains, this gives a stress σ_x and a strain ε_x - much higher values than would in fact have been measured in this material. The error is clear. In (Ref. 14.1) it was shown that using local strains from one material to assess another can result in the stronger material being rejected, or a weaker material being selected. An analysis using nominal strains and Neuber's rule avoids this error.

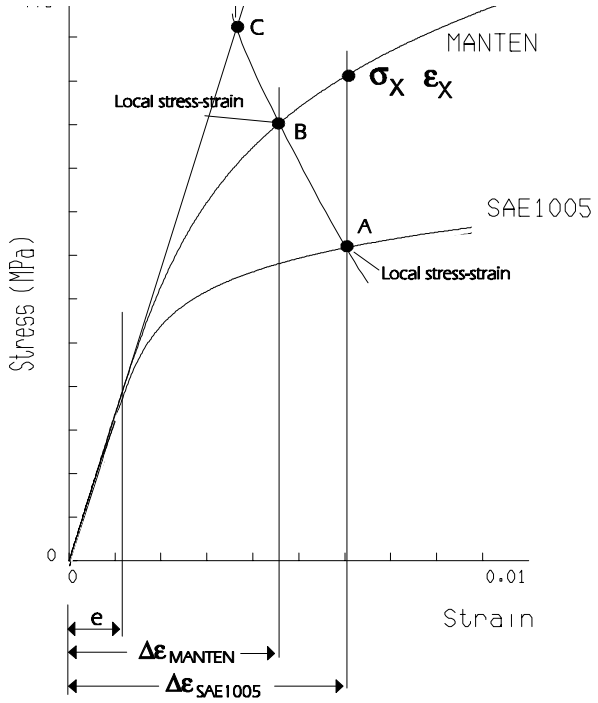


Figure 14.3 Error in using measured local strain for comparison of materials

If local strains have been measured, it is possible to covert them into the equivalent elastic strains at point C strains using Neuber's rule in reverse, and then use a Neuber analysis to calculate the local strains in another material. The effect of material memory must be included.

14.3 Data recording

The 'data chain' for engineering investigation is shown in *Figure 14.4*

Analogue tape has been used to record measured signals since MIRA first interfaced a tape recorder to an instrumentation system in the 1950's. The data is stored as frequency-modulated (FM) data, in that the variable amplitude signal is used to modulate a carrier frequency so that the data may be stored in constant- amplitude variable-frequency form. For computer analysis the data must at some stage be replayed and digitised. Although FM tape has some advantages, it is probably true to say that it will be superseded for many engineering applications by digital data storage. FM tape systems are limited in the number of channels that can be recorded. The channel calibration information must be recorded and used to replay the data. The signal-to-noise ratio is limited compared with digital systems, which means that the user must try to ensure that each channel is scaled to use as much of the available range as possible, and this requires prior knowledge of the amplitudes which will be measured. The subsequent digitisation process adds another step to the data chain before the signals can be analysed. However, FM tape has a wide frequency bandwidth, and the data can be visualised in analogue form on an oscilloscope. Perhaps most importantly, the user is not committed in advance to a specific sample rate with consequent limiting frequency.

