*ASM Handbook*, Volume 11, *Failure Analysis and Prevention*
B. Miller, R. Shipley, R. Parrington, D. Dennies, editors
DOI 10.31399/asm.hb.v11.a0006779



# Fatigue Failures

Revised by Lothar Issler, University of Applied Sciences Esslingen (Germany)

FATIGUE FAILURES may occur in components subjected to fluctuating (time-dependent) loading as a result of progressive localized permanent damage described by the stages of crack initiation, cyclic crack propagation, and subsequent final fracture after a given number of load fluctuations.

Fatigue has an overwhelming importance in the assessment of the safety and integrity of structures and consequently in failure analysis as well. Many service failures of cyclic-loaded components still occur by fatigue, despite a well-defined characterization of the fatigue properties in the laboratory and a proper fatigue design. In part, the difficulty is that fatigue behavior is a complex subject influenced by many variables, such as component size and shape, material fatigue properties, the presence of mean stresses including residual stresses, surface conditions, local stress concentrations due to notches and stiffness changes, as well as material and manufacturing imperfections or discontinuities. The problem of fatigue is the sensitivity to even small geometric and metallurgical imperfections and cracks, which may accumulate into a critical situation.

Another specific and crucial problem in fatigue design is the often insufficient information about the exact loading situation of a structure, for example, the type, sequence, magnitude, and number of the fluctuating (multiaxial) stresses often in combination with complex environmental influences (temperature, corrosion, wear). These factors complicate the laboratory simulation of actual application conditions and often require full-sized component tests under variable loading and environmental influences, where, due to time and cost constraints, the statistical analysis of the structural behavior may be problematic.

Thus, due to its complex and varied conditions, prevention of fatigue failure requires a profound theoretical and practical background as well as experience and awareness in all stages of the life of a component. This includes the critical assessment of design methods, the proper selection of materials and manufacturing methods, and the knowledge of anticipated service conditions, including manufacturing and operation inspections as well as maintenance and repair. An important link in the improvement chain and the prevention of future failures is a consequent, profound, and competent analysis of fatigue failures.

## Fatigue Properties and Design Life

The three design methods or philosophies in fatigue are given in Table 1. The methods should ideally include the assessment of the whole component life, with the conditions for crack initiation, cyclic crack growth, and final fracture. It must be distinguished between an infinite-life design, where the number of cycles is not important, and a finite-life design, in which the cycles up to the end of life must be defined and limited. The conventional design methods are supplemented by the fracture mechanics methods, in which cracklike flaws in the structure are anticipated and included in the design. These methods have all contributed to a better understanding of fatigue crack nucleation in regions of high stresses and strains, the subsequent crack growth mechanisms, and final fracture. They have also contributed to the optimization of materials, manufacturing, and component design as well as to the conduction of a proper failure analysis.

## Infinite-Life Criterion (*S-N* Curves)

The safe-life/infinite-life philosophy is the oldest of the approaches to fatigue. The design method is based on the stress-level/lifetime relationship, which is expressed with the *S-N* curve (stress amplitude versus number of cycles to failure). The necessary additional information about the influence of the mean stress on the tolerable stress amplitude is expressed by the fatigue diagram.

**Table 1 Methods and material characteristics for fatigue design concepts**

| Design philosophy | Design concept | Material characteristics |
|---|---|---|
| Safe-life, infinite-life | Nominal stress or local stress concept | *S-N* curve, $S_a = f(N_F)$<br>Fatigue diagram, $S_a = f(S_m)$ |
| Safe-life, finite-life | Local strain concept | Low-cycle fatigue curve, $\varepsilon_a = f(N_i)$<br>Cyclic stress-strain ($\sigma_a$-$\varepsilon_a$) curve |
| Damage tolerance | Fracture mechanics concept | Threshold value, $\Delta K_{th}$<br>Crack growth diagram, $da/dN = f(\Delta K,R)$<br>Fracture toughness, $K_{Ic}$ |

The stress ($S$)/time ($t$) cycle is schematically shown in Fig. 1. The cycle is described by several parameters. The main characteristics of a stress cycle are the stress amplitude, $S_a$, and the mean stress, $S_m$. The amplitude, $S_a$, is determined from the maximum and minimum stress of the cycle:

$$S_a = (S_{max} - S_{min})/2 \quad \text{(Eq 1)}$$

The range of stress, $\Delta S$, is the algebraic difference between the maximum and minimum stress:

$$\Delta S = 2 \cdot S_a = S_{max} - S_{min} \quad \text{(Eq 2)}$$

The mean stress is the mean value of the maximum and minimum stress:

$$S_m = (S_{max} + S_{min})/2 \quad \text{(Eq 3)}$$

In a completely reversed cycle ($S_{min} = -S_{max}$), the mean stress is 0; in a pure tension pulsating load, the mean stress is $S_m = S_a$; and in a pure compressive load, the mean stress is $S_m = -S_a$.

The stress ratio is the algebraic ratio of two specified stress values in a stress cycle. Two commonly used stress ratios are the ratio, $A$, of the alternating stress amplitude to the mean stress ($A = S_a/S_m$) and (preferably) the ratio, $R$,

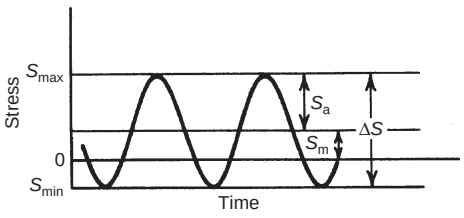


**Fig. 1** Characteristics of a stress ($S$)/time ($t$) cycle

of the minimum stress to the maximum stress ($R = S_{min}/S_{max}$).

The correlation of mean stress and stress amplitude in terms of $R$ is given by:

$$S_m = S_a \cdot (R + 1)/(R - 1) \quad \text{(Eq 4)}$$

Examples of typical structures loaded with different $R$-ratios are shown in Fig. 2.

If the stresses are fully reversed ($S_{min} = -S_{max}$), the stress ratio $R$ becomes $-1$. If the stresses are partially reversed, $R$ becomes a negative number less than 1. If the stress cycle is purely pulsating between a maximum stress and complete unloading ($S_{min} = 0$), the stress ratio $R$ becomes 0. If the stress is cycled between two positive tensile stresses, the stress ratio $R$ becomes a positive number between 0 and 1. A stress ratio $R = 1$ indicates no variation in stress, which corresponds with static loading.

The $S$-$N$ curve is the basis for fatigue design and shows the relationship between a constant stress amplitude, $S_a$ (or stress range, $\Delta S$), and the number of cycles to failure, $N$. The regions of the curve can be separated into the quasi-static, finite-life, and infinite-life regimes (Fig. 3).

Failure in $S$-$N$ testing is typically defined by fracture, which means by the total separation of the structure. The fatigue limit and the high-cycle regime of the $S$-$N$ curve are usually within the linear-elastic range and can therefore be based on loads, nominal stresses, structural stresses, or local stresses. For elastic-plastic behavior, local strains should be the basis of the assessment.

The finite-life regime in a double-log plot (log $S_a$-log $N$) can be approximated by a straight line. Using any reference point on the finite-life line, $P_R$ ($N_R/S_{aR}$), the equation of the finite-life line is:

$$N = N_R \cdot (S_{aR}/S_a)^k \quad \text{(Eq 5)}$$

where $k$ is the slope of the finite-life line. The slope $k$ is typically characterized by values between 3 and 15. A low slope corresponds with a low relationship between the fatigue strength, $S_{AF}$, and the static strength, $S_u$ (e.g., $S_{AF}/S_u = 0.05$), which corresponds with an unfavorable fatigue behavior.

Because of the significant scattering of the results of fatigue tests, the $S$-$N$ curve must be evaluated statistically (Fig. 3). The mean number of cycles, $N_{50}$, for a failure probability of 50% and a total of $m$ tests on the stress level $S_{ai}$ with single results $N_i$ is given by the equation:

$$\lg N_{50} = 1/m\,\Sigma \lg N_i \quad \text{(Eq 6)}$$

The scattering is defined by the standard deviation:

$$s = \sqrt{1/(m - 1\Sigma)(\lg N_i - \lg N_{50})^2} \quad \text{(Eq 7)}$$

The number of cycles for a given failure probability, $P_F$, is:

$$\lg N_{P_F} = \lg N_{50} + s \cdot u \quad \text{(Eq 8)}$$

The quantile (safety span), $u$, depends on the type of distribution (e.g., Gauss, Weibull, logit, arc sin $\sqrt{p}$) and the failure probability. Table 2 shows $u$-values for different failure probabilities for the most often applied Gauss log distribution.

The scatter of fatigue tests in terms of number of cycles is usually defined by the scattering range, $T_N$ , which is calculated from the number of cycles for the failure probabilities $P_F = 90\%$ and 10%:

$$T_N = N_{90}/N_{10} \quad \text{(Eq 9)}$$

Typical scattering ranges in fatigue tests that indicate low scattering are $T_N = 2$ to 5; medium scattering, $T_N = 5$ to 10; and high scattering, $T_N > 10$. The corresponding scattering range in terms of stresses, $T_S$, can be expressed with the slope $k$ as $T_S = T_N^{1/k}$.

## Synthetic S-N Curve

The fatigue strength of a structure is dependent on many variables. The most important influences are material type, material strength and toughness, component size, surface finish, surface treatment, notch effect, mean stress, and environmental influences. A typical equation to determine the fatigue strength (nominal stress amplitude for infinite life) from the static tensile strength, $S_u$, is:

$$S_{aF} = C_{material} \cdot S_{size} \cdot C_{finish} \cdot C_{surface} \cdot C_{notch} \cdot C_{mean} \cdot C_{environment} \cdot S_u \quad \text{(Eq 10)}$$

The equations and correlations for the influencing factors are described in literature, standards, and codes. An example of the correction factor $C_{finish}$ for surface roughness is plotted in Fig. 4 (Ref 1).

Equation 10 is useful in failure analysis because on one side it provides hints for the possible reasons for fatigue failures, and on the other side, useful information for the necessary prevention measures to avoid future failures. Consequently, an optimized fatigue design is mainly characterized by a suitable fatigue-proof material with an adequate combination of strength and toughness, optimized surface finish, favorable surface treatment, mild notches and soft stiffness changes, compressive (residual) stresses, and the absence of a detrimental environment.

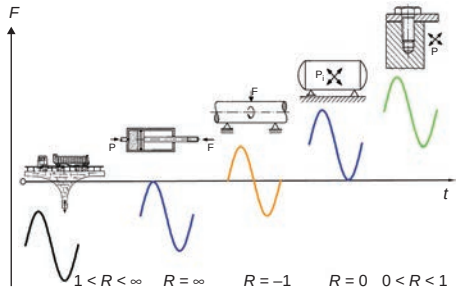


**Fig. 2** Examples of structures with typical $R$-ratios

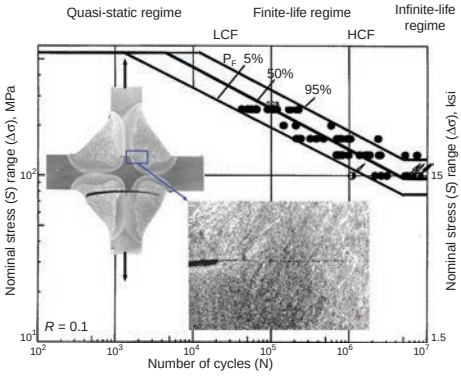


**Fig. 3** $S$-$N$ curve for cruciform metal-active-gas-welded joints (structural steel S355, ASTM A572 grade 5). LCF, low-cycle fatigue; HCF, high-cycle fatigue; $P_F$, probability of failure

**Table 2 Safety span for different failure probabilities (Gauss log distribution)**

| Failure probability, % | Safety span |
|---|---|
| 0.0001 | −4.75 |
| 0.001 | −4.26 |
| 0.01 | −3.72 |
| 0.1 | −3.09 |
| 1 | −2.33 |
| 5 | −1.64 |
| 10 | −1.28 |
| 99.9999 | 4.75 |
| 99.999 | 4.26 |
| 99.99 | 3.72 |
| 99.9 | 3.09 |
| 99 | 2.33 |
| 95 | 1.64 |
| 90 | 1.28 |

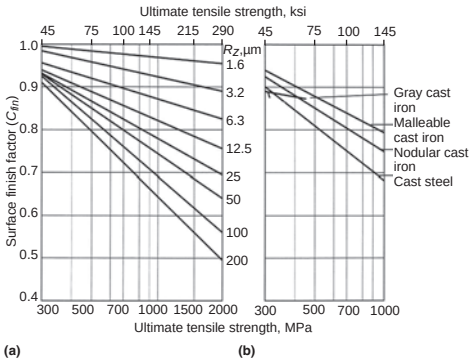


**Fig. 4** Example of correction factor for surface finish for steels. Source: Ref 1

Care should be taken in applying Eq 10, because an increase of the static strength, $S_u$, may not be the proper reaction and remedy for a fatigue failure. The selection of an alternative material with a higher strength (and possibly lower toughness) will not necessarily result in an improvement of the fatigue strength, because of the observation that an increased strength of the structure, $S_u$, will be connected with a decrease in many of the correction factors, such as a lower material factor, increased sensitivity to the surface quality (Fig. 4), lowering of the supporting factor for notch stresses (Eq 39), and higher sensitivity to the mean and residual stresses and environmental influences (corrosion).

## *Mean Stress Effect*

The tolerable fatigue stress amplitude is significantly influenced by the acting mean stress. With an increasing positive mean stress, the permissible amplitude steadily decreases, whereas an increasing negative mean stress increases the permissible stress amplitude.

To assess the mean stress influence, a series of fatigue tests must be conducted at various mean stresses ($R$-ratios) and the results plotted as a series of $S$-$N$ curves. Different presentations of the mean stress effect in fatigue diagrams have been proposed by Smith ($S_{max}$ and $S_{min} = f(S_m)$), Komerell ($S_{max} = f(S_{min})$), and Moore ($S_{max} = f(R)$).

The results of fatigue tests with different mean stresses are plotted preferably as a Haigh diagram (Ref 2), which shows the stress amplitude ($S_a$) versus the mean stress ($S_m$) (Fig. 5).

For general consideration of mean stress effects, various models of the mean stress-amplitude relationship in the Haigh presentation have been proposed. Figure 6 shows different approximations of the fatigue diagram. The shape of the diagram depends mainly on the material, the geometry of the component, and the type of loading (axial, bending, torsion, and shear). The lines are determined by the ultimate tensile strength, $S_u$, and the alternating fatigue strength amplitude, $S_{alt}$ ($R = -1$). The connection of $S_u$ and $S_{alt}$ with a straight line is known as a Goodman line (Ref 3):

$$S_a = S_{alt}[1 - (S_m/S_u)] \quad \text{(Eq 11)}$$

A parabolic connection leads to the Gerber parabola (Ref 4):

$$S_a = S_{alt}\sqrt{1 - (S_m/S_u)} \quad \text{(Eq 12)}$$

The behavior of brittle materials, such as gray cast iron, is mainly modeled with the Goodman line, whereas ductile materials are preferably described with the Gerber parabola. The diagrams for shear stresses (torsion, shear force) must be symmetrical to the $y$-axis, which leads, for example, to an elliptical failure line with the vertices $T_u$ (static torsion strength) and $T_{alt}$ (alternating torsion strength).

A more sophisticated and material-specific version of the Haigh fatigue diagram is an approximation with a straight line through the point of alternating fatigue strength, $S_{alt}$ ($R = -1$). The declension of the straight line of the fatigue diagram is given by the material-specific mean stress sensitivity, $M$ (Fig. 7):

$$S_a = S_{alt} - M \cdot S_m \quad \text{(Eq 13)}$$

At high mean stresses, the line is even flattening to the declension $M/3$ ($0 \leq R \leq 0.5$) and 0 ($R \geq 0.5$).

Values for the mean stress sensitivities, $M$, for different metals are plotted in Fig. 8 (Ref 5). The mean stress sensitivity depends strongly on the type of material and increases with the static tensile strength, $S_u$, of the materials.

Tests with constant $R$-ratios are characterized in the Haigh diagram by lines through the origin with the slope (Eq 4):

$$S_a/S_m = (1 - R)/(1 + R) \quad \text{(Eq 14)}$$

In addition to the fatigue failure line, the static limits must be included in the fatigue

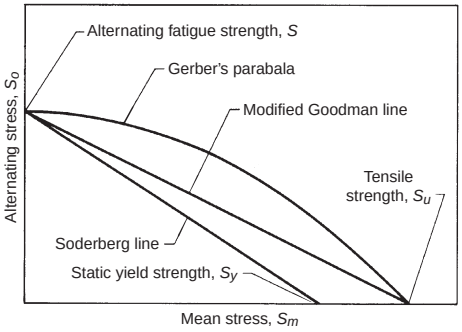


**Fig. 6** Effect of mean stress on the alternating stress amplitude, as shown by the modified Goodman line, Gerber's parabola, and Soderberg line

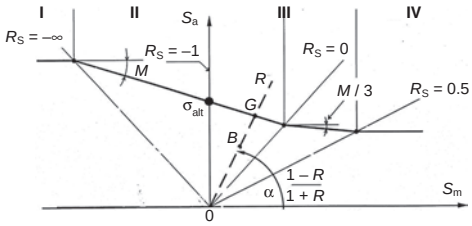


**Fig. 7** Haigh diagram based on mean stress sensitivity, $M$. Source: Ref 1

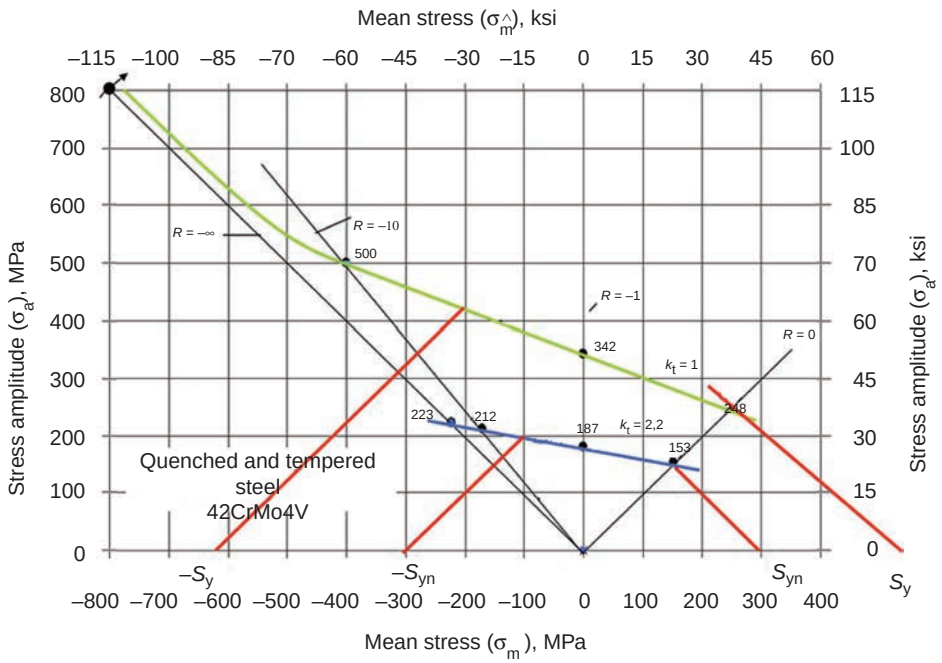


**Fig. 5** Haigh fatigue diagram for unnotched and notched round bars of quenched and tempered 42CrMo4V steel (AISI 1140)

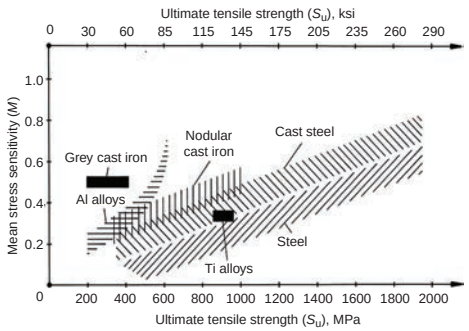


**Fig. 8** Mean stress sensitivity, $M$, for different metals. $K_t$ = 1 to 5; $N = 10^4$ to $10^6$; probability of failure = 50%. Source: Ref 5

diagram on the tension and compression side. This refers to the failure condition for yielding (yield stress, $S_y$) and for collapse (ultimate strength, $S_u$), as in Fig. 5. For example, the yield line is given by the condition:

$$S_{ay} = S_y - S_m \quad \text{(Eq 15)}$$

The fatigue diagram has resulted in many varied and useful treatments for improving fatigue life based on the fact that an increased compressive mean stress increases the fatigue strength. Because most fatigue cracks originate at the surface of the component, placing the hot spot at the surface under compressive stress is beneficial. This can be achieved by surface treatments such as nitriding, carburizing (case hardening), shot peening, needling, surface rolling, and mechanical overstressing. When these treatments are properly applied, the surface zone is—in addition to an increase in hardness (ultimate strength)—in a state of compressive residual stress, which is acting as a mean stress and considerably increases the lifetime in the finite-life range and the fatigue strength. However, if not properly applied, the treatments can have a detrimental effect on fatigue life. For materials that do not have a fatigue limit or for finite-life assessment of structures, the fatigue strength can be substituted and supplemented by the stress amplitude for a given number of cycles (finite-life fatigue diagram).

### Variable-Amplitude Loading

Because in most technical structures the stress amplitude is often not constant but will randomly vary under actual loading conditions, it is necessary to predict fatigue life not only under constant but as well as under variable stress amplitudes. The principle of lifetime determination for variable-amplitude loading is the comparison of the stress spectrum for the considered operation period with the *S-N* curve (Eq 5). The spectrum (number of cycles on different stress levels) is generated from the stress-time or strain-time cycles by application of different counting methods. The superior method is Rainflow counting, because it results not only in the amplitude but also the mean values of the different counted cycles.

The first hypothesis for lifetime assessment under block and variable-amplitude loading assuming a linear damage accumulation was suggested by Palmgren (Ref 6) and restated by Miner (Ref 7). According to the Palmgren/Miner rule, the application of $n_i$ cycles at a stress amplitude $S_{ai}$, for which the number of cycles to failure in the *S-N* curve is $N_i$ (Eq 5), causes a fatigue damage that is expressed by the damage sum:

$$D_i = n_i/N_i \quad \text{(Eq 16)}$$

Fatigue failure is assumed to occur when the total accumulated damage sum, $D_{tot}$, reaches a critical value, $D_{crit}$, which was proposed by Palmgren/Miner to be 1:

$$D_{tot} = \sum(n_i/N_i) = D_{crit} = 1 \quad \text{(Eq 17)}$$

The plot of the calculated number of cycles to failure for different maximum stress amplitudes, called the lifetime curve, is shifted several decades toward a longer lifetime compared to the *S-N* curve, depending on the severity of the stress spectrum.

The life prediction based on this method is often disappointing; numerous alternative methods and modifications of the damage calculation have been proposed (Ref 8). The main modifications of the original Palmgren-Miner rule are, for example, the adoption of an individual critical damage sum, $D_{crit} < 1$ (e.g., 0.3), the damage contribution of cycles below the fatigue limit, the effect of mean stress changes, the inclusion of the cycle sequence effect, including the consideration of the steadily increasing damage, and the separation of the damage calculation and lifetime assessment into the stages of crack initiation and cyclic crack propagation.

Because of the still-existing weaknesses of lifetime prediction with the—even modified—damage-accumulation theories, the behavior of structures under variable-amplitude loading should be evaluated and verified experimentally. Tests under variable-amplitude loading should be conducted with servohydraulic or electromechanical testing equipment on component-like structures under the real loading history. Alternative tests with simplified block programs are conducted as well, but in most cases, these block tests do not simulate the real service conditions and will usually considerably overestimate the lifetime (up to a factor of 10) compared to the real tests with random loading.

### Advantages and Limitations

The fatigue analysis based on *S-N* curves is more reliable because more of the test samples are component-like and simulate the parameters given in Eq 10. In addition, the fatigue limit at a high number of cycles under benign environmental conditions, typically $5 \times 10^6$ for steels and $\geq 10^7$ for aluminum alloys, provides a useful and beneficial result of *S-N* testing. However, other materials do not always exhibit this infinite-life response. Instead, many materials display a continuously decreasing stress-life response, even at a high number of cycles (up to $10^8$ to $10^9$), which is more correctly described by a tolerable fatigue stress amplitude for a given number of cycles. The finite-life behavior in the high-cycle range especially must be considered under corrosive environments (corrosion fatigue) and for crack initiation at inclusions beneath the surface, such as carbides, oxides, and sulfides. The initial crack-growth rate in the so-called very-high-cycle fatigue range is extremely low, for example, $da/dN = 10^{-8}$ mm/cycle = 4* $10^{-10}$ in/cycle.

### Assessing Fatigue S-N Properties

Given the extensive history of the stress-life method, substantial property data are available, but one should beware of the testing conditions employed in producing older data. The usefulness of property data is a critical point due to the numerous variables that influence fatigue results. In many cases, insufficient information is available for the effective use of older *S-N* data. Many necessary details are often missing. A list of important testing parameters may include:

- Material properties (microstructure, mechanical characteristics)
- Type of test sample (smooth or notched specimen, component, size)
- Specimen orientation
- Surface condition
- Stress concentration
- Environment (temperature, corrosion)
- Type of loading (axial, bending, torsion)
- Test-controlling parameter (load, displacement, strain)
- Frequency and waveform
- Mean stress (*R*-ratio)
- Type of stress for evaluation (nominal or local, amplitude or range)
- Damage criteria (crack initiation, change of stiffness, fracture)
- Shape of spectrum for variable-amplitude loading
- Handling and representation of runouts
- Determination of fatigue limit (pearl chain, staircase, probit)
- Statistical evaluation and quantification of scatter

An example of what should be considered important as supporting facts can be found in ASTM International E 468, "Presentation of Constant Amplitude Fatigue Test Results for Metallic Materials." It provides guidelines for presenting information other than just final data.

## Finite-Life Criterion (ε-*N* Curves)

If in cyclic loading the local stresses in the hot spot regions clearly exceed the yield strength of the material, the local strain approach is the suitable and superior method employed for response to cyclic loading in the finite-life regime. This concept is primarily intended to address strain-controlled damage in the low-cycle fatigue (LCF) area (e.g., from approximately $10^2$ to $10^4$ cycles).

The lifetime determination in the LCF range is based on the local strain concept, in which the load-induced elastic/plastic stress-strain cycles in the hot spot of the component must be constructed and counted on the basis of the cyclic stress-strain curve. For the damage calculation, the strain amplitude of each closed loop is compared with a strain-amplitude number of cycles to crack-initiation curve (LCF curve), including a mean stress correction with

damage parameters. The necessary strain amplitude ($\varepsilon_a$)/number of cycles ($N_i$) data are similar to those for the *S-N* curve. However, because strain is the governing damage parameter in LCF, the stress amplitude is replaced by the strain amplitude, which requires strain-controlled testing in the characterization of fatigue crack initiation. The LCF tests are carried out on small, unnotched specimens under pure alternating strain cycles ($R_\varepsilon = -1$) (Fig. 9).

The LCF curve is usually expressed by the Manson-Coffin equation (Ref 9, 10), which expresses the total strain amplitude as the sum of the elastic and the plastic strain as a function of the number of cycles to crack initiation, *N*:

$$\varepsilon_a = \varepsilon_{ael} + \varepsilon_{apl} = \sigma_f'/E(2N)^b + \varepsilon_f'(2N)^c \qquad \text{(Eq 18)}$$

The material parameters $\sigma_f'$ and $b$ are derived from the elastic line of the LCF curve, and the parameters $\varepsilon_f'$ and $c$ from the plastic line.

Whereas the *S-N* curve is mostly based on the total separation of the tested part, the failure criterion in strain-controlled testing is crack initiation, which is reflected in any of the following characteristics (in accordance with ASTM International E 606): modulus ratio, crack of a certain length, or percentage of maximum load drop. This flexibility allows the characterization of fatigue behavior in the low-cycle regime.

Low-cycle fatigue testing is usually more complex and time- and cost-consuming than load-controlled (stress-controlled) *S-N* testing. For example, monitoring and controlling the strain requires servohydraulic or electromechanical testing equipment, continuous extensometer measurements, and a relatively low testing frequency. Further details on testing are given in the article "Fatigue, Creep Fatigue, and Thermomechanical Fatigue Life Testing" in *Mechanical Testing and Evaluation*, Volume 8 of the *ASM Handbook*, 2000.

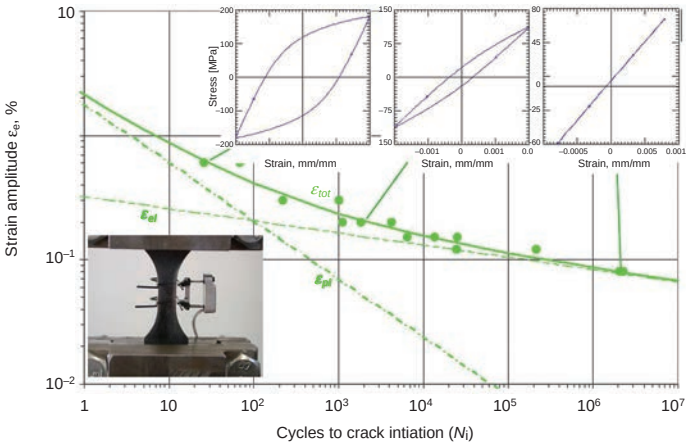


**Fig. 9** Example of low-cycle fatigue curve for a die-cast aluminum alloy

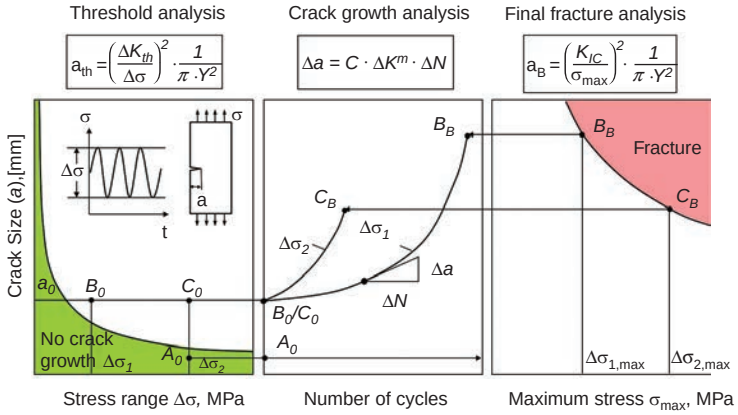


**Fig. 10** Application of fracture mechanics in failure investigation

## Damage Tolerance Criterion

In the latter half of the 20th century, the fracture mechanics approach gained acceptance in the safety evaluation of flawed structures, including the prediction of fatigue life. In this approach, a hypothetical or real cracklike defect in the structure is assumed, which means the period of crack initiation is neglected and only the crack propagation to final failure is considered. In failure analysis, the application of fracture mechanics helps to answer basic questions (Fig. 10) such as:

- Will an existing cracklike defect start to propagate under the cyclic operating load?
- If yes, how fast will the crack grow?
- What are the conditions for the final fracture?

This method can provide a conservative estimate of fatigue life and has considerable success in the lifetime assessment of cracked structures. However, the treatment of short cracks, the propagation of cracks for negative *R*-ratios and for irregular loading, the reaction to complex microcracks, the response to single overloads, the crack closure effects, and the behavior under multiaxial loading still provide significant challenges. Reference 11 and *Fatigue and Fracture*, Volume 19 of the *ASM Handbook*, 1996, cover these topics in more detail.

### *Fatigue Crack-Growth Behavior*

The limitation of the *S-N* and ε-*N* methods is the inability to quantify the effect of cracklike flaws. In traditional stress analysis, basic elasticity calculations show that both stress and strain approach extreme values at a discontinuity such as a crack, where the tip radius tends to approach zero, far exceeding any recognized limiting levels. Even invoking plasticity still leaves inordinately large values or, conversely, extremely low tolerable loads. The solution is the characterization and quantification of the stress field at the crack tip in terms of the stress-intensity factor (*K*) in linear elastic fracture mechanics (LEFM) as well as crack-tip opening displacement (CTOD) or *J*-integral in elastic-plastic fracture mechanics (EPFM).

The application of LEFM is restricted to a very small plastic zone ahead of the crack tip, which restricts the application to the onset of brittle fracture and the subcritical cyclic crack growth in brittle and ductile structures. The stress-intensity factor, *K*, uniquely characterizes the elastic stress field near a crack and represents the crack-driving force. The stress-intensity factor depends on the type of loading, the applied remote nominal stress, σ, the component and crack geometry (geometry factor, *Y*), and the crack size, *a*. The equation for the stress-intensity factor for crack-opening mode I is:

$$K_I = \sigma\sqrt{\pi \cdot a} \cdot Y \qquad \text{(Eq 19)}$$

An important application in failure analysis of flawed structures is the equation for a panel loaded in tension and/or bending with a semi-elliptical surface crack with depth *a* and length

$2c$ (Ref 12); see also American Society of Mechanical Engineers (ASME) XI:

$$K_I = \sigma\sqrt{\pi \cdot a/Q} \cdot M \qquad \text{(Eq 20)}$$

The flaw shape factor, $Q$, depends on the crack geometry ($a/2c$), and the geometry correction factor, $M$, is a function of the crack geometry and the crack-depth-to-wall-thickness ratio, $a/t$. Figure 11 shows diagrams for the flaw shape factor, $Q$, and the membrane geometry factor, $M$, for a tension-loaded panel with a semielliptical surface flaw.

In cyclic loading, the fracture mechanics methods are applied to determine the crack-growth behavior of preexisting cracklike defects. The crack-growth velocity, $da/dN$, is governed by the range of the stress-intensity factor, which is calculated according to Eq 20 and based on the stress range, $\Delta\sigma$:

$$\Delta K = K_{max} - K_{min} = \Delta\sigma\sqrt{\pi \cdot a/Q} \cdot M \qquad \text{(Eq 21)}$$

Because a crack will propagate mainly under opening of the crack flanks, only the positive part of the stress cycle is generally used for $\Delta\sigma$ in Eq 21.

Besides the range of the stress-intensity factor, other variables that influence fatigue crack-growth behavior are factors such as the stress ratio, $R$, material chemistry, heat treatment, material anisotropy (sample orientation), temperature, and corrosive environment. The loading frequency is usually not a decisive factor for metals in a benign environment.

Fatigue crack-propagation behavior of metals generally can be divided into three regions, as shown in Fig. 12 (Ref 13); see also Fig. 10. In threshold region 1, characterized by the threshold value $\Delta K_{th}$, crack growth is zero or at least negligible for an almost unlimited number of cycles. In crack growth region 2, the crack growth velocity $da/dN$ is expressed by the parameters $C$ and $n$. Region 3 or the final fracture is characterized by the fracture toughness $K_{Ic}$.

The complete S-shaped crack-growth curve $da/dN = f(\Delta K)$ in Fig. 12 can be expressed by the Erdogan-Ratwani equation (Ref 14):

$$da/dN = C_E \cdot (\Delta K - \Delta K_{th})^{n_E}/[(1 - R) \cdot K_{Ic} - \Delta K] \qquad \text{(Eq 22)}$$

This equation includes the stress-intensity range, $\Delta K$, according to Eq 21 and, as fracture mechanics material properties, the threshold value $\Delta K_{th}$, the crack-growth parameters $C_E$ and $n_E$, and the fracture toughness, $K_{Ic}$. Usually, the crack-growth parameters are adapted to $K$-values expressed in MPa$\sqrt{m}$ (ksi$\sqrt{in.}$).

An even more sophisticated crack-growth law is the equation according to Forman and Mettu (the NASGRO equation), which considers the shape of the transition regions 1 to 2 and 2 to 3 and the crack-closure effect by a crack-opening function, $\gamma$.

The crack-growth rate in the linear middle section of region 2 in the crack-growth curve can be characterized with the simplified power-law relationship according to Paris (Ref 15) (Fig. 12):

$$da/dN = C_P \cdot \Delta K^{n_P} \qquad \text{(Eq 23)}$$

The fatigue crack-growth threshold value, $\Delta K_{th}$, is an important design parameter for low-stress, high-frequency fatigue-loading applications where small cracks exist and no crack extension during service can be permitted. No cyclic crack propagation of an existing crack can be expected if:

$$\Delta K = \Delta\sigma\sqrt{\pi \cdot a/Q} \cdot M < \Delta K_{th} \qquad \text{(Eq 24)}$$

ASTM International E 647 defines $\Delta K_{th}$ as that asymptotic value of $\Delta K$ at which $da/dN$ approaches zero. For most materials, an operational, although arbitrary, definition of $\Delta K_{th}$ is that $\Delta K$ which corresponds to a fatigue crack-growth rate of typically $da/dN = 10^{-7}$ mm/cycle ($4 \times 10^{-9}$ in./cycle). The fatigue crack threshold, $\Delta K_{th}$, is a function of a number of variables, including the material, the $R$-ratio, and the environment.

## Short Cracks

These types of cracks can grow at stress-intensity levels below the threshold level for macroscopic cracks, as demonstrated by Pearson in 1973 (Fig. 13) (Ref 16). Small cracks

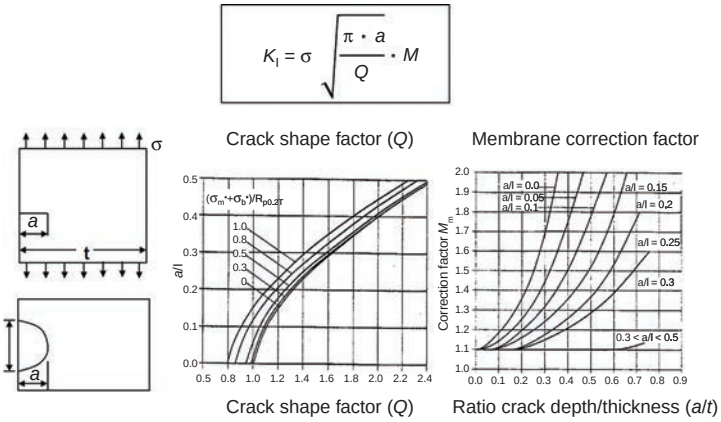


**Fig. 11** Flaw shape factor, $Q$, and geometry correction factor, $M$, for a tension-loaded bar with a semielliptical surface flaw. Source: Ref 12, ASME XI

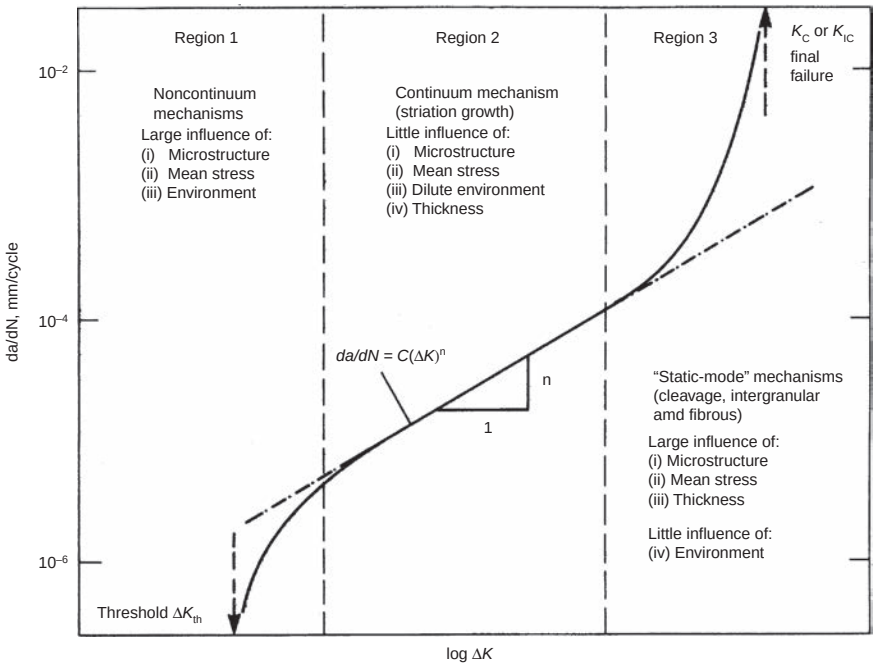


**Fig. 12** Schematic illustration of variation of fatigue crack-growth rate, $da/dN$, with alternating stress intensity, $\Delta K$, in steels, showing regions of primary crack-growth mechanisms. Source: Ref 13

have also been observed to grow at nonnegligible rates when the nominal applied $\Delta K$ is less than the threshold value determined from traditional (large-crack) test methods. Therefore, a structural life assessment based on large-crack analysis methods can be nonconservative if the life is dominated by small-crack growth.

Small-crack behavior is a complex subject, due to the variety of factors that may affect small cracks and the variety of microstructures used in engineering components. In contrast to large-crack growth rates, which generally increase with increasing $\Delta K$, small-crack growth rates are dominated by local conditions of the microstructure and sometimes are observed to increase, decrease, or remain constant with increasing $\Delta K$.

When small cracks are being evaluated, it is important to recognize that all small cracks are not the same (Ref 17). Different mechanisms are responsible for different types of small-crack effects in different settings. Criteria that properly characterize small-crack behavior in one situation may be entirely inappropriate in another situation. Therefore, before selecting suitable analytical treatments, it is critical to understand the different types of small cracks, such as microstructurally small, mechanically small, and chemically small cracks (Ref 17).

When the surface-crack length is larger than approximately 0.5 to 1 mm (0.02 to 0.04 in.), then a conventional fracture mechanics law, such as Eq 24, can be applied to calculate the threshold condition (Ref 18). However, as a crack under consideration is made smaller and smaller, there is a transition from LEFM treatment of long cracks at threshold to an endurance-limit-dominated behavior of short cracks. The Kitagawa diagram is used to depict this type of situation. Below a surface-crack length of 1 mm (0.04 in.), the threshold stress range departs gradually from its macroscopic value and approaches a constant stress range level that is approximately equal to the fatigue limit of unnotched smooth specimens (Ref 18).

## Large Cracks

*Region 1*

Region 1 cracks are characterized by the threshold value $\Delta K_{th}$ for various steels subjected to different stress ratios, $R$. The threshold value can be predicted (Ref 19) for $R \geq +0.1$ in Anglo-American units:

$$\Delta K_{th}(\text{ksi}\sqrt{\text{in.}}) = 6.4(1 - 0.85R) \quad \text{(Eq 25a)}$$

or in SI units:

$$\Delta K_{th}(\text{MPa}\sqrt{\text{m}}) = 7.0(1 - 0.85R) \quad \text{(Eq 25b)}$$

The respective equations for steels and $R < 0.1$ are:

$$\Delta K_{th}(\text{ksi}\sqrt{\text{in.}}) = 5.5 \quad \text{(Eq 26a)}$$

$$\Delta K_{th}(\text{MPa}\sqrt{\text{m}}) = 6.0 \quad \text{(Eq 26b)}$$

These equations indicate that the fatigue crack-propagation threshold values for steels in ambient environment are primarily a function of the stress ratio and are essentially independent of chemical or mechanical properties for a given material.

*Region 2*

Crack-growth behavior in region 2 is relevant to design situations involving a finite number of fatigue cycles. Extensive fatigue crack-growth rate data for various steels show that the primary parameter affecting growth rate in region 2 is the stress-intensity factor range and that the mechanical and metallurgical properties of these steels have only a small effect on the fatigue crack-growth rate in room-temperature air environment.

The data for martensitic steels fall within a single band (Fig. 14), and the upper bound of scatter can be obtained on the basis of the Paris law (Eq 23), as (Ref 19):

$$da/dN = 0.66 \times 10^{-8}(\Delta K)^{2.25} \quad \text{(Eq 27a)}$$

with $da$ in inches and $\Delta K$ in units of $\text{ksi}\sqrt{\text{in.}}$, or:

$$da/dN = 13.6 \times 10^{-8}(\Delta K)^{2.25} \quad \text{(Eq 27b)}$$

with $da$ in millimeters and $\Delta K$ in units of $\text{MPa}\sqrt{\text{m}}$.

Similarly, as shown in Fig. 15 (Ref 19), data for ferrite-pearlite steels fall within a single band (different from the band for martensitic steels), and the upper bound of scatter can be calculated from:

$$da/dN = 3.6 \times 10^{-10}(\Delta K)^{3.0} \quad \text{(Eq 28a)}$$

with $da$ in inches and $\Delta K$ in units of $\text{ksi}\sqrt{\text{in.}}$, or:

$$da/dN = 69.0 \times 10^{-10}(\Delta K)^{3.0} \quad \text{(Eq 28b)}$$

with $da$ in millimeters and $\Delta K$ in units of $\text{MPa}\sqrt{\text{m}}$.

The mean stress (stress ratio) has a smaller effect on the rate of crack growth in region 2 compared to region 1. Also, the frequency of cyclic loading and the waveform (sinusoidal, triangular, square, or trapezoidal) do not

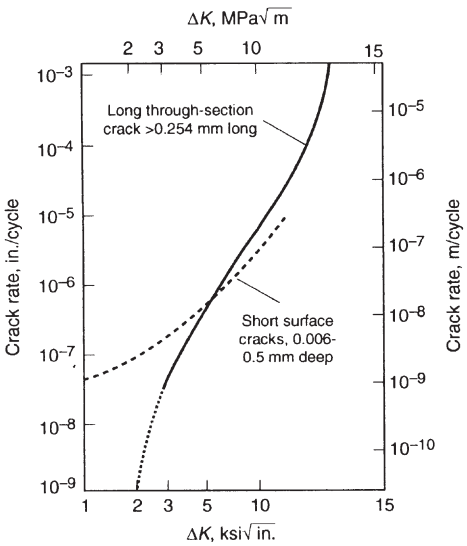


**Fig. 13** Pearson plot of crack-growth rate as a function of $K$ for short surface cracks and through-cracks. Source: Ref 16

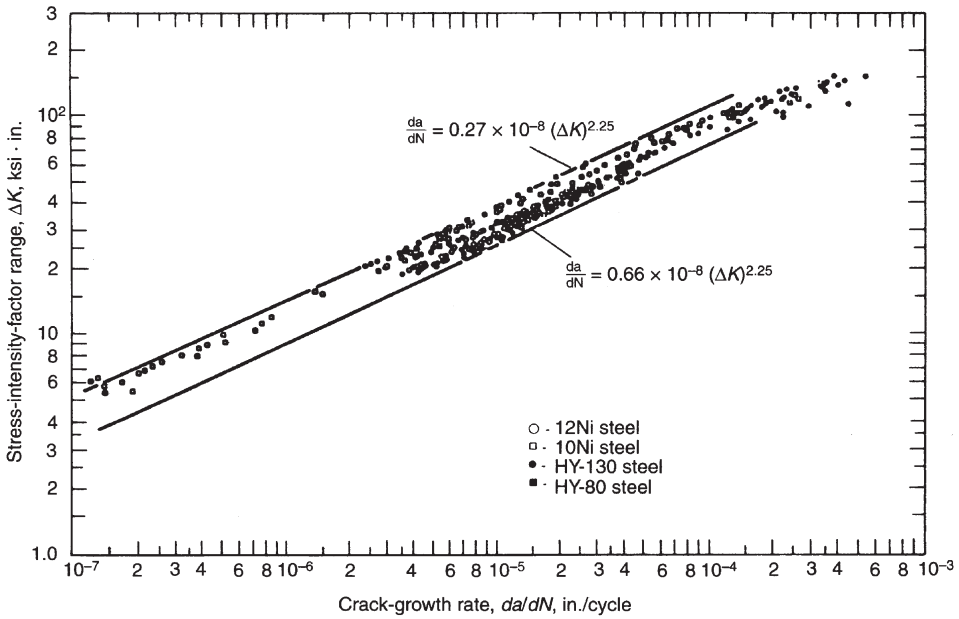


**Fig. 14** Summary of fatigue crack-growth data for martensitic steels. Source: Ref 19

significantly affect the rate of crack propagation for steels in benign environments (Ref 19).

The cyclic crack-growth behavior in region 2 and the lifetime calculation of cracked structures are based on the incremental crack growth, $\Delta a$, for a given number of cycles, $\Delta N$. Applying the Paris law (Eq 23) results in:

$$\Delta a = C \cdot \Delta K^m \cdot \Delta N = C \cdot \left(\Delta\sigma\sqrt{\pi \cdot a/Q} \cdot M\right)^m \cdot \Delta N \quad \text{(Eq 29)}$$

*Region 3*

The acceleration of fatigue crack-growth rates that determines the transition from region 2 to region 3 is caused by the superposition of a brittle or a ductile tearing mechanism onto the mechanism of cyclic subcritical crack extension. These mechanisms occur when the strain at the tip of the crack, the maximum stress-intensity factor, reaches a critical value. Thus, the fatigue rate transition from region 2 to region 3 depends on the maximum stress-intensity factor, the stress ratio, and the fracture properties of the material.

The final fracture condition in the crack-growth curve plot (Fig. 12) is given by a stress-intensity factor range of:

$$\Delta K_c = K_{Ic} \cdot (1 - R) \quad \text{(Eq 30)}$$

where $K_{Ic}$ is the fracture toughness of the material. In case of a brittle final fracture, LEFM can be applied to predict the critical fatigue crack depth using Eq 20 and 30:

$$a_{crit} = \Delta K_c{}^2 \cdot Q/(\sigma_{max}{}^2 \cdot \pi \cdot M^2) \quad \text{(Eq 31)}$$

The critical fatigue crack depth, which leads to final fracture, is consequently determined by the toughness, $K_{Ic}$, and by the maximum stress, $\sigma_{max}$; see examples of fracture surfaces in Fig. 16.

Equation 31 is only valid for LEFM behavior. Final fractures with increased ductility must be expressed for crack-tip-controlled failures in terms of EPFM (CTOD, *J*-integral) or in ligament-controlled failures, with collapse calculation (slip line theory).

## Environmentally Assisted Fatigue Crack Propagation

The environmentally assisted fatigue crack growth in metals is a complex phenomenon and is influenced by loading, material, and environmental variables. In a given material, in addition to the environmental effects, crack-growth rates can vary widely with variations in loading parameters, such as stress state, stress-intensity range, and stress wave shape and frequency. For example, crack size and geometry may alter the electrochemical environment near the crack tip and consequently change the crack-growth rate and may, in extreme cases, change the controlling process and the growth mechanism.

The crack-growth rate in a corrosive environment is usually significantly faster than under ambient conditions. For example, the fatigue crack-growth rate for a small crack in a high-strength steel (4130) in a 3% sodium chloride solution is approximately ten times faster than that of long cracks at similar stress-intensity levels in the same bulk environment, although the short-crack fatigue data obtained in air do coincide with those for long cracks. Therefore, the difference is purely chemical in nature.

The magnitude of the environmental enhancement of crack growth varies, depending on crack size and applied stress. At a low-stress range, crack-growth rates exceed those observed in air by 2 orders of magnitude and are faster than the values reported for long cracks, while at a high-stress range, the rate of cracking is nearly equivalent to that reported for long cracks. For low-strength steels, the crack-size effect does not appear to be as large. These factors and others are discussed in more detail in the article "Mechanisms of Corrosion Fatigue" in *Fatigue and Fracture*, Volume 19 of the *ASM Handbook*, 1996.