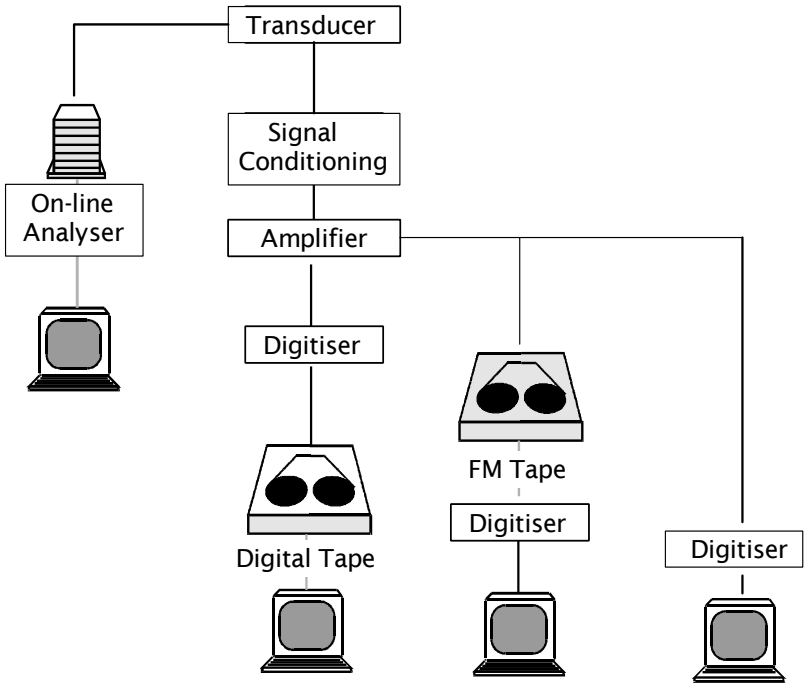


Figure 14.4 The data chain

Digital data storage is now available in many forms. Digital tape systems can acquire very large numbers of channels sampled at high frequencies. Typical applications may be from 48 to 256 channels sampled at several thousand samples per second per channel.

Less demanding applications can be met by computer-based systems using PC-compatible computers with commercial analogue-to-digital converters. Many mechanical engineering applications of up to 200 channels with up to 500,000 samples/second can be satisfied using PC-based systems with real-time data storage to hard disk combined with real-time graphics display of measured data from selected channels. Storage on hard disk means that the data is immediately available for display or analysis. As an example of hard disk capacity, a 1 GByte disk could store 17 hours of data from a sixteen channel system sampling each channel at 500 samples/second. Real time analysis, for example Rainflow cycle counting and PSD's, can also be included on PC - based systems.

14.4 Short term recording or long term analysis

Data acquisition on to data tape, for later replay and computer analysis, has been a standard method for engineering investigation. However, advances in electronics in the 1970's made other alternative strategies possible. The Datamyte of the 1970's was a small portable single-channel Rainflow cycle counter which digitised the analogue signal and carried out an on-line Rainflow cycle count into a cycle matrix stored in battery-backed memory. The Datamyte was a major change in thinking and, although limited by the computing power and memory then available, allowed engineers to monitor the service environment of a component over long periods of time with a single low-cost device. Ease of use allowed it to be used by mechanical engineers without extensive knowledge of instrumentation techniques.

A perceived limitation of the Datamyte was its inability to store significant lengths of recorded signals. This led to the wider concern about data quality in general, as once an event in the signal had been cycle counted and added to the cycle matrix, it was difficult to determine if it was a real event, or some form of data corruption.

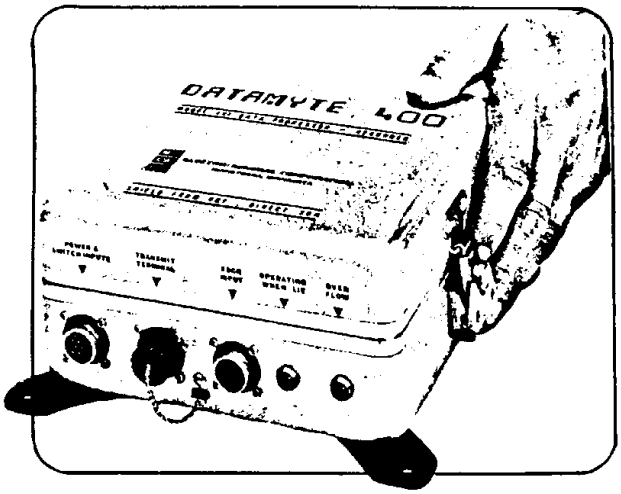


Figure 14.5 Datamyte 400 recorder/processor

In 1983 the Ground Vehicle Instrumentation Group, now the Measurement Systems Group of the Engineering Integrity Society, published a specification for an untended multi-channel data storage and long term on-line analysis device. Instrumentation companies produced systems which offered various combinations of data storage and analysis capability.