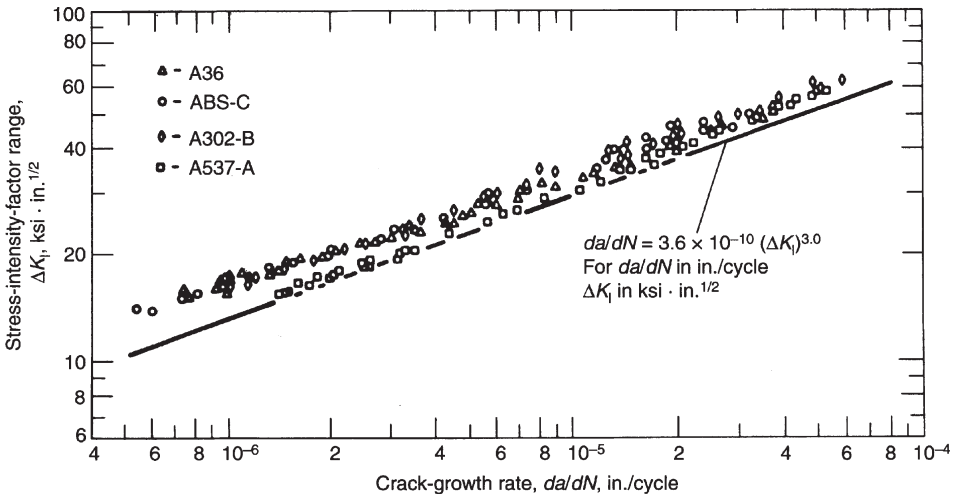


**Fig. 15** Summary of fatigue crack-growth data for ferrite-pearlite steels. Source: Ref 19

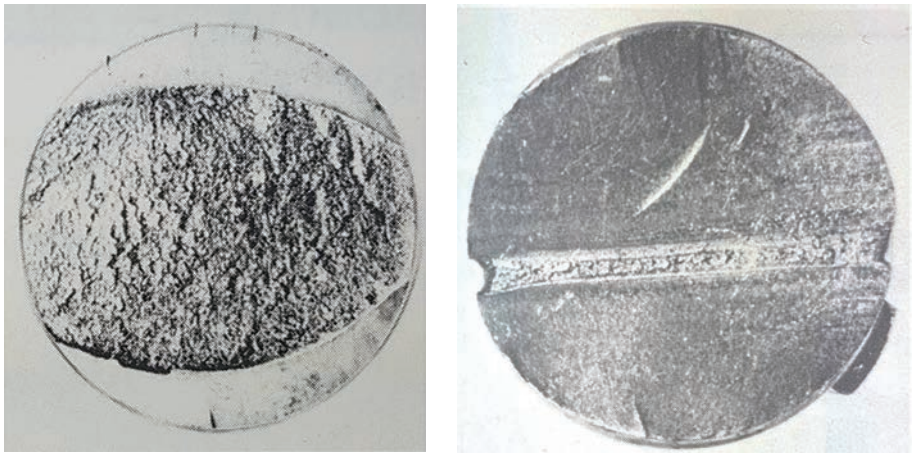


**Fig. 16** Fatigue failures with different fatigue crack size/final fracture ratios. (a) High stress level, low toughness. (b) Low stress level, high toughness

## Life Prediction

The fracture mechanics approach for characterizing fatigue crack growth can be used in design applications to estimate the maximum flaw sizes that allow a part to reach its design life. This approach is also very useful for conducting a failure analysis.

Because predicting service fatigue life often involves integrating crack-growth rates over a range of $\Delta K$ values, the $da/dN$ versus $\Delta K$ relationship can be represented by:

$$\frac{da}{dN} = \frac{A_l}{(\Delta K)^{nl}} + A_2\left[\frac{1}{(\Delta K)^{n2}} - \frac{1}{[(1-R)K_c]^{n2}}\right] \quad \text{(Eq 32)}$$

where $n_1$, $n_2$, $A_1$, $A_2$, and $K_c$ are best-fit material constants determined by regression analysis. This equation is general and applies to all three regions of crack growth.

### Example 1: Developing Inspection Criteria for Fan Shafts.

Industrial fan shafts are typically forged from medium-carbon steels. Primary fatigue stresses are generated from gravity bending and are fully reversed during each revolution at rates of 900 to 1800 revolutions per minute or 15 to 30 Hz. The maximum alternating stress in the fan shaft is typically ±52 MPa (7.5 ksi). Before shipping, shaft forgings are subjected to various types of nondestructive inspections to reveal inclusions and forging bursts or other defects. Fracture mechanics fatigue crack-propagation analysis can be used to evaluate the consequence of such defects.

*Investigation*

Specimens for mechanical testing were obtained by slicing a 580 mm (23 in.) diameter shaft forging into 75 mm (3 in.) thick disks. Tensile specimens and Charpy impact specimens were located at various radial distances from the center of the shaft cross section. Compact-tension specimens, 25.4 and 6.25 mm (1 and 0.25 in.) thick, were used for fracture toughness and fatigue crack-growth rate testing. In addition, 12.5 by 50.8 mm (0.5 by 2 in.) center-crack-tension specimens were also machined from the central region of another disk from the same forging for characterizing fatigue crack growth at $R = -1.0$.

Test results at several radial distances from the shaft center indicated that tensile strength increases slightly from the peripheral region of the shaft toward the center. Fatigue crack-growth testing was conducted on specimens taken from two extreme radial locations of the shaft, but no trend could be attributed to specimen location. Figure 17 shows the wide range of fatigue crack-growth rate behavior of the shaft material for load ratios $R = -1.0$, 0.1, and 0.5, which represent a spectrum of typical load histories. For a load ratio of $R = -1.0$, only the positive half of the load cycle was used to compute $\Delta K$. These data were used to determine the material constants $A_1$, $A_2$, $n_1$, and $n_2$ (Eq 32) using a standard multiple linear regression computer program.

*Discussion*

The number of fatigue cycles that accumulate on the fan shaft during the projected 40 year service life can be approximately $10^{11}$ cycles, but only very little crack extension during service is acceptable. Thus, the threshold value $\Delta K_{th}$ of the shaft material for completely reversed loading ($R = -1.0$) is selected as the maximum allowable value of $\Delta K$.

*Conclusions*

A growth rate of $10^{-12}$ mm/cycle ($4 \times 10^{-13}$ in./cycle) can be considered acceptable for this application because it implies a crack extension of 0.1 mm (0.004 in.) during $10^{11}$ cycles. The $\Delta K$ value corresponding to this growth rate can be determined from Eq 25 by using the appropriate fitting constants at $R = -1.0$. This value is 6.4 MPa$\sqrt{m}$ (5.8 ksi$\sqrt{in.}$), and a safety factor of 20% on the $\Delta K_{th}$ value reduces the allowable $\Delta K$ to 5.3 MPa$\sqrt{m}$ (4.8 ksi$\sqrt{in.}$).

Two common elliptical defect geometries are shown in Fig. 18 with the flaw shape factor, $Q$, for various values of the flaw aspect ratio, $a/2c$, and ratios of applied stress to the yield strength of the material. The value of $Q$ influences the magnitude of $K$ according to Eq 25. For fatigue, this relationship can be written as:

$$\Delta K = M_T \Delta\sigma \sqrt{\frac{\pi\alpha}{Q}} \qquad \text{(Eq 33)}$$

where $\Delta\sigma$ is the applied stress range, and $M_T$ is the finite-width magnification factor. Substituting the allowable value of $\Delta K$ results in the critical flaw size, $a_{cr}$:

$$a_{cr} = \left(\frac{5.1}{M_T \Delta\sigma}\right)^2 \frac{Q}{\pi} \qquad \text{(Eq 34)}$$

The $M_T$ value for deeply embedded internal flaws is 1.0, and for small surface defects in large shafts the value is 1.1. These values are used to calculate critical flaw sizes of various aspect ratios for several fatigue stress levels.

## Characteristics of Fatigue Fractures

After initiation, the fatigue crack propagates under the applied stress through the material until a complete fracture results. The process of fatigue consists of three stages:

- **Stage I:** Initial fatigue damage leading to crack nucleation and crack initiation
- **Stage II:** Progressive cyclic growth of a crack (crack propagation)
- **Stage III:** *(Stage I)* Final fracture of the remaining cross section

Fatigue cracks generally initiate at stage I in a highly stressed region of a component subjected to cyclic stresses of sufficient magnitude. On the microscopic scale, the most important feature of the fatigue process is the nucleation of one or more cracks under the influence of cyclic stresses, followed by the development of microcracks at persistent slip bands within grains or at grain boundaries. Subsequently, fatigue cracks propagate in stage II by a series of opening and closing motions at the tip of the crack that produce, within the grains, striations parallel to the crack front. Stage III occurs during the last stress cycles when the remaining cross section cannot sustain the applied load. The final fracture, which can be brittle, ductile, or a combination of the two (semiductile), is not a fatigue process and is discussed in more detail in the article "Monotonic Overload and Embrittlement" in this Volume.

Figure 19 shows the macroscopic characteristics of a typical fatigue fracture. The fatigue fracture should be examined by macroscopic and microscopic measures. The examination of the fatigue fracture surfaces usually begins visually by naked eye or magnifying glass or by use of an optical stereomicroscope at low magnification (up to 50×). Macroscopic examination of fracture surfaces involves relatively simple techniques. It can often be done at the failure site, requires little or no preparation of the specimen, requires minimal and relatively simple equipment, and does not necessarily destroy the specimen or alter the fracture

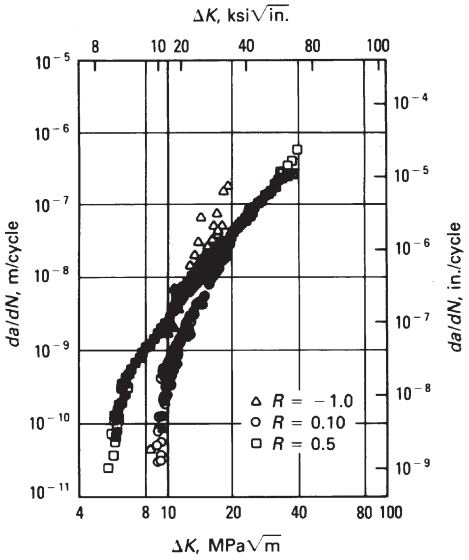


**Fig. 17** Room-temperature fatigue crack-growth behavior of ASTM International grade A293 steel. Levels of $R$ represent spectrum of typical in-service load history.

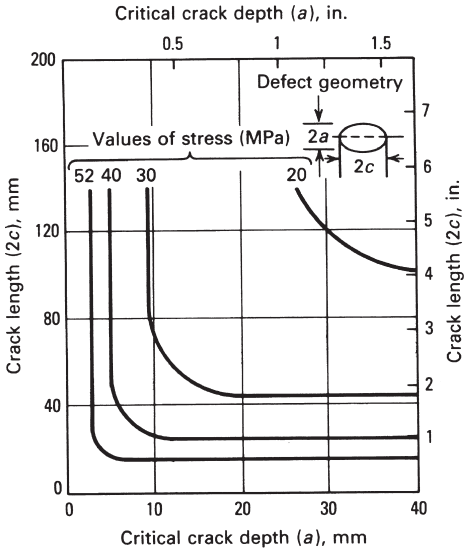


**Fig. 18** Flaw shape parameter curves for surface and internal cracks

surfaces. Macroscopic examination is particularly useful to identify the fracture as a fatigue failure, to explain the fracture location and direction of crack propagation, to detect the possible crack-initiation points, to determine the size of the fatigue crack, and to correlate the fracture-surface characteristics with the part size and shape as well as with the loading and environmental conditions.

Of special importance is the detection and characterization of the fatigue crack origin, because often tiny local effects, such as scratches, burrs, pits, or surface flaws, are the reason for the crack initiation. The crack origin can best be found by visual examination or by viewing the fracture surface at low magnifications (5 to 50×). The findings can later be confirmed by use of the scanning electron microscope (SEM). A metallographic examination of the cross sections through the suspected fatigue crack origin is useful to describe the surface condition, the microstructure, possible imperfections at the initiation site, and to analyze the crack path. The fatigue crack path is mostly transgranular, although intergranular fatigue does occur at times. Useful is the determination of (micro) hardness in the metallographic mount at or near the crack-initiation spot, in order to describe the condition of the surface zone (e.g., decarburization, surface hardening, cold working).

## *Crack Initiation*

In general, fatigue cracks initiate and propagate in regions where the stresses and strains are most severe. The initiation is often at the component surface, where the bending and torsion stresses reach maximum values and geometric notches may be present. Because in some cases the engineering materials contain defects, which form regions with stress concentrations, fatigue cracks often may be initiated at structural defects. Under the action of cyclic loading, a plastic zone develops at the defect tip. Local surface imperfections, such as scratches, mars, burrs, and other fabrication flaws, are the most obvious flaws at which fatigue cracks start. Other fatigue crack-initiation regions are corrosion spots or wear-induced surface damage at the component surface.

Surface and subsurface material inhomogeneities in critical locations also influence crack initiation. Nonmetallic inclusions, hard precipitate particles, and crystal discontinuities such as grain boundaries and twin boundaries are examples of microscopic stress concentrators in the material matrix. The fatigue crack may initiate below the surface in the case of mild or unnotched structures with internal hard particles, such as oxides, carbides, and sulfides, buried imperfections, and cracks. In surface-hardened structures, the crack may initiate at the interface (Fig. 20).

### Mechanisms and Appearances

The initiation stage is the submicroscopic alteration of the crystal structure of a metal by repetitive stressing of the component. In this stage, fatigue cracks nucleate and coalesce by slip plane fracture, extending inward from the surface on a plane of maximum shear stress at approximately ±45° to the principal stress axis. On the submicroscopic scale, dislocation density, lattice defects, and the orientation of cross slip planes control the formation of persistent slip bands, followed by intrusions and extrusions at the surface and dislocation cells, which ultimately control the nucleation mechanism.

Stage I is difficult to envision or describe in failure analysis because of its submicroscopic nature. Even with electron microscopes, this initial stage is difficult to examine and may not be discernible at all, depending on geometry, material, environment, and stress level. For example, a stage I fracture extends over two to five grains around the origin. In each grain, the fracture surface is along a well-defined crystallographic plane, which should not be confused with a cleavage fracture, although it has the same quasi-brittle appearance.

On a microscopic level, cracks develop at persistent slip bands or at grain boundaries. The local directions of propagation are controlled to some extent by crystallographic planes and may form on or may be parallel to slip bands in grains near the surface. In interior grains, cleavage cracks often form at the intersection of slip bands with grain boundaries. A variety of crystallographic features has been observed to nucleate fatigue cracks. In pure metals, tubular holes that develop in persistent slip bands, extrusion/intrusion pairs at free

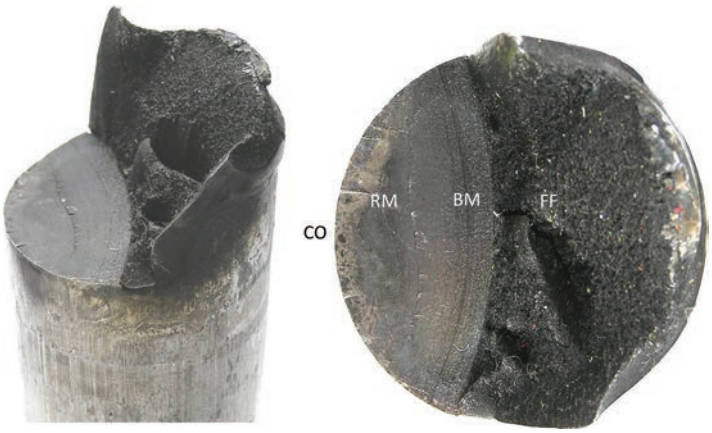


**Fig. 19** Macroscopic view of fatigue fracture in a bending-loaded bolt. CO, crack origin; RM, ratchet marks; BM, beach marks; FF, final fracture

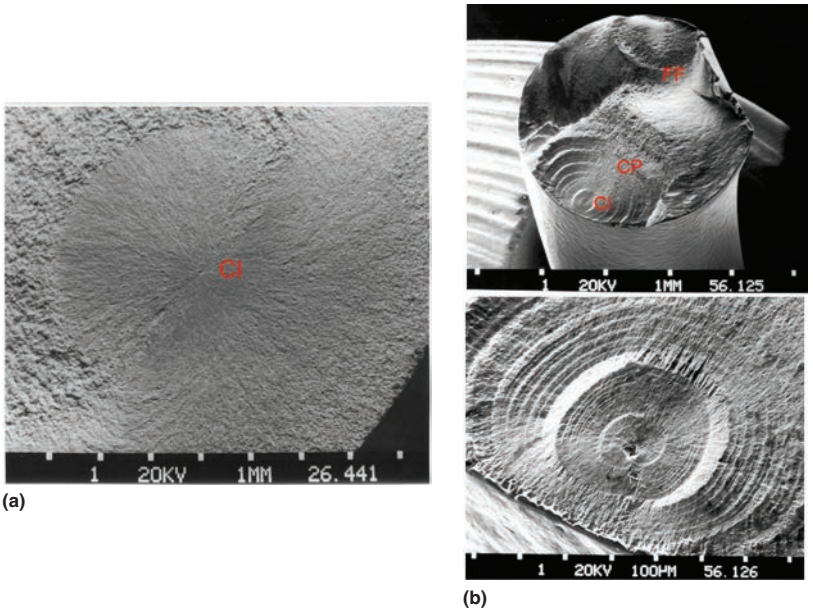


**Fig. 20** Examples of fatigue crack initiation below the surface. (a) Crack initiation (CI) at the interface of a case-hardened shaft and (b) at an embedded oxide in a helical spring. CP, crack propagation; FF, final fracture

surfaces, and twin boundaries are common sites for crack initiation.

Cracking may also initiate at grain boundaries in polycrystalline materials, even in the absence of inherent grain-boundary weaknesses. At high strain rates, the grain boundary seems to be the preferred site for crack nucleation. Nucleation at a grain boundary appears to be purely a geometrical effect related to plastic incompatibility, whereas nucleation at a twin boundary is associated with active slip on crystallographic planes immediately adjacent and parallel to the twin boundary.

The aforementioned processes also occur in alloys and heterogeneous materials. However, alloying and commercial production practices introduce segregation, inclusions, second-phase particles, and other features that disturb the structure and have a dominant effect on the crack-initiation process. In general, alloying that enhances cross slip or twinning and increases the work-hardening capability will retard crack nucleation. On the other hand, alloying usually raises the flow stress of a metal, which may speed up crack propagation, thus at least partly offsetting the potentially beneficial effect on fatigue crack nucleation.

In high-strength materials containing spheroidal second-phase particles, secondary crack nucleation ahead of the main crack front (slipless fatigue) occurs around such particles. In slipless fatigue, cracks form along lattice planes that are unfavorably located relative to the maximum principal stress.

### Relation to Environment

In observing locations of crack initiation, the possibility of environment-related mechanisms—including pitting corrosion, fretting corrosion, stress-corrosion cracking, and other effects of a hostile environment—must be considered. For example, a high number of fatigue failures in structures with low nominal stresses originate in fretted areas. In any structure having joints with micromotion, fretting provides a possible failure origin. Once a fatigue crack has initiated, its rate and direction of growth are controlled by localized stresses and by the structure of the material at the tip of the crack.

## *Fatigue Crack Propagation (Stage II)*

### Initial Propagation

The transition from stage I to stage II fatigue fracture is the change of orientation of the main fracture plane. The transition is from one or two shear planes in stage I to many parallel plateaus separated by longitudinal ridges in stage II. The plateaus are usually normal to the direction of the maximum principal stress (Fig. 21a). A transition from stage I to stage II in a coarse-grained specimen of aluminum alloy 2024-T3 is shown in Fig. 21(b). The presence of inclusions rich in iron and silicon did not affect the fracture path markedly. The fractured inclusions ranged from 5 to 25 μm in diameter. In Fig. 21(b), the stage II area shows a large number of approximately parallel fatigue patches containing very fine fatigue striations that are not resolved at the magnification used. Fine striations are typical in the initial stage II but are usually seen only under very high magnification.

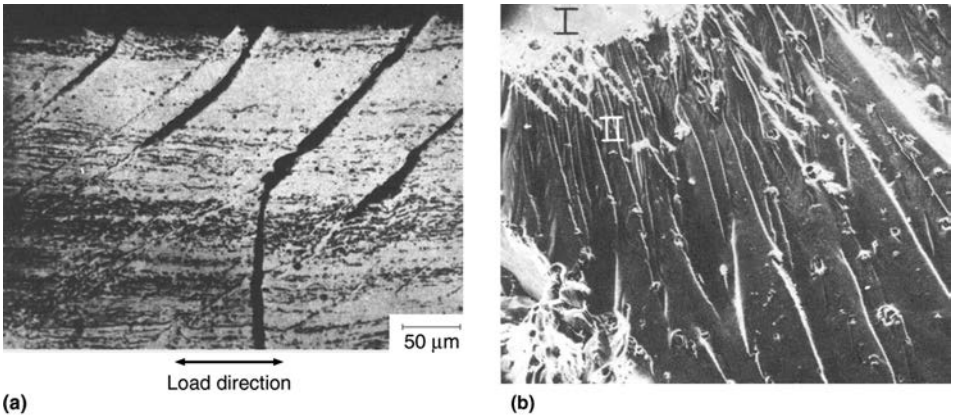


**Fig. 21** Transition from stage I to stage II fatigue. (a) Change in fracture path from stage I (top of photo) to stage II (bottom). (b) Transition from stage I to stage II of a fatigue fracture in a coarse-grained specimen of aluminum alloy 2024-T3. Source: Ref 20

Below the tip of fatigue cracks, the local stress state, due to the constraint effect of the sharp fatigue crack, is triaxial tension (plane strain). This reduces the local apparent ductility of the structure and helps control the orientation of the crack as long as the crack is small. After a crack has initiated and propagated to a larger size, it becomes a macroscopic stress raiser and can be more influential than any stress raiser already in the part. At this point, the crack tip will take over control of the fracture direction. Subsequently, the orientation of the crack surface will depend on the stress field at the crack tip and will often follow a series of void coalescences in advance of the crack front. Early crack extension occurs under plane-strain conditions and gives a typical fine-grained, flat-face surface that, when produced under random loading with sequences of high- and low-stress amplitudes, exhibits characteristic beach marks.

As a general rule, low stress in the high-cycle fatigue range produces even, flat-face (plane-strain) fractures. The fracture surface appears fine grained and lightly polished near the crack-initiation site, where the stress intensification is least. The surface becomes progressively rougher and more fibrous as the crack grows and the stress intensity increases. On high-stress, low-cycle fatigue surfaces, found in certain areas of all complete fatigue fractures, the surface is fibrous, rough, and more typical of plane-stress loading conditions, where the general fracture direction is changing to the maximum shear planes at ±45° to the maximum principal stress direction.

Crack propagation variations in anisotropic or inhomogeneous materials occasionally produce beach marks that are difficult to interpret. In highly anisotropic material such as spring wire, for example, a fatigue fracture surface may exhibit regions where the crack propagated in typical fatigue and other regions where it propagated preferentially in another mode along planes of weakness. For example, beach marks or striations, which are the identifying feature of fatigue in many components, are not always present in spring fatigue fracture surfaces. Beach marks are more likely to be present if the spring material is comparatively soft (40 to 45 HRC), but they are seldom present when higher-strength spring materials fail by fatigue (Ref 21).

In large structural components, the existence of a crack does not necessarily imply imminent failure of the part. Significant structural life may remain in the cyclic growth period of the crack to a size at which a critical failure occurs. The growth or extension of a fatigue crack under cyclic loading is principally controlled by the maximum stress range and the stress ratio. However, as in crack initiation, a number of additional factors may exert a strong influence, including environment, frequency, temperature, grain direction, and other microstructural factors.

### Beach Marks

The most characteristic features usually found on fatigue fracture surfaces are beach marks. The beach marks often show a semielliptical contour and are centered around a common point that corresponds to the fatigue crack origin. Also called clamshell, conchoidal, and arrest marks, beach marks are perhaps the most important characteristic feature in identifying fatigue failures in macroscopic examination. Beach marks can occur as a result of changes in loading or frequency or by oxidation or corrosive attack of the fracture surface during periods of crack arrest from intermittent service of the component.

Figure 22 shows an example of beach marks on the fracture surface of an aluminum alloy

7075-T73 forging. The beach marks are produced by changes in crack-growth rates due to the changing stress intensity at the crack tip. The initial crack-growth rate increases with the increasing stress intensity at the crack tip as the crack extends further.

Beach marks are also produced by other factors that cause changes in crack-growth rates during propagation. For example, Fig. 23 shows the surface of a fatigue fracture in a 4130 steel shaft that failed in service. This surface exhibits beach marks produced by oxidation of the fracture when the shaft was idle during a certain operation period.

Beach marks are usually produced by load variations during service, spectrum-load fatigue testing, or changes in stress intensity from geometric features or material discontinuities of the part. However, fatigue fractures produced under conditions of uninterrupted continuous crack growth and without load variations do not exhibit beach marks. Consequently, beach marks cannot be expected in laboratory fatigue tests under constant-amplitude loading (*S-N* tests) in a noncorrosive environment. Moreover, a fracture surface may sometimes have features that look like fatigue beach marks (Fig. 24).

### Ratchet Marks

These marks are macroscopic features that may be seen on fatigue fractures in the vicinity of the crack initiation. Ratchet marks are the result of multiple fatigue crack origins, each initially producing a separate small fatigue crack zone on different planes. As two approaching cracks on different levels meet, a small step is formed pointing perpendicular to the surface toward the crack-propagation direction (Fig. 25). The small steps around the periphery are the ratchet marks. Although ratchet marks are most apparent in the radial direction on the periphery of fractures in shafts, the stepped appearance is characteristic whenever fatigue cracks emanate from several origins and subsequently meet to form one principal crack front.

The occurrence of ratchet marks requires almost simultaneous initiation of several fatigue cracks, a situation that is favored by

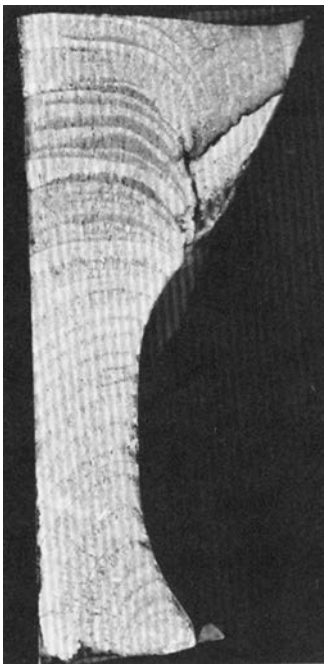

**Fig. 22** Beach marks on a fatigue fracture in aluminum alloy 7075-T73 forging. The light-colored reflective bands are zones of fatigue crack propagation. At high magnifications, thousands of fatigue striations can be resolved within each band. The dull, fibrous bands are zones of crack propagation by microvoid coalescence.

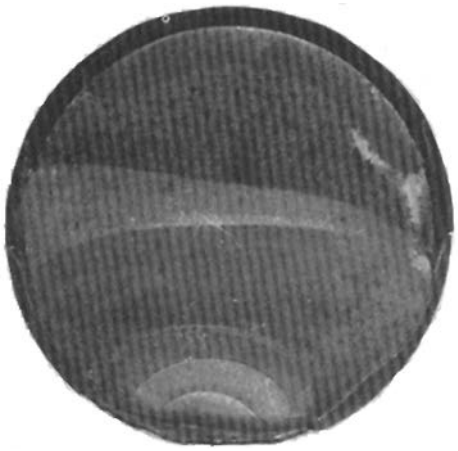

**Fig. 23** Fracture surface of steel shaft with beach marks produced by oxidation

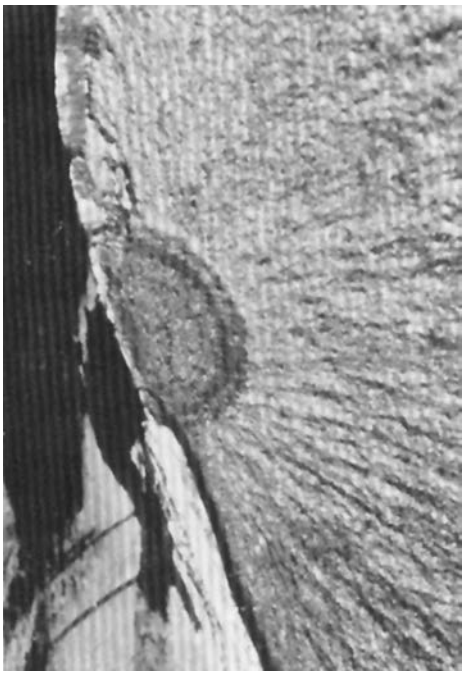

**Fig. 24** Beach marks on a 4340 steel part caused by stress-corrosion cracking. Tensile strength of the steel was approximately 1780 to 1900 MPa (260 to 280 ksi). The beach marks are a result of differences in the rate of penetration of corrosion on the surface. They are not related to fatigue marks.

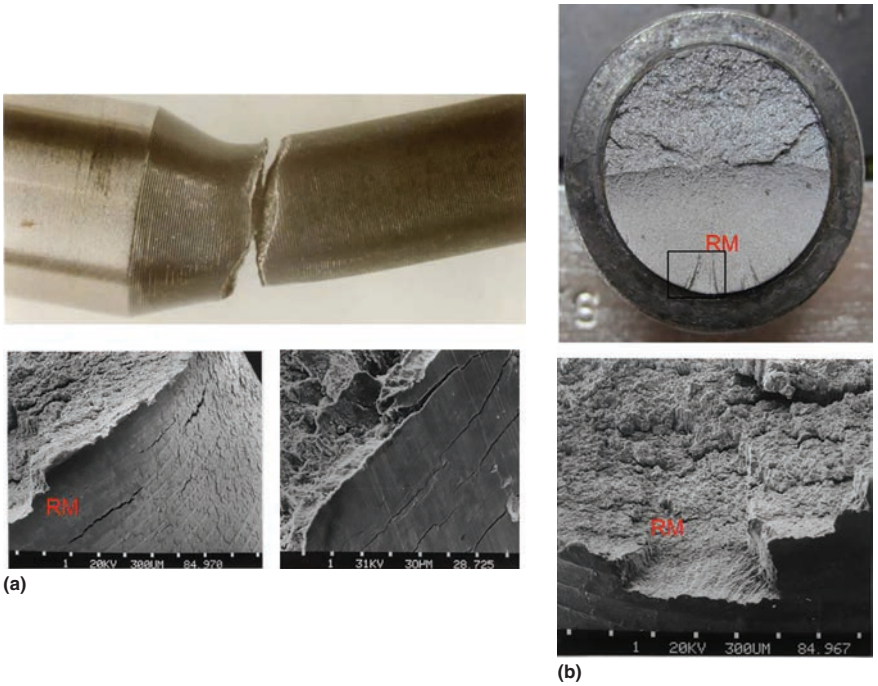


**Fig. 25** Examples of the formation of ratchet marks (RM) on fatigue fractures. (a) Screw shaft with rough surface finish. (b) Notch section of a fixing pin

the presence of multiple local high-stress concentration factors, such as machined grooves on the surface (Fig. 25), corrosion pits, or grinding burns.

### Striations

In the electron microscope examination of fatigue fracture surfaces, the most prominent features found are patches of finely spaced parallel marks, called fatigue striations. The fatigue striations are oriented perpendicular to the microscopic direction of crack propagation (Fig. 26a). In general, two types of striations can be recognized: ductile striations (Fig. 26b) and brittle striations (Fig. 26c). The more common ductile striations are due to the extended plastic zone being more pronounced and more easily detectable. The brittle striations seem to be a cleavage fracture along sharply defined facets. Numerous river markings separating these facets run normal to the striations, as seen in Fig. 26(c). Brittle striations in ductile aluminum alloys usually are an indication of a corrosive environment.

Striation spacing depends strongly on the level of applied stresses and stress-intensity factors. Under uniform loading, fatigue striations generally increase in spacing as they progress from the origin of fatigue. Each striation is the result of a single cycle of stress, but every stress cycle does not necessarily produce a striation.

In terms of fracture mechanics—under the presumption that one load cycle produces one striation—the striation spacing, $\Delta a$, is linked to the range of the stress-intensity factor, $\Delta K$. Applying the Paris crack-growth law, the striation distance for a semielliptical surface crack can be calculated from Eq 29, introducing $\Delta N = 1$ ($\Delta a$ in millimeters, $a$ in meters):

$$\Delta a = C \cdot \Delta K^m = C \cdot \Delta\sigma\sqrt{\pi \cdot a/Q} \cdot M)^m \qquad \text{(Eq 35)}$$

This equation is principally very useful in failure investigations, because it allows a correlation from the measured striation distance, $\Delta a$, to the acting stress range, $\Delta\sigma$. Care should be taken that the application of Eq 35 requires the determination of the spacing distance for a fracture plane, which is oriented as close as possible perpendicular to the electron beam.

A cyclic crack growth slightly above the threshold range with, for example, $\Delta K = 6.0$ MPa$\sqrt{\text{m}}$ (5.46 ksi$\sqrt{\text{in.}}$) results in steel structures according to Eq 27 and 35 with a striation distance of $\Delta a = 0.0077$ μm (0.3 μin.). Such extremely low crack-growth rates for the initiation stage of the fatigue crack will undershoot the resolution of commercial SEMs and, in practice, will not allow evaluation of the striations and measurement of the spacing distance. The investigation of striation spacing in the SEM consequently requires the evaluation of striations at regions of a larger crack size, $a$, which means at some distance from the origin.

The clarity of the striations depends on the type and the ductility of the material. Striations are more readily visible in ductile materials. Thus, patches of fatigue striations in body-centered cubic high-strength steel are less visible than in face-centered cubic aluminum alloys. Also, the striations are more prominent in stage II of the crack-growth curve (see the article "Fracture Appearance and Mechanisms of Deformation and Fracture" in this Volume).

At high rates of crack growth (e.g., 0.0025 mm, or $10^{-4}$ in., per cycle or more), the striations become wavy and develop a rough front. A large plastic zone exists in front of the crack, which may cause extensive secondary cracking. Each secondary crack propagates as a fatigue crack, creating a network of secondary striations. The local crack direction may differ markedly from the overall direction of crack propagation because of the changes in direction of the local fracture path.

The striations formed in aluminum alloys at very low crack-growth rates (less than $1.3 \times 10^{-4}$ mm, or $5 \times 10^{-6}$ in., per cycle) are difficult to resolve and often cannot be distinguished from the network of slip lines and slip bands associated with plastic deformation at and near the crack front as it propagates through the section. Care should be taken in identifying a fatigue fracture only on the basis of striations, because other sliplike lines that are not generated by cyclic loading, such as pearlite, graphite flakes, rubbing marks, or stable stepwise crack extension under quasi-static loading, may exist on the fracture.

## *Final Fracture (Stage III)*

Final fracture occurs when the crack has grown to the critical size for fracture. The size of the final fracture zone depends on the magnitude of the acting stresses as well as the loading type and the load-time cycle ($R$) (Fig. 16, Eq 31). The final fracture zone of a fatigue fracture surface is often fibrous, resembling the fracture surfaces of specimens in Charpy-impact or fracture toughness tests of the same material.

Two features of the final fracture zone aid in determining the origin of fracture. First, fatigue usually originates at the surface; therefore, the fatigue origin is not included in the shear lip zones of the final fracture region. In tough materials with heavy sections, the final fracture zone will consist of a fracture by two distinct modes: normal-stress-induced fracture (plane-strain mode) extending from the fatigue zone in the same plane, and shear-stress-induced fracture (plane-stress mode) at 45° to

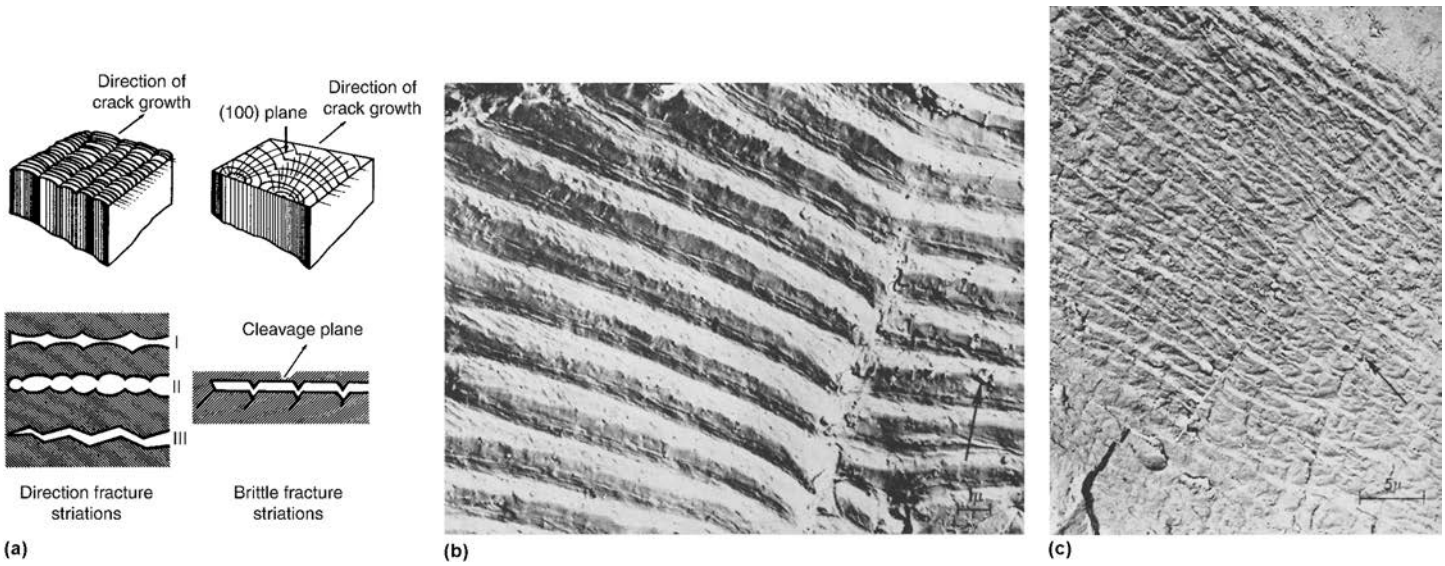


**Fig. 26** Ductile and brittle striations. (a) Schematics of different types of ductile and brittle striations. (b) Ductile striations in 718 aluminum alloy. (c) Brittle fatigue striations of 2014 aluminum alloy. Note cleavage facets running parallel to direction of crack propagation and normal to striations. Source: Ref 22

the surface of the part bordering the normal-stress fracture. In thin sheet structures with plane-stress conditions (uniaxial or biaxial stress state) having sufficient toughness, final fracture occurs somewhat differently. As the crack propagates from the fatigue zone, the fracture plane rotates around an axis in the direction of crack propagation until it forms an angle of approximately 45° to the principal stress direction and the surface of the sheet. The fracture can occur on either a single-shear or a double-shear plane (Fig. 27).

The second characteristic of a fast-propagating brittle fracture is chevron marks that point back to the origin of fracture. This is shown in Fig. 28 on a torsional fatigue fracture of an induction-hardened steel shaft. Beach marks fan out from the origin at a sharp corner in a keyway at the top. Beyond the beach marks, cracks progressed by fast fracture along the hardened perimeter, producing two sets of chevron marks that point back to the crack origin.

In an SEM investigation, the characteristics of final fracture are either dimples for the ductile fracture or cleavage for the brittle fracture. For semiductile behavior, a mix of both fracture types may be present.

## Effect of Loading and Stress Distribution

Initiation and growth of a fatigue crack and the features on the fracture surface are all strongly affected by the shape of the part, the type and magnitude of loading exerted on the part in service, residual stresses, as well as metallurgical and environmental factors. Fatigue cracks usually initiate at some region of stress concentration resulting from the presence of surface discontinuities (stress raisers), such as a step or shoulder, a screw thread, an oil or bolt hole, or a surface flaw. In addition, mechanically fastened joints and welded joints also require special attention concerning notch effects.

The magnitude of the nominal stress on a cyclic-loaded component is often expressed by the nominal stress in the cross section. Increasing the magnitude of the nominal stress means that the striation spacing is increased and the region of final fast fracture is increased. With very high-stress/strain amplitudes, low-cycle fatigue fractures are produced. The arbitrary but commonly accepted dividing line between high- and low-cycle fatigue is considered to be approximately $10^2$ to $10^4$ cycles.

A more basic distinction is made by determining whether the dominant component of the strain amplitude imposed during cyclic loading is elastic (high cycle) or plastic (low cycle), which in turn depends on the properties of the metal as well as the magnitude of the local stress. Often the transition from high- to low-cycle fatigue is defined as the intersection between the elastic and plastic line of the low-cycle fatigue curve (Fig. 9).

Design methods should account for the irregular nature of actual load applications, which requires the transfer of the load-time cycles into a load spectra, applying counting methods and the application of damage-accumulation concepts. For satisfactory correlation with the service behavior, full-sized or large-scale specimens should be tested under conditions as close as possible to those existing in service. This method is expensive, but it does provide valuable information. A less costly testing procedure is simplified laboratory testing of samples. By using the fatigue information obtained from the testing of standard specimens or by using models and applying the proper constitutive equations for configuration, surface finish, environment, and various other parameters, an approximation of the lifetime of the component can be estimated.

It should be emphasized, however, that this is just an approximation. For improving production designs, the target load/life test method of evaluating the design of a part has been used by automotive, farm equipment, construction equipment, offshore structure, and aircraft companies. In this method, numerous failures from actual operating machines are observed. From these observations, conclusions can be drawn for the possible loads that are causing a failure. A fatigue test is then set up to apply the appropriate loading. After obtaining failures during the test, certain adjustments can be made in the magnitude and method of applying the loads until the same type of failure is developed in the test as those experienced by the operating machines.

Target lives for the parts in the fatigue test can be established, based on the time-to-failure on the machines in service, the time-to-failure in the fatigue tests, and the desired life in service. One application in which this method has been used is the rear axles of automobiles. Maximum tractive torque is applied to the rear axle for 100,000 cycles. If the axle gears withstand this loading, they are considered satisfactory for customer use. The advantages of this procedure are that it makes use of the experience gained from many machines already in use and that it is quick. Its disadvantage is that sufficient machines must be placed in customers' hands to develop the failed parts for study and evaluation of test methods.

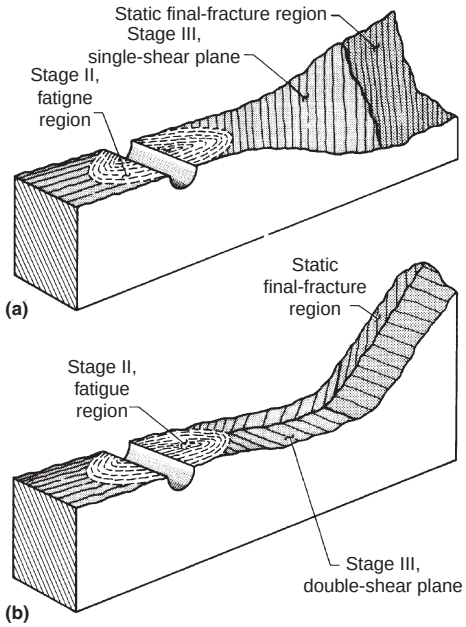


**Fig. 27** Fracture planes that are 45° to the direction of loading. (a) Single-shear plane. (b) Double-shear plane

**Fig. 28** Surface of a torsional fatigue fracture in an induction-hardened 1041 (1541) steel shaft, which fractured after 450 h of endurance testing

Moreover, the differences in the operating conditions, environment, fatigue properties of the parts, and processing methods between the production design and new design must be small to produce accurate results. Therefore, this technique is primarily used when there is ample experience, as in improving parts that have a history of failure in service or in evaluating machines that are similar to existing production models.

## *Load Conditions*

Stress analysis is relatively easy in a beam of simple uniform cross section subjected to axial, bending, or torsional loads. In the case of pure axial loading of a smooth bar, the stress is constant across the cross section of the beam, and a fatigue crack may initiate at a discontinuity within the member at or beneath the surface. However, due to superimposed bending stresses caused, for example, by misalignment of the structure or eccentric loading, pure membrane stresses are rarely found in service, and the appearance of fatigue cracks caused by cyclic axial loads is often similar to that described for bending loads. In bending and torsion, stresses are greatest at the surface, where fatigue cracks are thus more likely to initiate, as described in more detail for unidirectional, alternating, and rotational bending.

Loading of short components with rectangular section differs from that of long, cylindrical components. In flat plates, biaxial tension is more common and torsion is less so. In sheet or plate materials with heavy sections, the fatigue crack front may extend under plane-strain conditions to give a wholly flat-face fracture, as shown in Fig. 29(a). Fatigue fractures in very thin sheet materials subjected to high stress intensities may shift from flat face (plane-strain condition) to shear face (plane-stress condition), as shown in Fig. 29(b).

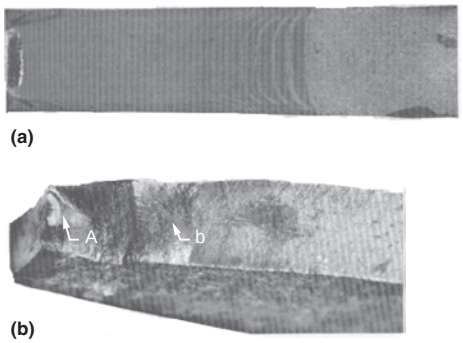


**Fig. 29** Fatigue fracture zones in aluminum alloy 7075-T6 plates. (a) Fatigue crack that grew as a flat-face fracture with a shallow convex crack front. (b) Change in orientation of fatigue fracture from plane strain (arrow A) to plane stress (arrow B)

### Unidirectional Bending

A beam subjected to cyclic unidirectional bending may be subjected to a bending moment that is uniform along the length of the beam; the tensile fiber stress in a smooth bar is also uniform along the length of the beam. Therefore, a fatigue crack may be initiated at any point along the beam; in fact, several fatigue cracks may have been active before one of them became large enough to cause the final fracture.

Another form of loading is cantilever loading, in which the bending moment, and therefore the tensile fiber stress, varies linearly along the length of the beam, and shear stress exists due to the transverse force. Fracture will initiate at the point of highest positive bending stress, adjacent to the rigid mount. Because of the superimposed shear stress, the direction of the maximum principal stress has a small angle with the axis of the beam (Eq 35). Consequently, a fatigue crack will usually propagate inclined at the fixed mounting.

Figure 30 illustrates fatigue beach marks produced from single origins at low and high nominal stresses. The cracks start from multiple origins in a smooth cylindrical bar or shaft subjected to a cyclic unidirectional bending moment evenly distributed along the length. The cracks initiate at minor stress raisers, such as surface discontinuities. Both the single origin and the smallness of the final fracture zone in Fig. 30(a) reflect that the nominal stress was low. The crack front, which forms beach marks, is symmetrical relative to the origin and retains a concave form throughout. The larger final fracture zone in Fig. 30(b) reflects a higher nominal stress. Figure 30(c) shows a typical fatigue crack originating as several individual cracks that ultimately merged to form a single crack front. Radial steps (ratchet marks) are present between the crack origins. Figures 30(d–f) show typical fatigue beach marks that result when a change in section in a uniformly loaded shaft provides a moderate stress concentration. With a low nominal stress, the crack front changes from concave to convex before rupture (Fig. 30d). At higher nominal stresses, the crack front flattens and may not become convex before final fracture (Fig. 30e, f).

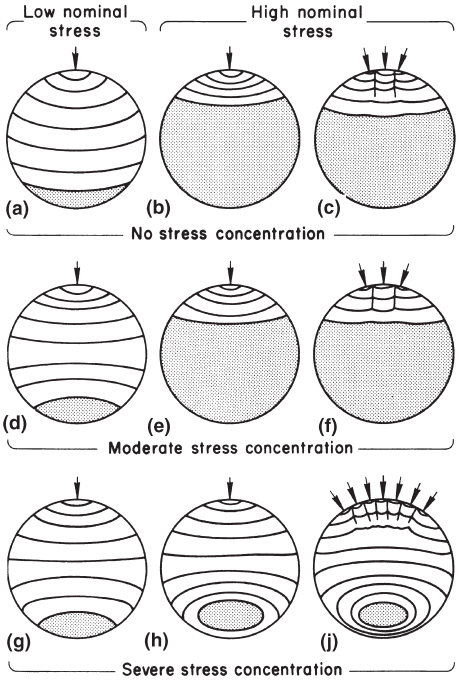


**Fig. 30** Fatigue marks produced from single origins at low and high nominal stresses and from multiple origins at high nominal stresses. Fatigue marks are typical for a uniformly loaded shaft subjected to unidirectional bending. Arrows indicate crack origins; final fracture zones are shaded.

A change in section in a uniformly loaded shaft that produces a severe stress concentration will lead to a pattern of beach marks such as that shown in Fig. 30(g, h, or j). An example of a severe stress concentration is a small-radius fillet at the junction of a shoulder and a smaller-diameter portion of a shaft or at the bottom of a keyway. Such a fillet usually results in the contour of the fracture surface being convex with respect to the smaller-section side. The crack front pattern shown in Fig. 30(g) was produced by a low nominal stress. The crack front in Fig. 30(h) developed more rapidly because of a higher stress in the peripheral notched zone. Multiple crack origins, high nominal stress, and unidirectional bending usually produce the beach mark pattern shown in Fig. 30(j).

Figure 31 is an example of a shaft that failed from unidirectional bending fatigue. The shaft, made from a free-machining grade of A6 tool steel, was part of a clamping device on a tooling assembly used for tube bending. The shafts were subjected to a membrane stress imposed by the clamping force and a bending stress resulting from the nature of the operation.

**Fig. 31** Steel shaft from a tube-bending machine that failed by fatigue fracture. The fracture surface shows regions of fatigue crack propagation and final fracture.

Unidirectional bending stresses were imposed on the shafts when a right-hand bend was made in the tubing as well as when a left-hand bend was made. The tensile stress on the shafts was also cyclic, because the clamping force was released after each bend.

Visual examination of the fracture surface revealed both a smooth area and a coarse, granular area (Fig. 31). The dull, smooth area is typical of some fatigue fractures and resulted as the crack was opened and closed by the bending stress. Beach marks on the smooth area of the fracture surface also indicate fatigue fracture. The coarse, bright, and crystalline-appearing area is the final fracture zone. The smooth-textured fatigue zone is relatively large compared with the crystalline-textured final fracture zone, which indicates that the shaft was subjected to relatively low nominal stresses. The final fracture surface at the bottom shows that a one-sided cyclic bending load was involved. The fatigue crack was initiated in a 0.25 mm (0.010 in.) radius fillet at a change in section. Cracking was nucleated at a nonmetallic inclusion that intersected the surface at a critical location in the fillet ("notch in the notch"). The small radius of the fillet at the change in section resulted in a stress concentration that, in conjunction with the oxide/sulfide inclusion that intersected the surface of the fillet, initiated a crack. New shafts were made with a 2.3 mm (0.09 in.) radius fillet at the critical change in section.

Another example of a bending-loaded shaft is shown in Fig. 32. The loading was mainly by unidirectional (pulsating) bending fatigue, causing the main fatigue crack initiation on the tension-stressed right side. However, secondary origins (C and D in the figure) on the compression side suggest that a small reversed bending load or a backlash may have been present. Another reason for crack initiation on the compressive side may be a compressive plastic deformation at the notch, with the consequence of tension residual stresses after unloading (see the section "Residual Stresses" in this article).