The introduction of high speed notebook PC's, combined with rugged high-capacity hard disks, offers an alternative approach, with advantages in standard operating systems and components, high channel counts, ease of programming and readily available post-processing software.

These many different devices are attempts to resolve the conflicts which face the engineering investigator. Recording measured signals as time histories - the raw data - is obviously desirable. The engineer can assess data quality by visualising the signals, and can run a wide range of analysis software as an investigation proceeds. Conclusions on the behaviour of a component can be drawn simply by being able to see the signals. However, this process is time consuming for even short lengths of data, and is quite impractical if long term monitoring is to be used to derive duty cycles or investigate rare events. This is because of the engineering time required and because of the vast quantity of data which must be stored.

Modern on-line analysers avoid these disadvantages by allowing signals to be acquired and processed continuously over very long periods of time - months or years. Several different types of analysis may be run on each channel simultaneously, and the recording of raw data can be event-driven so that only essential data is stored. These systems also provide instant access to analysis results. However, doubts about data quality do exist, and the methods of on-line analysis must be selected in advance.

With the systems now available, a combined approach is often adopted. In the early stages of an engineering investigation, a large number of channels may be recorded as raw data over short time periods, perhaps recording specific types of events which the component is known to experience. Off-line analysis of these signals may be combined with supporting modal or finite element analysis to provide an understanding of the component behaviour. This approach may then be supplemented by long term on-line analysis of a smaller number of channels to confirm the results over longer time-

scales. One aim of instrumentation designers is to develop devices which allow both the short term recording and long term analysis functions to be carried out by the same data acquisition system.

14.5 Sampling signals

At some stage in the data analysis process, analogue signals from strain gauges or other transducers must be digitised for computer analysis. The selection of a sampling frequency can have a major influence on the accuracy of the subsequent fatigue life calculations. Standard references on signal processing suggest a sampling rate of at least twice the maximum frequency present in the signal. This is sufficient to allow the frequency content of the signal to be determined, but fatigue analysis requires an accurate definition of the amplitudes, and significantly higher sampling frequencies are necessary.

As an example, a sine wave sampled at four times its frequency could produce many different sets of sampled values. Two possible sets of samples are shown in *Figure 14.6*.

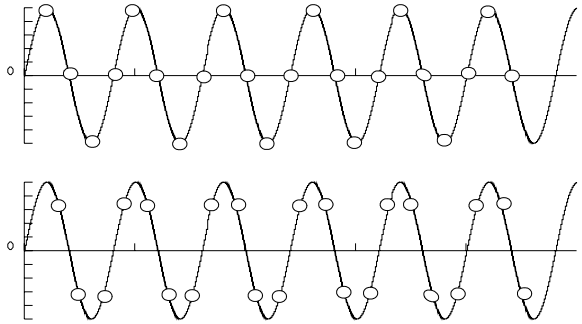


Figure 14.6 Possible samples from a signal sampled at four times the signal frequency

In the upper diagram the samples have occurred (by chance) on the peaks and valleys in the signal. In the lower diagram the peaks and valleys have not been accurately defined. The sample rate of four times the frequency of the signal is not adequate to define the amplitudes of the peaks and valleys.

In (Ref. 14.2), a narrow band Gaussian signal was used to investigate the error in fatigue life prediction which results from under-sampling a signal. Taking a sample rate of 100 times the signal frequency as giving the 'correct' answer, and normalising all other calculated lives by this value gave the results shown in *Figure 14.7* for local stress-strain analysis.

In a separate analysis for this, a broad band signal was created by low-pass filtering Gaussian white noise. The filter frequency was chosen to give 100 samples per cycle at the cut-off frequency. Again, fatigue lives were calculated with various reductions in sample frequency. Both sets of results are shown in *Figure 14.7*.