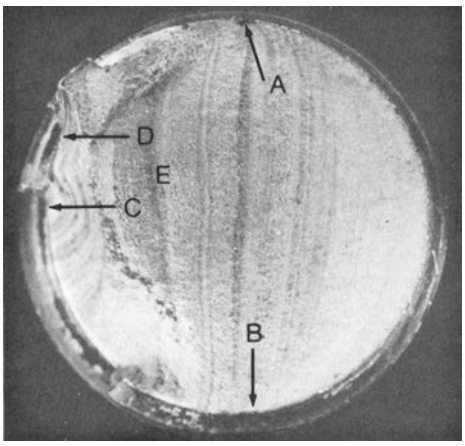


**Fig. 32** Fatigue fracture of a steel bolt. Interpretation of the surface indicates that loading was primarily by unidirectional bending. However, secondary origins (C and D) indicate the possibility that a small reversed bending or backlash may have been present. Many closely spaced origins along the thread root (right side, between A and B) occurred from this region of high stress concentration. Nominal stress was low, resulting in a small region of fast fracture (at E).

### Alternating (Reversed) Bending

When the applied bending moment is reversing, the part is subjected alternately to tension stress and compression stress. The points on one side of the plane of bending are in tension, and the points on the opposite side are in compression. If the bending moment is of the same magnitude in either direction, two cracks of approximately equal length usually develop from origins diametrically opposite each other and often in the same transverse plane. If the bending moment is greater in one direction than in the other, the two cracks will differ in length. The cracks on opposite sides of the beam or shaft are not necessarily in the same plane; therefore, final fracture in the center of the cross section may occur at some angle other than 90° to the axis of the beam. The size of the final fracture zone is indicative of the relative nominal stress level of the beam.

There is one significant difference between the appearance of fatigue cracks formed under pulsating load and those formed under alternating load. With an alternating load, each crack is opened in one half-cycle and closed and pressed together during the other half-cycle, which rubs and polishes the elevated points on opposite sides of the fatigue crack. Under unidirectional load, the rubbing—generated by compressive residual stresses—is less pronounced. Rubbing may sometimes be sufficient to obliterate many of the characteristic marks, and the crack surfaces may become dull or polished. Consequently in failure investigations, beach marks and striations are usually more readily observed in pulsating loading than in alternating loading.

Figure 33 shows schematically the typical formation of the fatigue crack front on the fracture surface of a stationary (nonrotating) shaft subjected to a reversed bending moment. The crack origins (arrows) are shown diametrically opposite each other, but sometimes they are slightly displaced by minor stress raisers. The pattern shown in Fig. 33(a) is typical of that for a single-diameter shaft with minor stress concentration. A large-radius fillet at a change in shaft diameter imposes a moderate stress concentration. Figure 33(b) shows the pattern on the surface of a fracture through such a fillet. A small-radius fillet at a change in shaft diameter results in a severe stress concentration. Figure 33(c) shows the typical pattern on the surface of a fracture through a small-radius fillet. The reason for the fatigue pattern changes is that the fatigue crack propagates faster in more severe stress concentrations at the surface than in the interior.

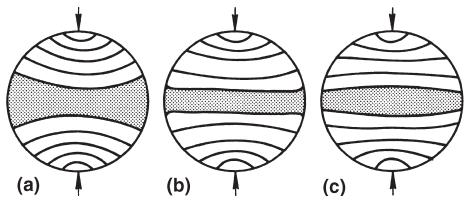


**Fig. 33** Typical fatigue marks on the fracture surface of a uniformly loaded nonrotating shaft subjected to reversed bending stresses. (a) No stress concentration. (b) Moderate stress concentration. (c) Severe stress concentration. Arrows indicate crack origins; shaded areas are final fracture zones.

Another fracture surface from reversed bending is shown in Fig. 34. Stress concentration was mild, indicated by several origins on both sides and the straight crack front (Fig. 33). Nominal stress was relatively high, indicated by the large region associated with fast fracture.

### Rotational Bending

A machine component that is commonly subjected to bending stresses is a rotating round shaft. A unique feature of rotational bending loading is that during one revolution of the shaft, both maximum and minimum stresses are exerted around the entire circumference of the shaft in the region of maximum bending stresses. Because the loading is axially symmetrical, a fatigue crack can be initiated at any point, or at several points, around the periphery of the shaft. Multiple cracks do not necessarily occur in the same plane and are separated by radial ridges (ratchet marks). The presence of multiple cracks is typical for rotational bending.

The essential difference between a stationary shaft and a rotating shaft subjected to the same bending moment is that in a stationary shaft the maximum tensile stress is confined

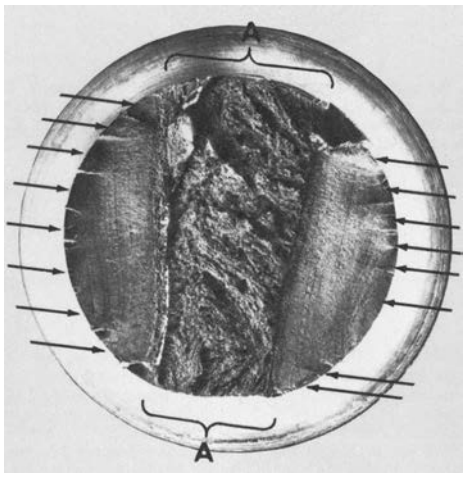


**Fig. 34** Fatigue fracture of steel shaft (50 mm, or 2 in., diameter) from reversed bending, indicated by fracture-initiation sites (arrows) on both sides. The multiple initiation sites indicate a mild amount of stress concentration. Fast fracture region is in area between brackets A.

to a portion of the periphery only. In a rotating shaft, every point on the periphery sustains a tensile stress once every revolution. Therefore, one difference arising from rotation is the distribution of multiple initiation sites at points on the periphery (Fig. 35b, d), depending on the relative magnitude of the stresses at different locations determined by conditions of balance or imbalance imposed on the shaft. Under low or moderate nominal stresses, a rotating shaft may fail as a result of a single fatigue crack.

Another important difference introduced by rotation is asymmetrical development of the crack front from a single origin. There is a marked tendency of the crack front to extend preferentially in a direction opposite to that of rotation, typically approximately 15° or more to the load line (Fig. 35a, c). When the shaft is always rotated in the same direction, the crack usually advances asymmetrically, with the beach marks shifted in a direction opposite to that of shaft rotation. Asymmetric fatigue crack growth indicates rotational bending and the direction of rotation. Under periodically reversing rotating-bending loads, fatigue cracks grow symmetrically. Even without reversing the direction of rotation, a rotating-bending fatigue crack may advance symmetrically.

A common result of final fracture in rotating bending is that slight movement of one side of the crack relative to the other side frequently causes severe damage to the fracture surfaces and tends to obliterate many characteristic marks. However, although the high spots on one surface may rub the high spots on the other, the marks in the depressions are retained and can be used for examination, for example, by SEM. Because the depressions are negative images of the damaged high spots on the opposing surface, they provide useful evidence. Therefore, it is desirable to examine both parts of a cracked or fractured part.

The similarity in macroscopic appearance of fractures in shafts resulting from rotating-bending fatigue and from monotonic overload torsional shear of a relatively ductile metal possibly results in misinterpretation. The fracture surface shown in Fig. 36(a) was the result of fatigue, as evidenced by the ratchet marks around the periphery and the pronounced beach marks. Under the low magnification shown in Fig. 36(b), beach marks are not visible because of constant-amplitude loading or because they were obliterated by rubbing. The presence of ratchet marks around the periphery is also an indication of rotating-bending fatigue. However, the metal smearing apparent on the quasi-static fracture surface and the twisting deformation of the shaft shown in Fig. 36(c) indicate monotonic torsional shear and would preclude mistaking this fracture for a fatigue fracture. The fracture shown in Fig. 36(d) also exhibits a superficial similarity to a fatigue fracture. However, it is evident that this fracture was the result of monotonic torsional shear, because the entire fracture surface has a smooth texture and no well-defined final fracture area.

In splined shafts, fracture resulting from torsional shear is frequently accompanied by deformation of the splines not engaged by the mating part. However, the portion of the shaft not engaged by the mating part is sometimes unavailable for examination. When macroscopic examination affords only inconclusive evidence, use of an SEM may reveal either fatigue striations or shear dimples. Also, metallographic examination of a section through the fracture surface may reveal microdeformation from torsional shear in the rotary direction that would not be present in a fatigue fracture.

### Example 2: Failure of an Overhead Crane Drive Shaft due to Rotating-Bending Fatigue.

In this example (submitted by J.R. Emmons, Rocketdyne Division, Rockwell International, Canoga Park, CA), a drive shaft used as part of the drive train for an overhead facilities

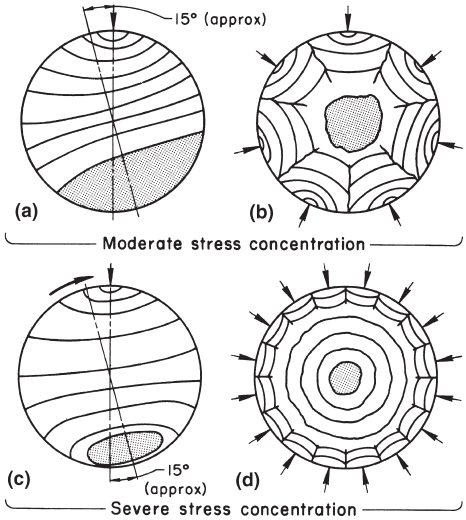


**Fig. 35** Typical fatigue marks on the fracture surface of a uniformly loaded rotating shaft. Marks are produced from single and multiple origins (arrows) having moderate and severe stress concentration; shaded areas are final fracture zones. Shaft rotation is clockwise

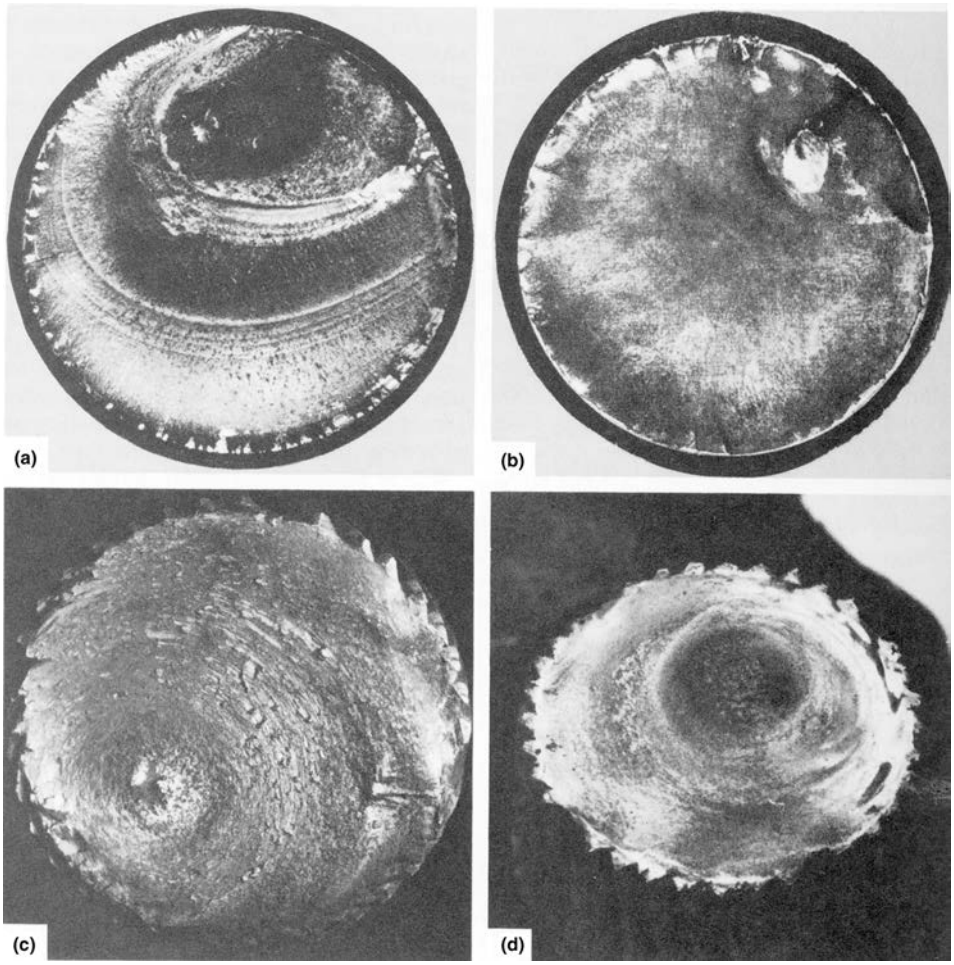


**Fig. 36** Fracture surfaces of failed shafts. (a) and (b) Failure by fatigue. (c) and (d) Failure by torsional shear. See text for discussion.

crane fractured due to low-cycle rotating-bending fatigue, initiating and propagating by combined torsional and reversed bending stresses. The part submitted for examination was a principal drive shaft that fractured in the location of a 90° fillet, where a wheel hub is machined down to 34.9 mm (1.375 in.) diameter from a 60.3 mm (2.375 in.) diameter main shaft. A 9.5 mm (0.375 in.) wide by 3.2 mm (0.125 in.) deep keyway was machined into the entire length of the hub, ending approximately 1.6 mm (0.062 in.) from the 90° fillet (Fig. 37). The fractured hub had separated from the main shaft and lodged in the shaft hole of the wheel when the wheel fell away from the shaft. The wheel, together with the fractured hub and the shaft, were submitted for analysis.

*Investigation*

The preservice and postfailure radiographic and penetrant inspection records for the shaft were reviewed and found to be satisfactory. No additional cracks or indications were evidenced. The crack that led to the failure was not present when the shaft was installed. It was estimated that installation had been 1 to 2 years prior to the failure. The shaft material was verified as tempered 4150 steel by using wet chemical analysis and metallographic techniques. The hardness and the microstructure were surveyed by preparing metallographic specimens from the fractured hub section. The microstructure was confirmed to be a tempered martensite structure. The surface of the crack was examined at low and high magnifications using optical and scanning electron microscopy.

Two additional shafts had not fractured to failure but were radiographically and die penetrant inspected as part of a verification analysis to determine whether the remaining drive shafts were serviceable. One of these shafts was found to have a radial penetrant indication, this time in the chamfered fillet in the 25.4 mm (1 in.) diameter machined motor drive at the end of the shaft, which is the end opposite to that which fractured in the principal shaft (Fig. 37). This shaft was sectioned diametrically to expose a polished cross section of the indication. One quarter of the radial indication was broken open to expose the crack fracture features for observation to determine the mode of fracture. Because of the resultant similarity between this crack and the fracture in the principal shaft, no further forensic work, other than basic fractography, was performed on this second shaft.

Figure 38 shows a magnified view of the fractured hub that caused the wheel to break off the principal drive shaft of the overhead crane. The concentric circles revolving around the final overload fracture, along with the ratchet marks around the outer circumference pointing toward the final fracture, are typical of crack growth induced by rotating-bending stresses. The ratchet marks expose the fatigue-initiation sites around the outer perimeter, as verified by high-magnification SEM performed later in the investigation. The individual cracks propagated toward a single crack front and then toward the final overload area. Figure 38 shows the darker final overload region slightly off-center of the hub. The final overload was probably off-center due to the rotating-bending stresses that initiated the fracture on one side of the hub (most likely the side opposite the final fracture), while at the same time driving the crack around the outer circumference, producing multiple inward-growing crack fronts.

It was originally suspected that the overall fracture had initiated at the keyway, which is usually the case in shafts where keyways are present, but the forensic analysis provided the contributing evidence that the fatigue cracks actually propagated toward the keyway rather than initiating from it.

*Conclusions*

A drive shaft used to assist in driving an overhead crane for the movement of medium- to heavy-weight components inside a manufacturing building fractured at the mating interface with the drive wheel, in the location of a reduced-diameter wheel hub machined at a sharp 90° fillet from the main bulk shaft. The sharp 90° fillet coexisted with a keyway in the vicinity of the fracture, although the keyway did not appear responsible for the failure. A second shaft was cracked in a sharply chamfered region where the main shaft was machined to a reduced diameter to form the motor drive hub opposite the wheel hub at the end of the shaft. The fracture mode for both shafts was rotating-bending fatigue initiated and propagated by combined torsional and reverse bending stresses. All drive shafts

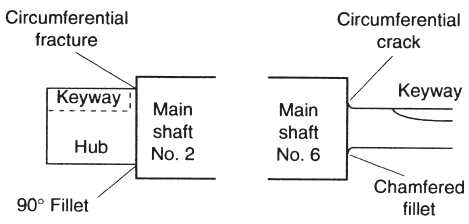


**Fig. 37** Locations of observed fractures in a shaft assembly

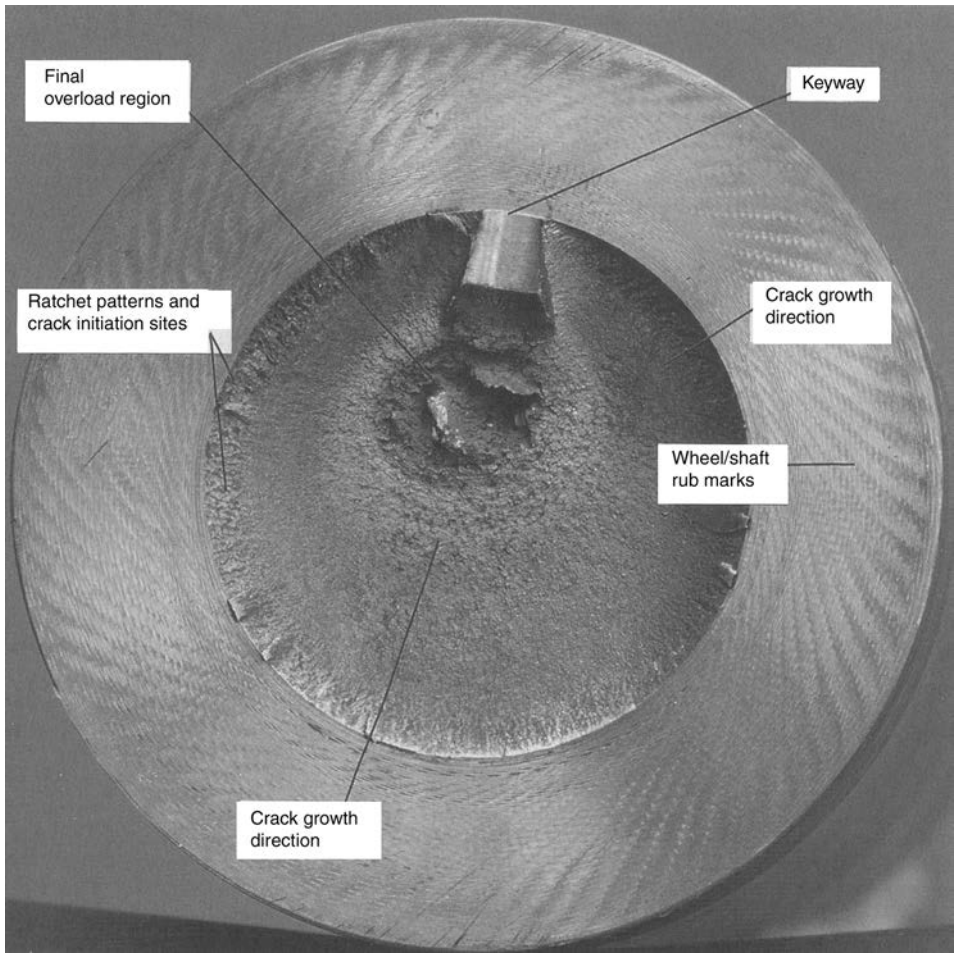


**Fig. 38** Fracture section of steel shaft. Courtesy of J.R. Emmons, Rocketdyne Division, Rockwell International

were replaced with new designs that eliminated the sharp 90° chamfers in favor of a more liberal chamfer, which reduced the stress concentration in these areas.

### Torsional Loading

Under pure torsional loading of a shaft, the maximum principal stress is oriented at 45° to the axis of the shaft. Under alternating (reversed) torsional loading ($T_m = 0$, $R = -1$) in a normal stress-governed failure, two sets of fatigue cracks under ±45°, perpendicular to each other, may develop. For example, the presence of perpendicular fatigue cracks in a component subjected to fluctuating torsion probably indicates the presence of alternating torsional loading, for example, due to vibration.

Shear-stress-induced torsional fatigue cracks, for example, due to relative high stresses in the low-cycle fatigue regime, are in the longitudinal (Fig. 39a) or the transverse (Fig. 39b) direction, where the maximum shear stresses of the same magnitude are present. Because of the directionality of structures with a weaker longitudinal rolling or forging direction, longitudinal shear cracks are more common.

At later stages of fatigue crack growth, one crack of the pair usually grows much faster than the other and eventually causes rupture of the shaft.

If two cracks develop mutually at right angles, it can be presumed that the torque was of a reversing character. The relative extent of development of two torsional fatigue cracks mutually at right angles can indicate the magnitude of the torque reversals that have been applied. If the cracks are of approximately the same length, the indications are that the torque reversals have been of equal magnitude, but only if the cracks are in a comparatively early stage of development. Beyond this stage, one crack usually takes the lead, and such inferences are no longer justified. Fatigue cracks arising from torsional stresses also show beach marks and ridges.

Under a fluctuating torsional load ($T_m \geq T_a$, $R \geq 0$), the fatigue cracks develop normal to the maximum principal stress and typically show a fatigue crack in only one direction (45°).

If an additional bending stress is applied to a shaft that is transmitting torque, the angle at which the fatigue crack develops is modified. The direction of the maximum principal stress, $\sigma_1$, to the longitudinal axis $x$ of a shaft under combined bending stress, $S_b$, and torsion stress, $T_t$, is:

$$\varphi_{x,1} = 0.5 \cdot \text{arc tan}(2 \cdot T_t/S_b) \quad \text{(Eq 36)}$$

In pure torsion loading ($S_b = 0$), the fracture surface is oriented at 45° relative to the shaft axis, as shown in Fig. 40 from a test specimen in a torsional fatigue test. If the angle differs significantly from 45° to the shaft axis (between 45° and 0°), the presence of a bending stress is indicated. Pure bending ($T_t = 0$) produces a normal-stress-controlled fatigue fracture in the transverse direction of the shaft.

Longitudinal stress raisers are comparatively harmless under bending stresses but are important under torsional loading. This sensitivity of shafts loaded in torsion to longitudinal stress raisers is of considerable practical importance, because keyways, machining grooves, and inclusions in the shaft material are oriented in the longitudinal direction of the shaft axis. It is not unusual for a torsional fatigue crack to originate at a keyway corner or spline, a surface mark, or a longitudinal inclusion and to branch outward at approximately 45° (Fig. 39).

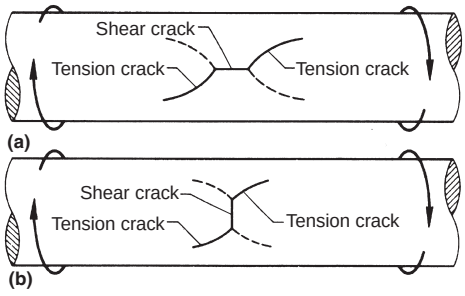


**Fig. 39** Schematic of the initiation of torsional fatigue cracks in a shaft subjected to (a) longitudinal shear and (b) transverse shear. Dashed lines indicate other cracks that can appear when torsional stresses are reversed.

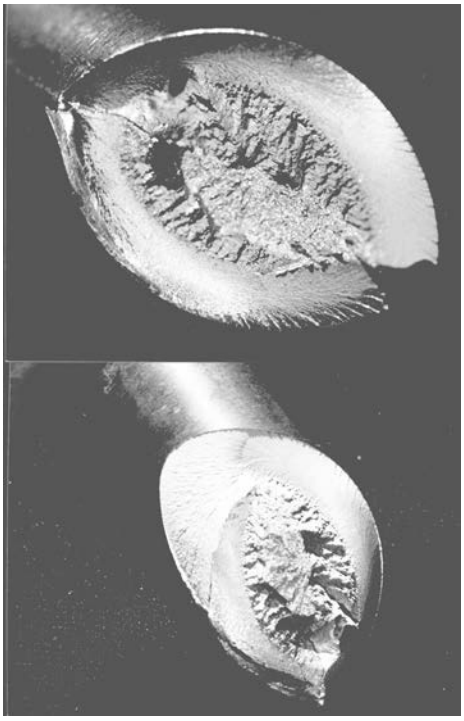

**Fig. 40** Fracture surfaces of a torsional fatigue test specimen. Courtesy of G. Fett, Dana Corporation

### Example 3: Torsional Fatigue Fracture of a Large 4340 Steel Shaft Subjected to Cyclic Loading and Frequent Overloads.

The 4340 steel shaft shown in Fig. 41(a), the driving member of a large rotor subject to cyclic loading with frequent overloads, broke after three weeks of operation. The shaft was also part of a gear train that reduced the rotational speed of the driven member. The driving shaft contained a predetermined breaking section that was designed as a shear groove at which the shaft should break if a sudden high overload occurred, thus preventing damage to an expensive gear mechanism. The rotor was subjected to severe chatter, which was an abnormal condition resulting from a series of continuous small overloads at a frequency of approximately 3 Hz.

*Investigation*

Examination disclosed that the shaft was broken at the shear groove and that the fracture surface contained a star-shaped pattern (Fig. 41b). Figure 41(c) shows the fracture surface with the pieces fitted back in place. The pieces were all nearly the same size and shape, with indications of fatigue, cleavage, and shear failure in approximately the same location on each piece. The cracks were oriented at approximately ±45° to the axis of the shaft, which indicated that final fracture was caused by the principal stress normal to the 45° plane and not by the longitudinal or transverse shear stress that had been expected to cause a ductile overload failure.

Examination of the surfaces of one of the pieces of the broken shaft revealed small longitudinal and transverse shear cracks at the smallest diameter of the shear groove. Also, slight plastic flow had occurred in the metal adjacent to these cracks. Cracking occurred at many points in the groove in the shaft before several of the cracks grew to a critical size.