It can be seen that sampling at 10 times the signal frequency gave calculated lives of 1.1 times the true value for a broad band signal, and up to 1.5 times the true value for a narrow band signal. A sample frequency of 10 points/cycle is now widely used in industry, as it offers a reasonable compromise between quantity of data and accuracy of analysis.

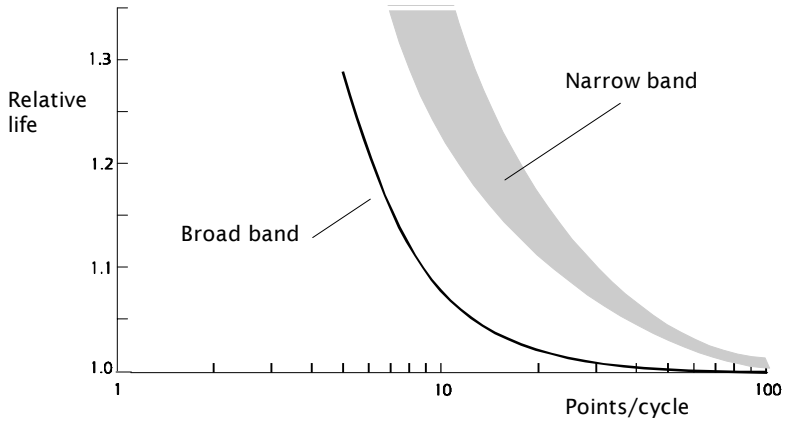


Figure 14.7 Effect of sampling frequency on fatigue life estimation

14.6 Interpreting analysis results

The end result from a complex and expensive instrumentation, data acquisition and computer analysis project can be a single number - the fatigue life. This provides no information on why the fatigue life is a particular value, nor, if the life is too short or too long, does it help in the search for a more optimum design.

A display which indicates the approximate distribution of fatigue damage in the signal can help in finding the most damaging sections of a service history. A possible solution to an unacceptably short life may be to eliminate or reduce the frequency of occurrence of these events.

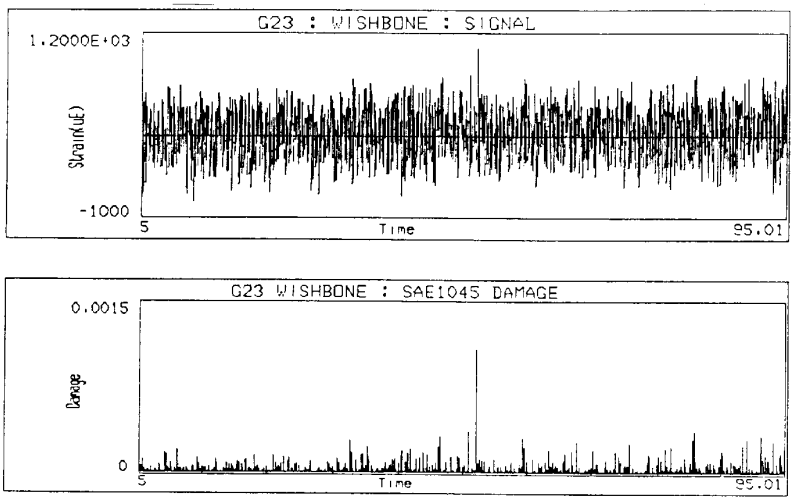


Figure 14.8 Display showing damaging sections of a signal

Cycle histograms provide a description of the cycles present in the signal, and damage histograms *(Figure 14.9)* show which cycles contribute most to the total damage. More importantly, they show whether the damage is dominated by a small number of large cycles, or a larger number of small cycles. This is valuable in the selection of alternative materials.