No surface irregularities were present in the shear groove at any of the shear cracks. The microstructure of the steel was normal, with a uniform Rockwell hardness of 30 to 30.5 HRC (equivalent to $S_u$ = 940 MPa, or 136 ksi) across the section, indicating a tensile strength in the expected range for quenched and tempered 4340 steel shafts. A hot acid etch showed the steel to be free from segregations or other irregularities.

*Conclusions and Corrective Measures*

The basic failure mechanism of the shaft was fracture by torsional fatigue, which started at numerous surface shear cracks, both longitudinal and transverse, that developed in the periphery of the root of the shear groove. These shear cracks resulted from high peak loads caused by chatter. Stress concentrations developed in the regions of maximum shear, and fatigue cracks propagated in a direction perpendicular to the alternating maximum principal stress, thus forming the star pattern at ±45° to the longitudinal axis of the shaft.

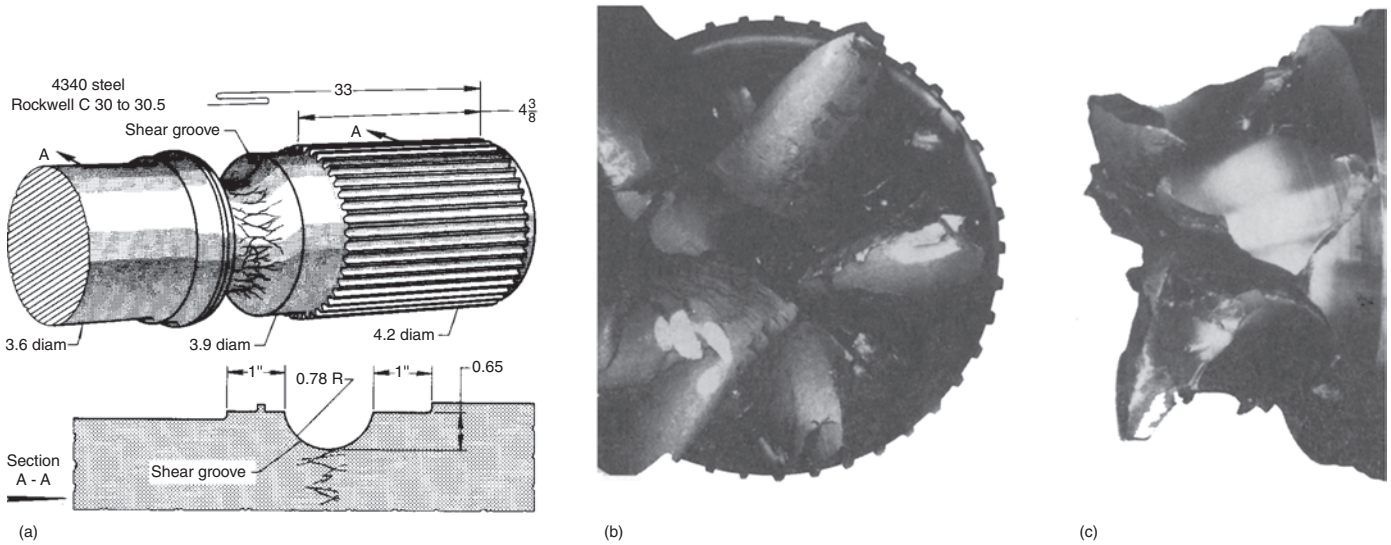


**Fig. 41** 4340 steel rotor shaft that failed by torsional fatigue. (a) Shear groove designed to protect gear mechanism from sudden overload (dimensions are in inches). (b) Star-shaped pattern on a fracture surface of the shaft. (c) Longitudinal and transverse shear cracks on the surface of the shear groove, which resulted from high peak loads caused by chatter

The shear groove in the shaft, designed to prevent damage to the gear train, had performed its function, but obviously at a lower overload level than intended.

The fatigue strength of the shaft was increased by shot peening the surface of the shear groove, and chatter in the machine was minimized.

### *Stress Concentrations*

Notches, grooves, holes, fillets, threads, keyways, and splines are common design features that act as stress raisers in a component. To visualize the distribution of stress at a change in section size or shape, it is helpful to consider the part in terms of water or electricity flowing through a chamber or a conductor of similar cross section. Stress flow can be represented as a series of parallel lines, where stress is inversely proportional to the distance between the lines for several types of stress raisers typically found in parts (Fig. 42). A single notch (Fig. 42j) introduces a greater stress-concentration effect than does a continuous thread (Fig. 42k). This is because of the softer flow due to adjacent threads. The stress concentration at the right of the arrow in Fig. 42(k) is very similar to that in the narrow collar in Fig. 42(h). To the left of the arrow, however, the first thread is relieved from one side only; consequently, there is more stress concentration, similar to that of the single notch in Fig. 42(j). This—and the fact that the first thread carries the highest load due to the sudden stiffness change—is the reason why bolts fracture in fatigue at the first load-carrying thread.

All such sectional discontinuities increase the local stress level above the nominal stress in the minimum cross-sectional area. In addition to magnification of the stress and the reduction of fatigue strength or fatigue life, increasing the severity of stress concentration has effects on fatigue crack features including:

- Beach marks usually become convex toward the point of crack origin.
- Under rotational loading, the beach marks may completely surround the final fracture zone.
- Multiaxial stress states may be introduced, thereby influencing the direction of crack growth.

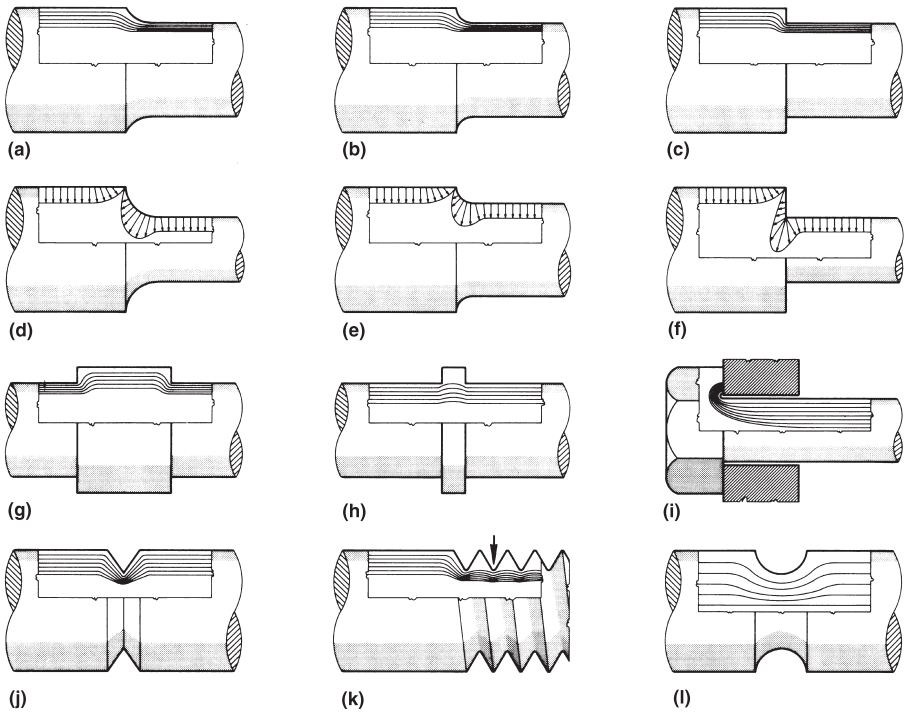


**Fig. 42** Effect of stress raisers on stress concentration and distribution of stress at several changes of form in components. (a) to (c) Progressive increases in stress with decreasing fillet radii. (d) to (f) Relative magnitude and distribution of stress resulting from uniform loading. (g) Stress caused by the presence of an integral collar of considerable width. (h) Decrease in stress concentration that accompanies a decrease in collar width. (i) Stress flow at the junction of a bolt head and a shank. (j) Effect of a single sharp notch. (k) Effect of a continuous thread. (l) Effect of a groove or gage

In the absence of stress concentrations at the surface, cracks propagate more rapidly near the center of a section than at the surface. The deformation constraints ahead of the crack tip cause the stresses to be triaxial and more severe away from the surface. However, when there is a stress-concentrating notch at the surface, such as a thread with a sharp root, the stress near this notch is more severe than it is farther below the surface. Under conditions of severe notching, W-shaped crack fronts are sometimes observed. In regions exposed to high stresses from loading or in the presence of a severe stress concentration, multiple crack origins will be seen. In most cases, the cracks from these origins will unite to form a single crack front. Before the single crack front is formed, the individual microcracks will be separated by small, vertical ledges called ratchet marks.

### Stress-Concentration Factor

Stresses are concentrated in components due to structural discontinuities, such as fillets in shafts, holes, scratches, and volumetric or planar defects, which act as stress raisers. Another reason for stress magnifications above the nominal stress are geometric imperfections (e.g., misalignments in welds) or stiffness changes (e.g., reinforcement) that generate local secondary bending stresses, which also increase the primary nominal stress.

The maximum stress in the root of the notch must be evaluated in general with the theory of elasticity, applying numerical methods such as finite-element or boundary-element methods. The stress can be a coordinate stress or an equivalent stress (maximum principal stress, maximum shear stress, von Mises stress). Alternatively, the stresses can be determined by experimental stress analysis with (small) strain gages or photoelastic methods.

For typical standard geometries, such as a plate with holes or fillets of shafts where a nominal stress can be defined, the maximum stress can be expressed with the stress-concentration factor, which again can be defined in coordinate stresses or equivalent stresses. The stress-concentration factor, $K_t$, is the ratio of the maximum stress at the root of the notch (or other stress concentrators), $S_{max}$, to the corresponding nominal stress in the notched cross section, $S_n$:

$$K_t = S_{max}/S_n \quad \text{(Eq 37)}$$

Equations and diagrams for numerous $K_t$ values are reported in standard reference books (Ref 23–27).

### Fatigue Strength-Reduction Factor

Stress concentrations affect the fatigue behavior of different materials and notch geometries differently. For example, cast irons, containing innumerable internal stress raisers, show little further adverse effects for mild notches from externally introduced stress raisers. In contrast, high-strength materials, such as quenched and tempered steels, are more susceptible to the effects of stress raisers than ductile low-strength materials, such as normalized or annealed steels. This can be explained by a model where the initial fatigue damage is not caused by the maximum stress directly in the root of the notch but within a limited process zone, the size of which depends on the material (toughness). Therefore, the effective fatigue-relevant stress depends not only on the stress concentration but also on the size of the process zone, which depends on the material (strength and toughness), and on the stress gradient ahead of the notch, which depends on the notch radius.

The fatigue damage is therefore not caused by the maximum stress, expressed by the theoretical stress-concentration factor, $K_t$, but by a reduced stress, expressed by the fatigue strength-reduction factor, $K_f$. This factor ranges from 1 to $K_t$ and is defined as the ratio of the fatigue strength of a smooth (unnotched) specimen to the fatigue strength of a notched specimen at the same number of cycles:

$$K_f = S_{A,unnotched}/S_{A,notched} \quad \text{(Eq 38)}$$

The correlation of $K_t$ and $K_f$ is given by the cyclic supporting factor, $n$, which depends on the material and the stress gradient:

$$K_f = K_t/n \quad \text{(Eq 39)}$$

Peterson (Ref 24) expressed the fatigue notch factor with the fatigue notch sensitivity, $q$. The fatigue notch sensitivity is defined with the proportion of stresses above the nominal stress:

$$q = (K_f - 1)/(K_t - 1) \quad \text{(Eq 40)}$$

where $q$ varies between 0 (no notch effect, $K_f = 1$) and 1 (full notch effect, $K_f = K_t$). With the value $q$, which is determined for a given material and notch geometry, the fatigue notch factor can be derived from Eq 40:

$$K_f = q(K_t - 1) + 1 \quad \text{(Eq 41)}$$

For example, fatigue tests on smooth and notched specimens ($K_t = 2.2$) for the quenched and tempered steel 42CrMo4V in Fig. 5 resulted in alternating loading ($R = -1$) in $S_{A,unnotched} = 347$ MPa and $S_{A,notched} = 187$ MPa. This corresponds with the following supporting factors: $K_f = 347/187 = 1.85$; $n = 2.2/1.85 = 1.19$; and $q = 0.85/1.2 = 0.71$.

### Reduction of Stresses

The detrimental effect of stress concentrations can be reduced by optimizing the geometry of the notch or by inducing compressive stresses. These stresses may be introduced by plastic deformation of the surface, such as in cold working (rolling, shot peening); by phase transformation, such as in case hardening; or by proof loading with a load high enough to cause local plastic flow in the notch but low enough to avoid general yielding.

Extreme or incorrect cold working may have adverse effects under fatigue conditions, because it may give rise to folds or minute cracks in the surface of the material, or at least make the surface prone to such cracks. If the correct amount of cold working is applied, however, improvement in the fatigue strength is appreciable. For example, a screw thread produced by rolling is more resistant to fatigue failure than one that has been cut. Surface rolling and shot peening, especially for notched sections of crankshafts and springs, are other examples of the successful application of cold working.

Other methods of balancing the effect of stress concentrations are to produce compressive residual stresses by surface hardening, such as flame hardening, case hardening, or nitriding. If performed correctly, any of the processes is beneficial in raising the fatigue strength significantly, but the extent of the treated region (hardening depth) must be carefully chosen; otherwise, failure is likely to occur at the interface between the treated and untreated regions (Fig. 20).

### Residual Stresses

Residual stresses are stresses acting in externally completely unloaded components. The integration of the stress distribution, $\sigma$, across the cross section, $A$, results in the axial force, $F$. In the case of residual stresses, $\sigma_{res}$, the acting load is zero, which means that the residual stress must be internally balanced (self-balancing stress):

$$F = \int \sigma_{res} dA = 0 \quad \text{(Eq 42)}$$

A precondition for the generation of residual stresses is a local plastic deformation (usually at a hot spot). The plastic deformation may result from manufacturing processes or from operational loads.

Practically all manufacturing processes will induce residual stresses, such as casting, rolling, forging, forming, grinding, polishing, machining, and heat treatment. These processes result in a combined effect of residual stresses and strengthening by work hardening, so that the direct isolated effect of the residual stresses is difficult to quantify. The operational load causing residual stresses may be mechanical or thermal. Thermal-induced stresses are of special relevance; plastic deformation is more likely because the yield stress is reduced with increasing temperature. The stresses due to thermal influence are related to stresses as a consequence of restricted thermal expansion, temperature differences across the wall thickness, and/or volume changes due to phase transformation.

The sign of mechanically induced residual stresses in a hot spot is contrary to the responsible loading stress, which means that tension overloading results in hot spots in compressive

residual stresses and vice versa, a compressive overload generates tension residual stresses.

Failures are often induced, or at least favored, by tension residual stresses at the surface. Residual stresses are an important reason for most manufacturing defects (casting cracks, grinding cracks, quenching cracks, welding cracks), for brittle fractures, for fatigue failures, and for stress-corrosion cracking. In addition, residual stresses may lead to distortion, cracking, and even fracture during and after machining, which disturbs the internal equilibrium (Eq 42) and causes plastic strains during stress redistribution.

The influence of residual stresses on fatigue strength is, in principle, similar to that of an externally applied mean stress. However, the mean stress influence due to residual stresses is slightly less than the influence of an external mean stress. This can be explained by the fact that residual stresses are gradually reduced during cyclic loading and by the extreme gradients of residual stresses. The reduction during cyclic loading is explained by local plastic deformation, cyclic creep and relaxation, and local stress-relieving effects due to crack formation.

Similar to Eq 13, the combined influence of external stresses, $S_m$, and residual stresses, $S_{res}$, on the fatigue strength, $S_A$, can be expressed by:

$$S_A = S_{alt} - M \cdot S_m - M_{res} \cdot S_{res} \quad \text{(Eq 43)}$$

The residual mean stress sensitivity, $M_{res}$, is usually lower than the external mean stress sensitivity, $M$ (Fig. 8); for example, for steels up to $S_u$ = 1500 MPa, $M_{res} = M - 0.1$.

From the fatigue diagram in Fig. 7 and from Eq 43, it can be concluded that a compressive residual stress increases the fatigue strength, and a tensile residual stress reduces the fatigue life and strength. The residual-stress effect on fatigue is more pronounced for high-strength materials because—compared to low-strength materials—the residual stresses are usually higher (limited by the yield stress), and the mean stress sensitivity, $M$, is increased with the strength (Fig. 8). Consequently, procedures that generate compressive residual stresses at the surface, such as rolling, needling, shot peening, and surface hardening, are more effective and beneficial in high-strength compared to low-strength materials.

The change of sign between the causative load stress and the generated residual stress can explain fatigue fractures under operational external pure-pulsating compression stresses, in that a possible compressive overload with plastic flow in the notch will, after unloading, shift the stress level in the positive regime, with the result of tension-pulsating loading.

Residual compressive surface stresses improve fatigue strength mainly under conditions involving bending and torsional stresses and under the presence of surface notches. Where the stress is purely tensile, the fatigue strength may be reduced because of the internal tensile stress that balances the compressive stress at the surface (Eq 42). In tensile loading of unnotched structures, the stress is uniformly distributed across the section and is augmented by the residual tensile stress; failure is likely to take place by cracking just beneath the surface (Fig. 7).

The interaction of mechanical, thermal, and residual stresses generally causes a multiaxial stress state (biaxial at the surface), which requires the consideration of residual stresses in the theories of strength (Eq 44–47).

The critical judgment about the influence of residual stresses in failure investigations requires information about the magnitude and direction of the acting residual stresses. This can be achieved by numerical calculations (e.g., finite element) or (preferably) experimental stress analysis, such as the drill-hole technique or the x-ray diffraction method. However, the problem of experimental determination of residual stresses in cracked or even fractured parts is the fact that after the material has separated, possible residual stresses are partly or totally released. Therefore, the residual stresses should principally not be measured on already fractured components. The residual-stress measurement should consequently be carried out on comparable, nonfractured reference structures.

## Multiaxial Stress State

To compare the multiaxial stresses in a component with the uniaxial material property (yield stress, $S_y$; ultimate tensile strength, $S_u$; fatigue strength, $S_A$), the multiaxial stress state must be transferred by application of a theory of strength to a damage-equivalent uniaxial stress, $S_{equ}$. However, the well-known theories of strength, which are historically introduced for static loading, cannot be directly applied to fatigue loading.

In the special case of pure alternating synchronous stresses with principal stress amplitudes $S_{1a}$, $S_{2a}$, and $S_{3a}$, where $S_{1a} \geq S_{2a} \geq S_{3a}$ and $S_A$ is the alternating uniaxial fatigue strength, the fatigue damage condition can be expressed with the failure conditions such as:

- Maximum principal stress criterion: $S_{a,equ} = S_{1a} = S_A$
- Maximum shear stress criterion: $S_{a,equ} = S_{1a} - S_{3a} = S_A$
- Von Mises criterion: $S_{a,equ} = 1/\sqrt{2} \cdot \sqrt{(S_{1a} - S_{2a})^2 + (S_{2a} - S_{3a})^2 + (S_{3a} - S_{1a})^2} = S_A$ (Eq 44)

For example, complex three-dimensional (multiaxial) combinations of stress and strain occur in the vicinity of surfaces and notches. Because fatigue cracks usually propagate from the surface, where one of the three principal stresses perpendicular to the load-free surface is zero, only the biaxial stress state must be considered. It is often convenient to determine the suitability of the various criteria by fatigue tests under torsion (biaxial stress state) and bending (uniaxial stress state) loading.

Theories of strength based on equivalent strains, $\varepsilon_{equ}$, are dependent on the principal strain amplitudes, $\varepsilon_{1a}$ and $\varepsilon_{3a}$ ($\varepsilon_{1a} > \varepsilon_{2a} > \varepsilon_{3a}$):

- Normal strain criterion: $\varepsilon_{a,equ} = \varepsilon_{1a} = \varepsilon_A$
- Shear strain criterion: $\varepsilon_{a,equ} = (\varepsilon_{1a} - \varepsilon_{3a}) = 2 \cdot \gamma_{max} = \varepsilon_A$ (Eq 45)

where $\gamma_{max}$ is the maximum shear strain, and $\varepsilon_A$ the tolerable alternating strain amplitude determined in strain-controlled tests.

The general application of the theories of strength for complex fatigue loading requires the separate treatment of stress amplitudes and mean stresses on one side, and on the other side, not linking the equivalent stress with the principal stress system in order to consider the possible change of direction of the principal stress system. If the directions of the principal stress axis are changing as a function of time, the fatigue damage is not reflected in the equivalent stresses, because they are principally defined as scalar values, for example, von Mises. A typical practical example for a nonconstant principal stress direction is a bending- and torsion-loaded shaft, such as a crankshaft, in which the ratio of shear stress, $T_t$, and normal stress, $S_b$, as a function of time is not constant (Eq 36).

The most effective solution for these load cases is to analyze the amplitudes and mean stresses in all coordinate-oriented cutting planes and to compare the stresses in the cutting planes with the fatigue strength in the fatigue diagram (the critical plane approach). In the critical plane, the cyclic damage is determined by the maximum normal stress amplitude (brittle behavior) and the shear stress amplitude (ductile behavior), including an additional term that describes the mean stress influence.

In the high-cycle fatigue (HCF) regime of ductile materials, the stresses in the critical plane can be expressed according to Findley (Ref 28) by the shear stress amplitude, $\Delta\tau/2$, which is modified with the normal stress, $\sigma_\perp$, in the critical plane to include the mean stress effect. The damage condition is defined by comparing the modified shear stress amplitude with the $S$-$N$ curve for $R = -1$ or the elastic part of the LCF curve (Eq 18), which must be transferred into shear stresses, $\sigma_f'$:

$$(\Delta\tau/2 + k \cdot \sigma_\perp)_{max} = \tau_f'/G(2N)^b \quad \text{(Eq 46)}$$

where the material constant $k$ = (0.2 − 0.3), $\tau_f' = \sigma_f'/\sqrt{3}$, and the shear modulus $G = E/[2(1 + \mu)]$.

The treatment of strain-controlled multiaxial fatigue in the LCF regime of ductile components (Ref 29) is based on the maximum shear strain, $\gamma_{max}$, in the critical plane and the normal stress, $\sigma_{\perp max}$, perpendicular to this plane. The failure condition is given by comparing with the shear LCF curve:

$$\gamma_{max}/2(1 + \sigma_{\perp max}/\sigma_{flow}) = \tau_f'/G(2N)^b + \gamma_f'(2N)^c \quad \text{(Eq 47)}$$

The maximum shear strain is calculated from the principal strains $\gamma_{max} = \varepsilon_1 - \varepsilon_3$. The flow stress is approximated by $\sigma_{flow} \approx \sigma_f'$. The Manson-Coffin parameters for the shear LCF curve are $\tau_f' = \sigma_f'/\sqrt{3}$ and $\gamma_f' = \varepsilon_f'/\sqrt{3}$

More basic definitions of multiaxial effective stresses and strains and the differences between proportional and nonproportional loading are described in Ref 30, along with some basic correlations for multiaxial fatigue, including mean stress effects, sequences of stress/strain amplitude or stress state, nonproportional loading and cycle counting, and HCF limits. Reference 30 also covers the formation and propagation of small cracks (on the order of several grain sizes in diameter, typically less than 1 mm, or 0.04 in., in length) in initially isotropic ductile structural alloys.

### *Effect of Load Frequency and Temperature*

#### Frequency Effect

Exclusive of environmental effects, there are no distinguishing surface features of a fatigue fracture produced at high frequency that differentiate it from other types of fatigue fractures during visual or optical microscope examination. At best, when examined with an optical microscope, a fracture surface created at a high frequency will have a quasi-brittle appearance, showing mostly a platelike structure throughout the fatigue zone. Beach marks may or may not be present, depending on whether crack growth was steady or intermittent or if load variations occurred.

For typical frequency ranges from 60 to 6000 cycles/min (1 to 100 Hz), the fatigue strength of most materials, based on a given number of cycles to fracture, is not or only little affected by frequency as long as no creep effects (creep fatigue) and no corrosion effects (corrosion fatigue) exist. In general, there is a slight decrease in fatigue strength and fatigue life with a decrease in frequency. This is because the amount of plastic deformation that occurs during the stress cycle is slightly increased, due to the fact that at high frequencies there is less relaxation time during each stress cycle for deformation to occur. In turn, this results in less fatigue damage, which means increased lifetime. Another reason for the frequency effect can be related to the fact that even the surrounding air has a minor corrosive effect, which makes the damage cycle- and time-dependent.

#### Elevated Temperatures

In principle, the fatigue strength of metals decreases with increasing temperature, with some exceptions, such as lamellar cast iron. For temperatures below the recrystallization temperature, the fatigue strength is dominated by the reduction of the static strength, $S_u$.

For temperatures exceeding the recrystallization temperature, the damage mechanism creep becomes a decisive factor. The temperature for start of the creep range, $T_{creep}$, is often expressed as a function of the melting temperature, $T_{melt}$, of the material:

$$T_{creep}[K] = (0.35 \text{ to } 0.55) \cdot T_{melt}[K] \quad \text{(Eq 48)}$$

The damage above this temperature range is caused by the superposition of fatigue and creep effects, or creep fatigue (Eq 49). The fatigue damage in the creep fatigue range depends on the total time of the applied stress rather than solely on the number of cycles. This behavior occurs because of the loss of strain-hardening capability, which causes a continuous deformation under constant load.

Fractures resulting from fluctuating stresses at high temperatures may be similar in appearance to fatigue fractures and stress-rupture (creep) fractures or a mixture of the two, depending on the relative magnitudes of the mean and alternating stresses. In failure investigations of steel components, tempering colors visible on the component and fracture surfaces are an indication of the acting temperature. As creep damage becomes dominant, the transcrystalline crack path will change into an intercrystalline failure mode.