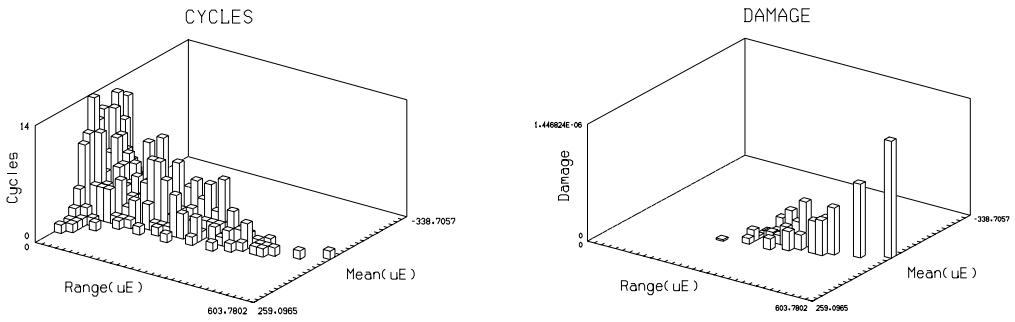


Figure 14.9 Cycle and damage histograms from a local strain analysis

In *Figure 14.9*, the histograms were obtained from a local strain analysis, and show that the damage is dominated by two large cycles. In itself this should be cause for concern about the validity of the analysis, because it suggests that the length of signal may be too short to define the magnitude of these cycles, or their true probability of occurrence. *Figure 14.10* shows the ten largest hysteresis loops from the same analysis, and emphasises the single large cycle.

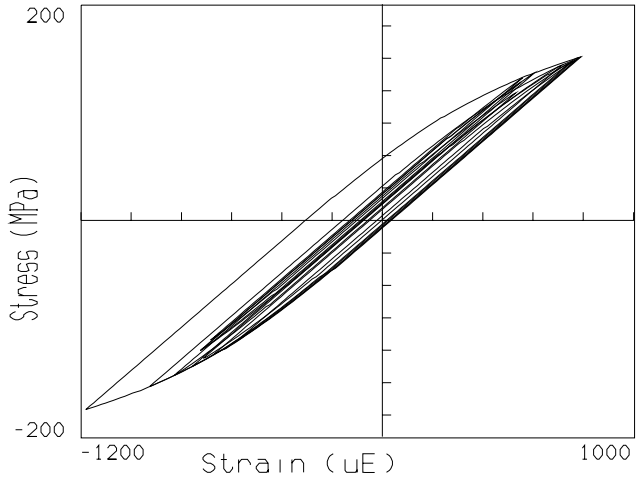


Figure 14.10 Hysteresis loops for the analysis of Figure 14.12

Figure 14.11 shows the same signal analysed using a BS7608 welded joint S-N curve. In this case the largest damage contribution comes from the relatively large number of medium-sized cycles, and this is quite typical of S-N curves for welded joints.

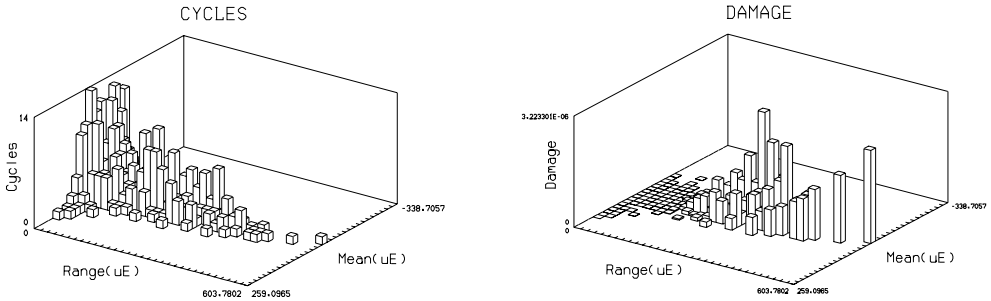


Figure 14.11 Cycle and damage histograms from a BS7608 analysis

Figure 14.12 shows a cycle histogram from a much longer length of signal, with a more continuous distribution of cycle ranges.