At very high temperatures, an important requirement for component resistance, besides resistance against fatigue and creep, is oxidation and other types of high-temperature corrosion. Fatigue strength at high temperature can be reduced by surface attack from fuel ash containing vanadium pentoxide, from leaded fuels, and from other contaminants. In general, however (excluding temperature embrittlement, such as aging, sigma phase, and creep embrittlement), materials are more ductile and consequently less notch sensitive at high temperatures than at room or lower temperatures. Extended operation at high temperatures may result in metallurgical changes in the alloy structure, which may also play a role in reducing fatigue strength. Generally, however, brief exposure to high temperatures, which does not result in such metallurgical changes as recrystallization, tempering, phase changes, coarsening, precipitation, melting, or diffusion, will not have a serious effect on fatigue life upon return to normal operating temperatures.

## Effect of Material Condition

The material condition plays an important role in fatigue assessment and failure analysis. The analysis should include not only the normal material response but also the consideration of material and manufacturing defects. Discontinuities (imperfections) within a metal, either at the surface or subsurface, can adversely affect fatigue strength. These discontinuities may arise from melting practices, primary or secondary working of the material, or may be a characteristic of a particular alloy system. Manufacturing practices may also introduce surface irregularities or induce residual stresses that act as crack-initiation sites. In wrought metals, the movement of metal during forming processes, whether performed at room temperature or at elevated temperatures, makes them common sources of surface discontinuities, such as laps, seams, and cold shuts. Oxides, slivers or chips of the base material, or foreign material can be embedded into the surface by rolling or forging. These surface imperfections produce notches, which act as stress raisers under load and will adversely affect fatigue strength.

### *Strengthening and Heat Treatments*

The microstructure of the material can influence fatigue crack initiation and cyclic crack growth by influencing the plastic deformation process. Under high cyclic plastic strains in the LCF range, the fatigue life of many metals is nearly independent of grain size. In contrast, under low cyclic strains in the HCF regime, the fatigue life of many metals is increased when grain size is reduced, for example, fine-grained steels. However, the effect of grain size on HCF properties is difficult to assess because these properties may be altered by the same treatments that alter grain size. In some alloys, an improvement in resistance to HCF achieved by a decrease in grain size may be partly offset by a deleterious effect on another property. For example, a decrease in grain size is thought to raise the smooth-bar fatigue limit in some steels. However, small grain size may increase the notch sensitivity in these steels; therefore, the net result may constitute no improvement in resistance to fatigue of notched structures. As another example, fine grain size in high-temperature alloys (which may be subjected to both fatigue and creep) may result in high fatigue life but low creep life at normal service temperatures; thus, an intermediate grain size may afford the longest service life.

The influence of chemical composition on fatigue strength of smooth bars is approximately proportional to its influence on tensile strength. The fatigue limit of plain carbon steel generally increases with carbon content. Molybdenum, chromium, and nickel have a similar effect. The fatigue limit of high-strength steel with tensile strength in the range of 1379 MPa (200 ksi) can be increased by the addition of copper. Although the phosphorus content of steel is generally kept at a minimum to prevent brittleness, steels with high phosphorus contents have greater fatigue strength. A sulfur content up to 0.01% has no effect on the fatigue limit of steels. Austenitic steels containing nickel and chromium have a relatively high fatigue limit, together with low notch sensitivity and high resistance to corrosion fatigue in some environments.

The fatigue strengths of titanium alloys are slightly higher (material factor 0.5) (Eq 10) than those of steels of the same strength (material factor 0.4 to 0.45). Some titanium alloys maintain considerable strength up to 500 °C (930 °F) and also have a high resistance to corrosion fatigue. Even though titanium alloys are particularly susceptible to hydrogen embrittlement, the presence of hydrogen without cracking does not affect the fatigue properties.

Aluminum alloys with a material factor of 0.3 (Eq 10) show a relative low fatigue strength compared to steel. Solid-solution strengthening for unnotched specimens results in an increase in fatigue strength to approximately the same degree as the corresponding increase in tensile strength. If the strength of magnesium alloys is increased by solute addition, the fatigue strength is also increased in proportion to the increase in tensile strength.

Second phases, which are often present in metallurgical systems, affect the fatigue crack-growth behavior. Second phases have a marked influence on the mechanism and kinetics of crack nucleation and propagation because they can accelerate or inhibit the crack-propagation rate under various circumstances. For example, investigations on aluminum alloys containing a large number of particles of precipitated intermetallic second phase showed that the particles acted as stress raisers, from which fatigue cracks initiated. The zones adjacent to grain boundaries of age-hardened alloys are generally free of precipitate particles and are relatively soft. Therefore, stress relaxation takes place along the precipitate-free zones, resulting in localized strain concentration and crack nucleation at grain-boundary triple points.

The instability of coherent precipitate particles is considered to be the most important factor responsible for the low fatigue life of high-strength aluminum alloys. Experimental results indicate that reversion takes place during the early stages of cyclic loading. Reversion under cyclic straining involves the passage of dislocations back and forth through the precipitate particles, which disintegrates the particles to subcritical size. The solute then goes back into solution or is distributed along the dislocation network, thereby softening the slip bands and causing crack initiation.

In alloys that are not strengthened by heat treatment (some copper alloys, aluminum alloys, and stainless steels), fatigue strength can be increased by cold working. Work-hardened copper-aluminum alloys (75 wt% Al) do not soften under cyclic loading. This results in lower rates of crack propagation in work-hardened alloys, thereby indicating a small amount of deformation at the crack tip during fatigue. Experimental results on annealed and cold-worked α-brass suggest that a work-hardened alloy is harder at the crack tip than an annealed alloy. However, no improvement in fatigue strength is obtained by cold working of age-hardenable alloys. Alloys that are hardened by transformation processes—for example, martensitic steels—exhibit a lesser degree of improvement by cold working than the nonhardenable alloys.

### Overheating

In heat treatment of metals, high temperatures generally cause large grains to develop. Depending on the metal and application, large grains may have undesirable attributes. In most metals, large grains will generally reduce fatigue strength. The properties of coarse-grained metals are impaired by the size of the grains and by changes that occur at the grain boundaries, such as the precipitation of solid impurities, which may form weak, continuous grain-boundary films. Damage from overheating is especially evident in high-carbon steels, in which all the harmful effects of coarse grains on properties are combined with the probability of cracking during quenching from the hardening temperature.

### Eutectic Melting

In aluminum and other age-hardenable alloys, solution heat treatment improves mechanical strength by developing the maximum practical concentration of the hardening constituents in solid solution. The solubilities of these constituents increase markedly with temperature, especially just below the eutectic melting temperature. Consequently, the most favorable temperature for effecting maximum solution is very near to that at which melting occurs. Melting in age-hardenable alloys produces an intergranular network of nonductile eutectic products and intragranular circular spots (rosettes), both of which reduce ductility and fatigue strength. Similar effects of melting have been observed in high-temperature alloys, such as Stellites and high-speed tool steels.

### Quench Cracks

The origin of quench cracks in steel—a possible cause of failure in hardened steel parts—is attributed to thermal and residual stresses as well as sudden volume changes. The transformation of austenite into martensite is accompanied by volume expansion, which generates tension residual stresses in the surrounding areas and, under unfavorable conditions, may result in cracking. Conditions that promote the formation of quench cracks include too-high hardening temperatures, large grain size, a severe temperature gradient (too-short cooling cycles, severe quenching medium), sharp notches, and rough surface finish.

Sometimes, the cracks do not appear immediately but are delayed and take time to become visible on the outside. Delayed quench cracks are the result of additional transformation of retained austenite in steel. In high-alloyed steels, delayed cracking can be avoided by tempering immediately after quenching.

Quench cracks are generally intergranular. If a quench crack is open to the surface during tempering, the walls of the crack may be covered with scale and may be decarburized.

### Decarburization

This is a loss of carbon at the surface of a ferrous alloy as a result of heating in a medium that reacts with carbon. Unless special precautions (shielding gas, carbon-enriched atmosphere) are taken during heat treatment, the risk of losing carbon from the surface of steel is always present in any heating to high temperatures in an oxidizing atmosphere. A marked reduction in tensile strength and fatigue strength is noted in steels with decarburized surfaces. The effect of decarburization is more pronounced on high-tensile-strength steels than on steels with low tensile strength.

In failure analysis, decarburization can be shown by examining the microstructure (carbon-free zones) and by determining the hardness distribution, for example, Vickers $HV_1$, perpendicular to the decarburized surface.

## *Subsurface Discontinuities*

### Gas Porosity or Shrinkage Porosity

Gas porosities, or gas holes, are rounded cavities (spherical, flattened, elongated, or partially collapsed) that are caused by the generation or accumulation of gas bubbles in the molten metal as it solidifies. Shrinkage cavities result from varying rates of contraction while the metal is changing from the molten to the solid state. Shrinkage porosity is mainly characterized by jagged holes or spongy areas lined with dendrites. The fatigue strength of cast alloys is less reduced by the presence of buried shrinkage or gas porosity, but it is much reduced when the porosity extends to the surface, regardless of severity, because of the increased notch effect.

### Inclusions

The presence of nonmetallic inclusions at or close to the surface is detrimental, because the inclusions act as stress raisers and may form points for the initiation of fatigue cracks. The possibility of fatigue failures increases with the number of discontinuities. However, there does seem to be a limit to the number of inclusions that could cause fatigue. If discontinuities such as graphite flakes in cast iron or sulfide inclusions in free-cutting steels are very numerous, mutual stress relief occurs, based on the principle of a single groove versus a continuous thread, and the material tends to become less prone to fatigue.

Inclusions in ferrous alloys are usually oxides, sulfides, carbides, and silicates. Many inclusions, however, are of a more complex intermediate composition. Every commercial metal has its own characteristic inclusions. In cast metals, the shape of the inclusions is approximately the same in all polished sections. In wrought metals, the shape depends on the orientation of the polished surface. Deformation in mechanical working causes

inclusions to deform plastically to elongated shapes and to appear in longitudinal sections as stringers or streaks, although in transverse section the shape is globular or flat. Hard, refractory, and very small inclusions, such as alumina inclusions in steel, are not deformed by mechanical work but are often broken up and distributed parallel to the direction of working.

Fatigue properties of high-strength alloys are degraded by inclusions, with a more marked effect on the transverse fatigue properties of wrought alloys than on the longitudinal properties. An extreme effect of inclusions, such as manganese sulfides in steels, is observed in loading of rolled plates with fillet welds in the thickness direction perpendicular to the surface (lamellar tearing).

In general, however, the effect of inclusions depends on the size and shape of the inclusions, on their resistance to deformation and their orientation relative to the stress, and on the tensile strength and toughness of the alloy. Soft steels, for example, are much less affected by inclusions than are hard steels.

Inclusions, like most discontinuities, are most damaging when they intersect the surface. However, subsurface inclusions can also be responsible for premature fatigue failure when the stress level at the inclusions is high enough to initiate a crack. The fracture surface of a hardened-steel connecting rod is shown in Fig. 43. Large inclusions that intersect the surface are indicated by arrows. Loads on this rod were transverse to the forging grain flow. Figure 44 shows the surface of a torsional fatigue fracture that initiated at a large subsurface nonmetallic inclusion in a hardened-steel valve spring.

Internal bursts in rolled and forged metals result from the use of equipment that has insufficient capacity to work the metal throughout its cross section. If the working force is not sufficient, the outer layers of the metal will be deformed more than the inside metal, sometimes causing wholly internal, intergranular fissures that can act as stress concentrators, from which fatigue cracks may propagate.

### Alloy Segregation

Distribution of alloying elements in industrial alloys is not always uniform. Localized deviations from the average composition originate from specific conditions during solidification of the alloys. Hot working and soaking tend to equalize the compositional differences, but these differences sometimes persist into the wrought product. Inhomogeneous distribution of alloying elements in heat treated alloys is particularly objectionable, because—in addition to the local changes of strength and toughness—it may lead to the formation of thermal cracks caused by uneven contraction or expansion in heating and cooling.

### Banding

When segregation in an alloy occurs in layers or bands, the alloy is said to have a banded structure. Banding can lead to discontinuities that can in turn cause premature fatigue failure. Figure 45 shows the fatigue fracture of a carburized and hardened steel roller. Banded alloy segregation in the metal resulted in heavy banded retained austenite, particularly in the carburized case after heat treatment. When the roller was subjected to service loads, the delayed transformation of the retained austenite to martensite caused microcracks near the case/core interface. These internal microcracks nucleated a fatigue fracture that progressed around the circumference of the roller, following the interface between case and core.

### Flakes

Also called fisheyes, flakes are bright, crystalline, internal fissures in ferrous metals. They are attributed to stresses produced by localized transformation and decreased solubility of hydrogen during cooling after hot working. For alloy steels, flakes are often associated with high hydrogen content in the steel; chemical-element segregation, producing regions of high alloy content; and rapid cooling from the hot working temperature.

## *Effect of Manufacturing Practices on Fatigue Strength*

### Machining

Most components subjected to fatigue conditions have been machined. The fatigue strength is influenced by the surface-finish quality, for example, roughness (Fig. 4). In addition, heavy cuts and residual tool marks from rough machining can promote fatigue failure in a component. Surface irregularities produced by rough machining act as stress raisers. However, a series of parallel grooves, such as results from turning operations, is less severe in its effect than an isolated groove, because parallel grooves provide mutual stress relief.

Rough machining also changes the metal structure to an appreciable depth. If a machining tool shears off metal, rather than cutting it, it results in a torn and work-hardened surface to an extent that depends largely on the depth of the cut, the type and shape of the tool, and the characteristics of the metal. Components subjected to fatigue loading should be finished with a fine cut, or preferably ground, and the direction of the final cut or grind should be parallel to that of the principal tensile stress whenever practical.

Most mechanically finished metal parts have a shallow surface layer with compressive residual stresses. Apart from the effect it has on surface roughness, the final finishing process is beneficial to fatigue life when it increases the depth and intensity of the compressively stressed layer and will be detrimental when it decreases or removes the layer. Processes such as electrolytic polishing and chemical and electrochemical machining, which remove metal without plastic deformation, may therefore reduce the fatigue properties. Electric discharge machining can also be detrimental to fatigue properties without proper control and subsequent processing because of surface and subsurface microstructural changes and possible microcracking.

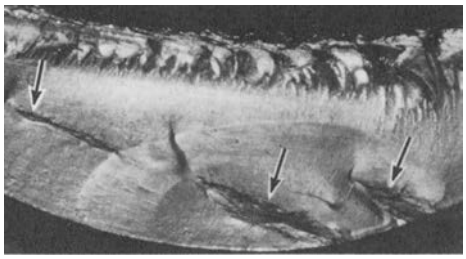

**Fig. 43** Fracture surface of a hardened-steel connecting rod. Arrows indicate large inclusions. Fatigue cracking initiated from the middle inclusion.

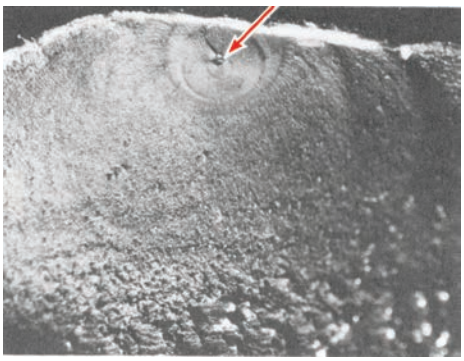

**Fig. 44** Fracture surface of a hardened-steel valve spring that failed in torsional fatigue. Arrow indicates fracture origin at a subsurface nonmetallic inclusion.

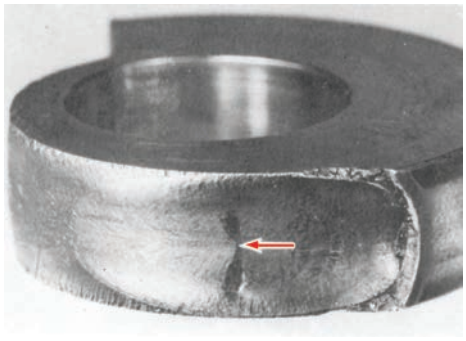

**Fig. 45** Fracture surface of a carburized and hardened steel roller. As a result of banded alloy segregation, circumferential fatigue fracture initiated at a subsurface origin near the case/core interface (arrow)

Improperly controlled grinding can have similar effects on fatigue properties.

Occasionally, during machining of a component, a tool may scratch or groove the surface. If the part is highly stressed in service, the result can be a premature fatigue fracture nucleating at the tool mark. Stress raisers frequently occur at a change in section, such as the shoulder between two shaft sections of different diameters. Rough machining marks, as well as steps resulting from improper blending of fillets with shaft surfaces, may serve as initiation sites for fatigue cracks.

### Drilling

The fatigue strength of components is reduced by the presence of a drilled hole due to the notch effect ($K_t$ = 3 for small holes). It is further reduced by lack of removing burrs (incurred during drilling) from the hole edges. Fractures originating at drilled holes are common in complex parts containing internal, intersecting machined passages because of the difficulty and expense of providing adequate break-edge radii at such locations.

### Grinding

Proper grinding practice produces a smooth surface that is essentially free of induced residual stresses or sites for the nucleation of fatigue cracks. However, abusive grinding, particularly in steels heat treated to high hardness, is a common cause of reduced fatigue strength and failure that results from severe induced tensile residual stresses, intense localized heating, or both. Intense localized heating due to high pressure of the grinding disc and insufficient cooling results in overtempering, formation of untempered martensite, and burning in combination with precipitous stress gradients, which may generate multiple tight and shallow surface cracks perpendicular to the grinding direction, referred to as grinding cracks.

### Straightening

Components may be unintentionally plastically deformed without cracking during manufacturing, transport, or in service. In the ductile condition, these parts can be straightened manually (by heating), in presses, or in roll straighteners. The initial deformation and subsequent working operations will introduce residual stresses or peak stress at notches. Nicks, scratches, or locally work-hardened surfaces are potential stress raisers. For example, aircraft propeller blades, which are damaged by deformation in minor ground accidents, are repaired by cold straightening, either manually or in a press. The straightening operations are restricted to bends of less than a specified angle, varying with the location along the blade.

### Plating

Plating on metal surfaces can be detrimental to fatigue strength of the plated parts. Carbon and low-alloy steels, particularly steels of high hardness, are susceptible to hydrogen damage due to absorption of atomic hydrogen during the plating cycle and the associated acid or alkaline cleaning procedures. A soft plating material, such as cadmium, may inhibit the escape of hydrogen from the steel and lead to hydrogen damage. However, the harmful effects of cadmium plating can be significantly minimized by following recommended procedures, which include plating at high current densities (6 to 8 A/dm$^2$, or 0.4 to 0.5 A/in.$^2$), followed by baking at approximately 190 °C (375 °F) for 8 to 24 h. A daily procedure for ensuring that the plating bath will give low hydrogen pickup is highly desirable. An apparatus for such a procedure is available in the form of an electronic electrochemical instrument that measures the relative amount of hydrogen generated during electroplating.

Hard plating materials are usually in a state of tensile stress after they have been deposited on the base metal. Cracks that may develop in the plating material act as stress raisers and produce (possibly delayed) cracking in the base metal. Several effective procedures for decreasing the residual tensile stress imposed by electroless nickel plating have also been developed, including the addition of saccharin and other organic compounds containing sulfur to the plating bath.

### Cleaning

Sometimes, cleaning of structures is necessary for the removal of oil, grease, or other contaminants from the surface of a workpiece. For production work, vapor degreasing or a dip into a solvent or a simple chemical cleaning solution may be employed. Many alkaline solutions that remove grease and oil from aluminum are not satisfactory for most cleaning purposes because they attack the surface. However, alkaline solutions that are suitably inhibited to eliminate the corrosive action can be used successfully. The cleaning treatment should be followed by thorough rinsing in clean water, followed by drying.

### Arc Burns

Localized regions of untempered martensite can be produced by arc burns that result when the surface of a component is affected by the ignition of a welding electrode or by electrodes used in magnetic-particle inspection. In steels, these procedures result in a local hardness increase, tensile residual stresses, and even micro cold cracking.

For example, fatigue cracking of several knuckle pins occurred within approximately 1 month in service (Fig. 46). Examination showed that the fractures initiated at circular cracks that apparently existed on the pins before their installation. Micrographs of sections through the cracked areas of both unused and failed knuckle pins revealed a remelting zone and an area of untempered martensite within the region of the cracks. The local areas of overheating were attributed to electric arc burning that caused circular cracks. The pins subsequently failed by fatigue fracture. The circular cracks were the result of arc burning, attributable to improper technique in magnetic-particle inspection. The conductor should be insulated to prevent arc burning.

### Identification Marking

Stress raisers and residual stresses may be introduced into components by identification marks (manufacturing date or lot number, steel heat number, size or part number). The location and method of applying the marks can be important. The marks should not be located in areas of high stresses.

Raised numerals or letters are preferable to those that are indented. Hot forging methods are recommended over cold forging, stamping, or coining. For marks that are made on machined surfaces, use of a marking ink is generally preferable to use of an electric etching pencil or vibrating mechanical engraver, but either of the latter two is less abusive than using a cold steel stamp. However, using steel stamps for numbers in a nonstressed area may be preferable to some other type of identification mark in a more highly stressed region.

Related to the direction of the maximum principal stress, characters with straight-line portions (such as "I," "H," or "0") (Fig. 47a) have the utmost tendency to cause cracks in tension and bending, whereas inclined characters (such as "N" or "Y") are most detrimental in torsion (Fig. 47b). Characters with rounded contours can also cause cracking. If steel stamps must be used, low-stress stamps (stamps with the sharp edges removed from the characters) or dull stamps cause the least trouble, particularly if light impressions are

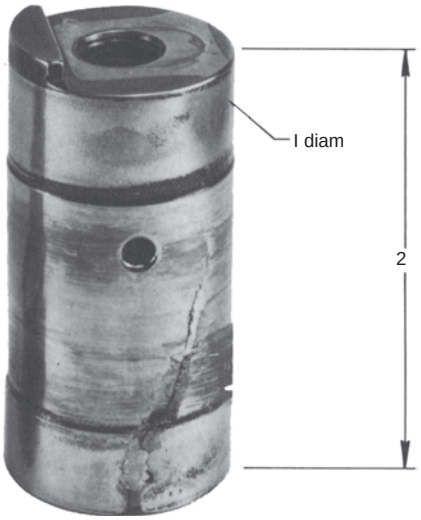


**Fig. 46** Steel knuckle pin that failed in service by fatigue cracking that originated at an arc burn at the bottom of a longitudinal oil hole during magnetic-particle inspection

made. As previously mentioned, the stamping should be located in known low-stress areas.

## Corrosion Fatigue

Corrosion fatigue is associated with alternating or fluctuating stresses in components subjected to a corrosive environment. This damage combination causes accelerated crack initiation and propagation at a location where neither the environment nor the stress acting alone would possibly be sufficient to produce a crack. An important fact of corrosion fatigue is that practically no material in the active condition as well as in the passive condition, such as stainless steels and aluminum and titanium alloys, is completely resistant against corrosion fatigue.

It should be distinguished between fatigue of corroded components (preceding corrosion) and simultaneous interaction of fatigue and corrosive environment (corrosion fatigue). The negative effect of cyclic-loaded, surface-corroded components is related to the corrosion-damaged surface by, for example, rust or pitting, which introduces stress raisers with possible crack initiation on the surface. Corrosion fatigue is the interaction of a mechanical fatigue mechanism and an electrochemical mechanism (anodic or cathodic reaction). The mechanism of corrosion fatigue can be described by the crack initiation at stress raisers at the surface under tensile stresses, including damage of the passive surface layer, the formation of persistent slip bands, and the following anodic dissolution of the blank material. For structures that are susceptible to hydrogen embrittlement (such as ferritic and martensitic steels) or for parts exposed to a continuous corrosive environment with intermittent applications of loading, the cracking mechanism is somewhat complex.

The safety relevance of corrosion fatigue consists of the fact that, compared to loading in a noncorrosive environment, the number of cycles to cause fatigue fracture in the finite-life range is significantly reduced. Of specific safety relevance is the fact that (for ambient temperatures more or less clearly defined) the fatigue strength decreases constantly with the number of cycles. This causes a finite-life situation even at low cyclic stress amplitudes. In terms of fracture mechanics, the corrosive environment reduces the threshold value, $\Delta K_{th}$, and significantly increases the cyclic crack-growth rate, $da/dN$.

In contrast to testing under an ambient environment, the test results under corrosion-fatigue conditions are strongly dependent on the frequency, in a sense that decreasing the frequency leads to a shorter life time and reduced fatigue strength. The increased testing time enhances exposure to the time-dominated corrosion mechanism. Corrosion-fatigue tests should consequently be carried out under the relevant medium and the real stress-time cycles (frequency, shape).

The remedies to avoid corrosion fatigue are mainly focused on an optimized design with minimized notch effect and gaps, protection against the corrosive medium (such as saltwater, chlorides, hydrogen), avoidance of stagnant conditions, application of surface coatings, lowering of tension residual stresses (stress relieving), generation of compressive residual stresses at the surface (e.g., shot peening), treatment of the medium (inhibitors), and use of cathodic protection.

### *Appearances*

Fatigue cracking in a corrosive environment is identified by the corroded surface, the presence of numerous small cracks adjacent to the fracture, and the compacted corrosion product on the fracture surface, which may damage and obscure fine surface details of the fracture. Corrosion-fatigue cracks generally initiate at the surface. Surface features at origins of corrosion-fatigue cracks vary with the alloy and with specific environmental conditions. In materials in the active condition, such as carbon steels, cracks often originate at hemispherical corrosion pits and often contain significant amounts of corrosion products (Fig. 48a). The cracks are often transgranular and may exhibit a slight amount of branching. Corrosion-fatigue cracks sometimes occur in the absence of pits, for example, in materials in the passive state, and follow grain boundaries or prior-austenite grain boundaries (Fig. 48b). Crack-propagation paths in steels may be intergranular or transgranular and in alternating sequence, depending on material, stress level, frequency, temperature, and environment.