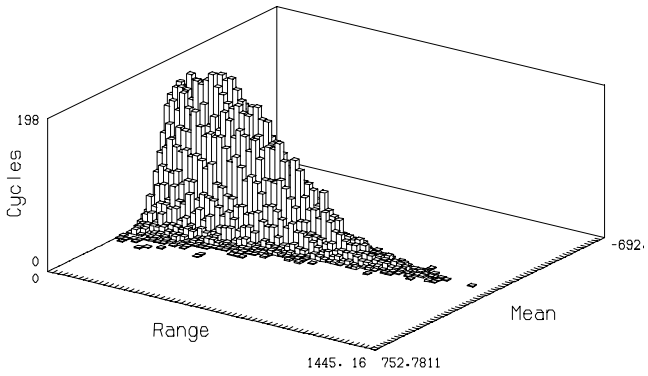


Figure 14.12 Cycle histogram from a long signal.

Cycle and damage histograms can be plotted on the basis of cycle range, in which case it is useful to convert the histogram into a density diagram by dividing by the width of the bins used to form the histogram. This produces a diagram which is independent of the bin width used in the analysis. The diagram may then be used for comparison with density diagrams from other signals, may be added to other density diagrams, and re-scaled to represent different lengths of data, with great flexibility. For a cycle density diagram, the area between any two cycle ranges represents the number of cycles between these two ranges. For a damage density diagram, the area between any two ranges is the total damage caused by cycles between these cycle ranges.

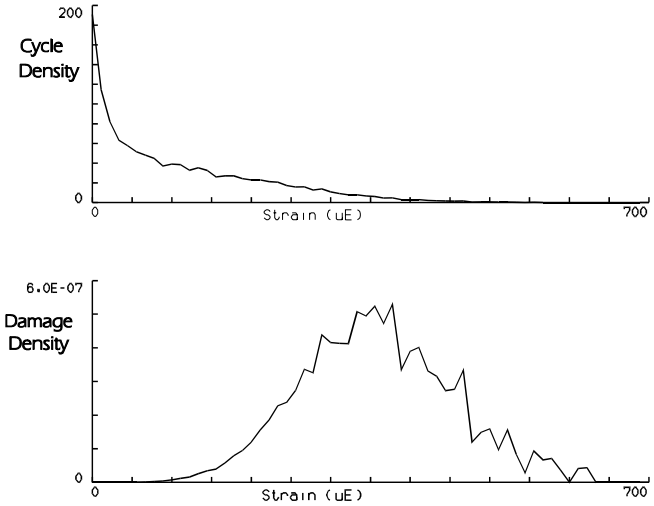


Figure 14.13 Cycle and damage density diagrams

The effect of design changes can be assessed using graphs which show the effect of stress concentration on fatigue life. Similar analysis can show the effect of an overall reduction in stress-strain which may result from a change in component dimensions or an overall reduction in the magnitude of the applied loads. Note that linear scaling of measured strain histories is not valid if plasticity is present in the original signals, as will be the case for measured local strains on machined components.

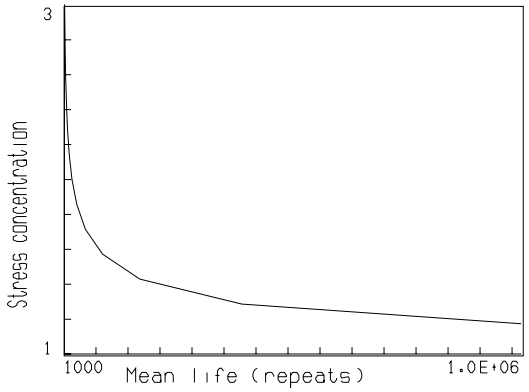


Figure 14.14 Effect of stress concentration on fatigue life

14.7 References

14.1 Draper J

Interpretation of Measured Strains for Materials Selection Using Critical Location Fatigue Analysis

EIS International Conference 'Measurement for Integrity, Buxton, 1988

Engineering Integrity Society

14.2 Morton K, Musiol C, Draper J
Local Stress-Strain Analysis as a Practical Engineering Tool
Proc. SEECO 83 'Digital Techniques in Fatigue'. City University, London 1983
Society of Environmental Engineers

Index