Extensive corrosion often damages fracture surfaces and can make a positive identification of the fracture mechanism difficult. At low frequencies and low temperatures, fatigue crack growth in hydrogen, for example, follows primarily prior-austenite grain boundaries, although some cleavagelike fracture is also observed. At high frequencies and temperatures, the crack tends to follow a transgranular path.

Both corrosion fatigue and mechanical fatigue may exhibit multiple origins and ultimately join to form a single crack front on a single plane. However, it is sometimes possible to distinguish corrosion fatigue from purely mechanical fatigue by the number of cracks propagating through the part. Frequently, several corrosion-fatigue cracks form and propagate simultaneously along parallel paths. Yet, mechanical-fatigue cracks may initiate at several points in the same general region on a part, but one crack usually becomes dominant, or several cracks join to form a single front, before cracking has progressed far into the part. Environmental effects can usually be identified by the presence of corrosion damage or corrosion products on fracture surfaces or within growing cracks. Corrosion products,

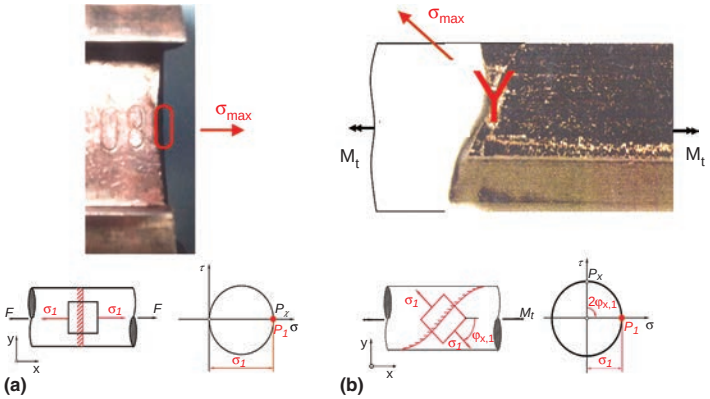


**Fig. 47** Fractures due to identification marking. (a) Keyhole of a rotating-bending-loaded shaft. (b) Fracture of a torsion-loaded hexagonal transmission shaft

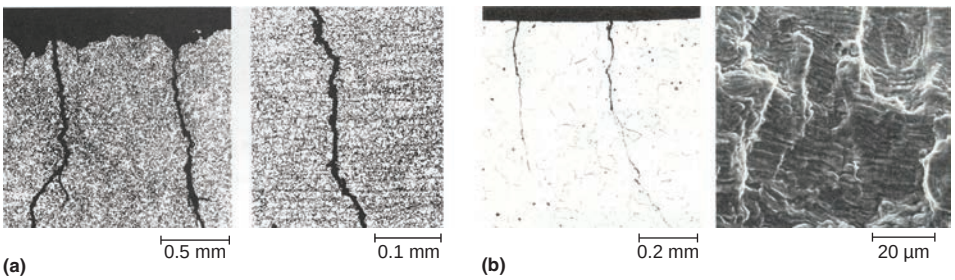


**Fig. 48** Corrosion-fatigue cracks. (a) Ferritic steel in active state. (b) Austenitic steel in passive state. Source: Ref 31

however, may not always be present. For example, corrosion-fatigue cracking of high-strength steel exposed to a hydrogen-producing gas, such as water vapor, may be difficult to differentiate from some other forms of hydrogen damage, such as hydrogen-induced slow-strain-rate embrittlement and sustained-load cracking. At sufficiently high frequencies, the fracture-surface features produced by corrosion-fatigue crack initiation and propagation do not differ significantly from those produced by fatigue in nonaggressive environments.

In aluminum alloys exposed to aqueous chloride solutions, corrosion-fatigue cracks frequently originate at sites of pitting or intergranular corrosion. Initial crack propagation is normal to the direction of the principal stress. This is contrary to the behavior of fatigue cracks initiated in dry air, where initial growth follows crystallographic planes. Initial corrosion-fatigue cracking normal to the axis of principal stress also occurs in aluminum alloys exposed to humid air, but pitting is not a requisite for crack initiation. Corrosion-fatigue cracking susceptibility in aluminum alloys also does not appear to be sensitive to grain directionality and grain shape. This is in contrast to the propagation of stress-corrosion cracks in susceptible wrought aluminum alloys, which show greater susceptibility in the short-transverse direction.

Corrosion-fatigue cracks in copper and various copper alloys initiate and propagate intergranularly. Corrosive environments have little additional effect on the fatigue life of pure copper over that observed in air, although they change the fatigue crack path from transgranular to intergranular. Alternatively, copper-zinc and copper-aluminum alloys exhibit a marked reduction in fatigue resistance, particularly in aqueous chloride solutions. This type of failure is difficult to distinguish from stress-corrosion cracking except that it may occur in environments that normally do not cause failures under static stress, such as sodium chloride or sodium sulfate solutions.

### Examination

Analysis of fracture-surface deposits can be vital to the identification of corrosion fatigue. The first step in the macroscopic investigation should be to strip the fracture surface with cellulose acetate tape or plastic replicating material to entrap any deposit for later microanalysis (energy-dispersive x-ray analysis). Then, whether complex electron optic or simple photographic documentation is to be used, the piece should be cleaned and degreased in an ultrasonic cleaner containing reagent-grade acetone, or similar organic solvent, as the tank medium.

The main advantage of reagent-grade acetone is that it leaves a minimal amount of organic residue on the fracture surface yet removes almost all traces of the cellulose acetate left from stripping. Also, acetone is an excellent solvent for wet magnetic-particle suspensions and for most liquid penetrants, which are often used in the field for nondestructive detection of cracks. If the part is too large or otherwise not amenable to ultrasonic cleaning, a solvent may be sprayed onto the surface of the part using a squeeze bottle or an aerosol can; the solvent should not be applied with a brush.

Features at the origin of a corrosion-fatigue fracture are often indistinct, because crack closure effects (favored by corrosion products) have forced mating fracture surfaces together and have formed an extremely rubbed, discolored origin. Also, the area of the origin is exposed to the environment for the longest time and thus may exhibit more extensive attack than the rest of the fracture surface. A corrosion product at fatigue origins can be misleading when viewed macroscopically. The oxide may be flat and tenacious, globular, or nodular. In the stereomicroscope at magnifications of less than 50×, globular or nodular oxide particles are easily mistaken for intergranular facets. Therefore, no conclusions should be drawn until other metallographic or fractographic techniques confirm the mechanism of crack initiation. Occasionally, oxidation is so severe that no information other than location of origin can be obtained.

Other features that can be observed macroscopically are secondary cracks, pits, and fissures, all of which are often adjacent to the main origin of a particular fracture. In corrosion-fatigue failures, cracks and fissures adjacent to the primary origin are indicative of a uniform stress at that location. Therefore, the primary fracture simply propagated from the flaw that either induced the most severe stress concentration or was exposed to the most aggressive local environment. Sometimes the primary crack origin is related more to the heterogeneous microstructure than to the stress distribution.

In the SEM investigation, striations may not be detectable for materials in the active state, because of the corrosion attack at the crack edges (Fig. 48a). However, cyclic crack growth in materials in the passive state, such as stainless steel and aluminum and titanium alloys, with oxide layers on the surface and the crack flanks, is usually characterized by striations (Fig. 48b).

## Contact Fatigue

Rolling-contact fatigue is a type of surface damage that results from repeated rolling or rolling and sliding contact between curved metal surfaces. The significant stress in rolling-contact fatigue is the maximum alternating shear stress that undergoes a reversal in direction during rolling. In pure rolling, this shear stress occurs on a plane slightly below the surface and can lead to initiation of fatigue cracks in the subsurface material (see the article "Rolling-Contact Fatigue" in this Volume). As these cracks propagate under the repeated loads, they reach the surface, producing pits.

When sliding is imposed on rolling, the tangential forces and thermal gradient caused by friction alter the magnitude and distribution of stresses in and below the contact area. The alternating shear stress increases in magnitude and is moved nearer to the surface by sliding forces. Thus, initiation of contact-fatigue cracks in gear teeth, which are subjected to significant amounts of sliding adjacent to the pitchline, is found to be in the surface material. These cracks propagate at a shallow angle to the surface, and pits result when the cracks are connected to the surface by secondary cracks. If pitting is severe, the bending strength of the tooth may be decreased to the point at which fracture can occur.

The existence of several distinctly different modes of contact-fatigue damage has been recognized, but the factors that control the initiation and propagation of each type are only partially understood. The contact fatigue of hardened steel has been the subject of extensive research using analytical and experimental methods, and publications have revealed a large variation in the contact-fatigue strength of similar materials when tested under different conditions of contact geometry, speed, lubrication, temperature, and sliding in combination with rolling. These variations result from a poor understanding of the basic mechanism of contact fatigue. Although the cause of the fatigue and the locations of the origins are somewhat different, these are true fatigue fractures that are caused by shear stress origins due to contact stresses between two metal surfaces.

In addition to an incomplete understanding of the basic mechanism, the bearing and gear industries have, in the past, used completely different nomenclature to describe identical fatigue damage modes. In 1996, American National Standard ANSI/AGMA 1010-E96 was published after several years' work to standardize nomenclature for failure modes that is compatible with both gear and bearing industries. Reference 32 also uses nomenclature for bearings that is similar to that used in ANSI/AGMA 1010-E96 for gears. To improve communication, it is recommended to maintain compatible nomenclature in accordance with ANSI/AGMA 1010-E96 such that:

- The term *macropitting* should be used instead of *spalling*. *Spalling* has been used in the past to describe several different failure modes and therefore is confusing. It should be avoided where possible.
- The term *micropitting* should be used instead of *peeling*. *Peeling* is not a good name for micropitting because it does not describe either the appearance or the mechanism of the failure mode. *Micropitting* is a much better name because it describes both appearance and mechanism and links the failure mode to macropitting, which is similar except for scale.

Table 3 lists terminology used by the bearing and gearing industries to describe the fracture appearance of contact-fatigue-failed components. The following sections describe the types of damage from contact fatigue based on the categories in Table 3 (Ref 33). However, general prevention and characteristics of damage represent the limit of load-carrying capacity with the following origins.

### Surface-Origin Pitting

This is the type of pitting characteristic of combined rolling and sliding, such as occurs on gear teeth slightly away from the exact pitchline, where only rolling is present. Also, cam and cam-follower assemblies suffer this type of failure on the cam follower itself. The maximum shear stress is no longer buried within the metal but is now brought to the surface because of the friction at the interface. Because the character of the surface is intimately involved, the lubricant becomes critical. It is for this reason that transmissions, final drives, and various other types of gear systems require lubricants with special additives. Everything possible must be done to reduce surface friction, expand the contact area, and reduce the load in order to reduce the contact stress and provide a hard, fatigue-resistant structure in the metal. Unfortunately, if those things are already near optimum, little can be done to improve the fatigue life. Only minimal improvements, if any, can be expected. Improvements can best be obtained, if at all, by changes in the lubricant.

### Subsurface-Origin Pitting

This type of pitting destroys the integrity of contacting surfaces where one surface rolls under pressure across another. This action is typical of antifriction bearings, such as ball, roller, tapered roller, and needle bearings, as well as other rolling devices such as roller cams. It is prevented by eliminating the hard and brittle angular inclusions. A principal application for vacuum-treated steels is the manufacture of antifriction, or rolling-element, bearings and other critical parts and is now common practice for such components.

### Subcase Fatigue

The previous types of contact-stress fatigue result in small, often tiny, pits in the contacting surfaces. Subcase-origin fatigue is different because very large cavities may be formed, apparently in a very short time. The fatigue originates near the case/core interface where the shear stresses are highest. Because the basic cause is inadequate metal strength in the subcase region, the logical solution should be to strengthen the metal there. This can be done by increasing the case depth and/or the core hardness but must be done carefully, because the changes could lead to through-hardening and brittle fracture, particularly on gear teeth.

### Macropitting

*Macropitting* is a general term that includes spalling and other forms of macroscale damage (Table 3) caused by Hertzian contact fatigue. Macropitting of bearing raceways or gear teeth is generally due to contact fatigue, which occurs from localized plastic deformation, crack initiation, and finally macropitting from crack propagation in and near the contact surface. Macropitting results from the subsurface growth of fatigue cracks, which may have a surface or subsurface origin. Figure 49 shows macropit formation by subsurface growth of surface-origin pitting on a gear tooth. The bottom of inclusion-origin macropits form first, and so forth. When circumstances cause surface-origin pitting, crack growth often occurs within and beyond the range of maximum shear stresses in the contact stress field (called hydraulic-pressure-propagation crack paths that may be due to chemical reactions at the crack tip as well as to lubricant viscosity effects).

When fully developed, the craters exist at a depth comparable to that of the maximum alternating Hertzian shear stress. A macropit does not result from the gradual enlargement of a small cavity, but rather it results from the subsurface growth of a fatigue crack, which eventually separates from the main body of material. In the example of a gear damaged by crater formation due to surface-origin macropitting (Fig. 49), the craters are generally steep walled and essentially flat bottomed, because the walls and bottom are formed by fracture surfaces. After the pits initiate at the surface, the bottoms of macropits usually form next, and then the crack extends through the contact surface, creating the walls of the crater. However, there are circumstances in which macropits are initiated at or near the contact surface and the crack propagates through the Hertzian shear stress field, forming the walls and then the bottom of the crater (Ref 34).

Macropitting fatigue life is inherently statistical because defect severity and location are randomly distributed among macroscopically identical contact components. Two major classes of macropitting damage are distinguished according to the location of the initiating defect: subsurface- and surface-origin macropitting.

#### Subsurface-Origin Macropitting

This type of pitting results from defects in the bulk material subjected to the Hertzian cyclic stress field. Subsurface-origin macropits are characterized by a smooth area parallel to the contact surface with a steep wall exit (inclined by more than 45° with respect to the contact surface). Sometimes the subsurface defect that caused the damage remains visible after the macropit forms. The defects that cause subsurface-origin macropits are inclusions and/or material microstructure alterations. The defects that most likely produce this type of macropit are located above the depth of maximum alternating Hertzian shear stress. Variables governing the life of a component with respect to this type of macropit are Hertzian shear stress level, material-matrix fatigue resistance, defect severity, and defect location.

Inclusion-origin fatigue is normally initiated at nonmetallic inclusions below the contact surface and is the primary mode of contact-fatigue damage in antifriction bearings. The initiation and propagation of fatigue cracks are the result of cyclic stresses that are locally intensified by the shape, size, and distribution of nonmetallic inclusions in steels. Oxide inclusions from deoxidation, reoxidation, or refractory sources are the most frequently observed origins of inclusion-origin fatigue damage. Sulfide inclusions alone are rarely associated with contact-fatigue cracks. The ductility of sulfide inclusions at hot working

**Table 3 List of contact-fatigue terminology used to describe the different fatigue mechanisms**

| Macroscale | Terms |
|---|---|
| Macropitting | Pitting<br>Initial pitting<br>Destructive pitting<br>Flaking<br>Spalling<br>Scabbing<br>Shelling<br>Fatigue wear |
| **Microscale** | **Terms** |
| Subcase fatigue | Case crushing |
| Micropitting | Microspalling<br>Frosting<br>Glazing<br>Gray staining<br>Surface distress<br>Peeling |

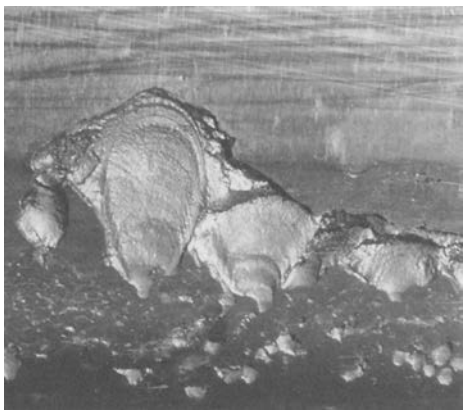

**Fig. 49** Fractograph showing the advanced stages of macropitting fatigue for a helical pinion tooth

temperatures and a relatively low interfacial energy make them less effective stress concentrators than the harder, incoherent oxide inclusions. In addition, oxide inclusions are often present as stringers or elongated aggregates of particles, which provide a much greater statistical probability for a point of stress concentration to be in an area of high contact stress.

When a nonmetallic inclusion as described previously results in significant stress concentration, two situations are possible: localized plastic flow occurs and/or a fatigue crack begins to propagate with no plastic flow, as observed microscopically. If localized plastic deformation occurs, microstructural alterations begin to develop, and the effective stress concentration of the inclusion is reduced. The probability of fatigue crack initiation/propagation is less than before the plastic flow occurred, so a macropit may not form. However, if a fatigue crack forms due to the inclusion, prior to any plastic deformation, the stress concentration is increased, and crack propagation results in macropitting fatigue. This hypothesis includes the possibility of a fatigue crack originating from an inclusion that is associated with some sort of microstructural alteration but also proposes that microstructural alterations are not a necessary step in the initiation/propagation of inclusion-origin fatigue.

Early macroscopic propagation of inclusion-origin fatigue takes on a characteristic pattern. Many branching subsurface cracks are formed around the inclusion, with propagation often being most rapid perpendicular to the rolling direction, with a resulting elliptical shape for the incipient macropit. In some instances, the cracked material around the inclusion macropits early and further propagation is relatively slow, leaving a small macropit with very few subsurface cracks to cause further damage. However, when material has macropitted from the contact surface, the result may be debris bruising and denting of other contact areas with subsequent surface-origin macropitting.

### Surface-Origin Macropitting

Surface-origin macropitting is caused by defects in the immediate subsurface material subject to asperity-scale cyclic stress fields and aggravated by surface tractive forces. These defects are either preexisting defects, such as nicks, dents, grinding furrows, surface discontinuities, and so forth, or microscale pits, which are described subsequently. The distinguishing features of a surface-origin macropit are in the entrance zone of the macropit. The entrance zone may exhibit a shallow-angle entry wall (inclined less than 30° to the contact surface), an arrowhead configuration (Fig. 50), the presence of a visible surface defect, or an association with a stress concentration due to design geometry. Surface-origin macropits have been classified by the nature of the defect: geometric stress concentration and point surface origin.

*Geometric Stress-Concentration Macropits*

Geometric stress-concentration (GSC) macropits are distributed on the contact surface at the ends of line contact. When the contact geometry, deflection under applied loads, and alignment cause the contact stress to be higher at the end of line contact, fatigue occurs within a narrow band in which the contact stresses are more severe than those associated with inclusions. The GSC macropits can propagate more rapidly in either direction, parallel or transverse to the rolling direction. This mode of contact fatigue is more frequently observed in life tests accelerated due to overload conditions, because contact geometry is designed to produce uniform loading under expected service loads.

Crowning of bearing rollers and gear teeth in line-contact situations has been shown to be an effective method of prolonging life or preventing the GSC mode of macropitting. The fact that GSC macropits can occur simultaneously at both ends of a line contact indicates that this mode of contact fatigue is the result of end stress concentrations and not simply misalignment. However, misalignment can aggravate the situation by increasing end contact stresses. Early stages of GSC macropitting exhibit cracks extending to the surface, but it is not clear whether the first cracks always originate at the surface or slightly subsurface. However, the origins are rarely associated with a nonmetallic inclusion, but occasionally an inclusion at the end of line contact will serve as the nucleus for a GSC fatigue macropit.

*Point Surface-Origin Macropits*

Point surface-origin (PSO) macropits occur randomly distributed on the contact surface, similar to inclusion-type macropits. However, PSO macropits are different in two important aspects: first, there is no consistent association with nonmetallic inclusions; second, the origin is located at or near the contact surface, whereas with inclusion-origin macropits, the initiation site is located subsurface. These characteristics result in a distinct arrowhead appearance, in which the tip of the arrowhead is the origin, and the advanced stages of the macropit develop in a fanned-out appearance with respect to the tip in the rolling direction. Figure 51 shows an example of a tapered roller-bearing cone that was damaged due to PSO macropitting.

**Fig. 50** Gear with arrow-shaped surface-origin pit. Source: Ref 35

## *Micropitting*

*Micropitting* is the preferred term for peeling fatigue, which is defined as microscale spalling fatigue. It is the damage of rolling/sliding-contact metal surfaces by the formation of a glazed (burnished) surface, asperity-scale microcrack, and asperity-scale micropit craters. Micropitting damage is a gradual type of surface fatigue damage that is a complex function of surface topography and its interaction with the lubricant. Even under high contact loads, a thin film of lubricant between the contacting surfaces from elastohydrodynamic (EHD) lubrication is present. Cumulative plastic deformation first results in a surface with a mirrorlike reflection, known as a glazed surface. The surface and immediate subsurface material of a glazed surface is heavily deformed and, with continued cyclic stressing, the ductility of the material decreases and microcracks form. The cracks tend to propagate parallel to the contact surface at depths comparable to that of asperity-scale shear stresses.

A microcracked surface is not usually distinguishable from a glazed surface by the unaided eye, but with microscopy, microcracks opening to the surface may be visible in glazed areas. As the microcracks multiply and propagate, the surface becomes undermined (in asperity dimensions), and multiple microscopic pits form. A micropitted surface appears frosted to the unaided eye, with black spots representing the micropits (Fig. 52). If extensive, the micropitting is considered a failure.

The occurrence of micropitting is controlled by the EHD film thickness-to-roughness ratio and surface microgeometry. In the presence of a sufficiently high EHD film-thickness-to-surface-roughness height ratio, micropitting does not occur because the film prevents high-contact microstress in asperity interactions. However, film thickness may be sufficient to prevent generalized micropitting but

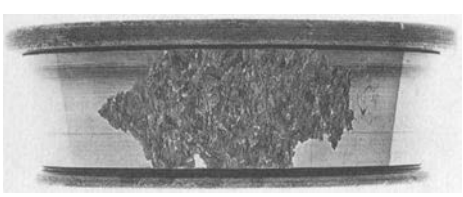

**Fig. 51** Fractograph showing the advanced stages of point surface-origin macropitting fatigue of raceway in which the origin is still visible

not local micropitting in the vicinity of surface defects such as nicks and dents that locally depressurize and thereby thin the EHD film. Tribology features that affect micropitting are both the height and sharpness (root mean square height and slope) of the asperities. There is also some evidence that the normal wear in boundary-lubricated gear contacts removes incipient surface-origin pitting, which is one reason gears may be less prone to macropitting from surface-origin macropitting under some circumstances.

Micropitting (or peeling) is characterized by the limited depth of cracking and propagation over areas rather than propagation in depth. Micropitting fatigue cracks usually propagate parallel to the surface at depths from 2.5 to 13 mm (0.1 to 0.5 in.) and rarely as deep as 25 mm (1 in.). The high stress gradient that exists in the vicinity of asperities explains why such cracks do not propagate to greater surface depths of the macro-Hertzian contact stress field. In essence, micropitting removes the asperity ridges and relieves the high stress gradient. An additional visual characteristic is that the micropitting occurs more or less in the rolling direction, following finishing lines.

Macropitting fatigue (spalling) is the most hazardous consequence of micropitting. The principal operating factor influencing the rate of macropitting from micropits is tractive interfacial stress. When friction is high, macropits form rapidly from the micropits on the surface. If early macropitting does not terminate the operation of a component that has suffered micropitting, then surface material delamination may occur. The original surface is thereby lost over wide areas, and the component becomes inoperative through loss of dimensional accuracy, noise, or secondary damage.

In many cases, micropitting is not destructive to the bearing raceway or gear-tooth surface. It sometimes occurs only in patches and may stop after the tribological conditions have been improved by run-in. Run-in is defined as an initial transition process occurring in newly established wearing contacts, often accompanied by transients in coefficient of friction, wear rate, or both, that are uncharacteristic of the long-term behavior of the given tribological system. Gears appear less sensitive to the consequences of surface distress than rolling-element bearings. This may be due to the greater ductility of lower-hardness materials.

**Fig. 52** Gear with micropitting (frosting) failure. Source: Ref 35

## *Subcase Fatigue*

Subcase fatigue is the fracture of case-hardened components by the formation of cracks below the contact surface within the Hertzian stress field. However, the depth at which the cracks form is much greater than those that cause macropitting fatigue and is a function of material strength in conjunction with the alternating Hertzian shear stresses. Subcase fatigue is also sometimes referred to as case crushing, but because it results from a fatigue crack that initiates below the effective case depth or in the lower carbon portion of the case, the former nomenclature is used.

In case-hardened components, produced by carburizing, nitriding, or induction hardening, a gradient of decreasing hardness (shear strength) exists from the case to the core. If the material shear strength gradient is steeper than the gradient in Hertzian shear stress, then the applied-stress-to-strength ratio is least favorable at some depth below the normal maximum Hertzian shear stress region. This depth is usually located at (or near) the case/core interface due to shear strength and stress gradients, in addition to unfavorable tensile residual stresses. The reversal in the sign of the residual stress from compression in the case to tension in the core occurs at or very near the case/core interface, and therefore, the Hertzian shear stresses are increased due to the tensile residual stresses.

The fatigue originates near the case/core interface where the shear stresses, which are still fairly high, exceed the strength of the relatively low-strength core metal in that region. Fatigue cracks propagate parallel to the surface but underneath the case, then they join and form cracks that come up through the case to the surface, resulting in the removal of relatively large chunks of metal. Figure 53 shows micrographs of a gear tooth and a roller specimen in which subcase fatigue cracks had formed and propagated during testing. A typical example of this type of failure in gear teeth is shown in Fig. 54.

Subcase fatigue is usually caused by a too-small case depth, such as on a nitrided gear tooth, by insufficient core hardness on carburized or induction-hardened parts, or in heavily loaded case-hardened rolling-element bearings. Factors that tend to promote the various types of contact fatigue are listed in Table 4. Note that three of the five modes of fatigue damage are influenced by lubrication, including all surface-origin damage modes.

Figures 55(a) to (d) show Hertzian shear stresses and the material shear strength as a function of depth below the contact surface.

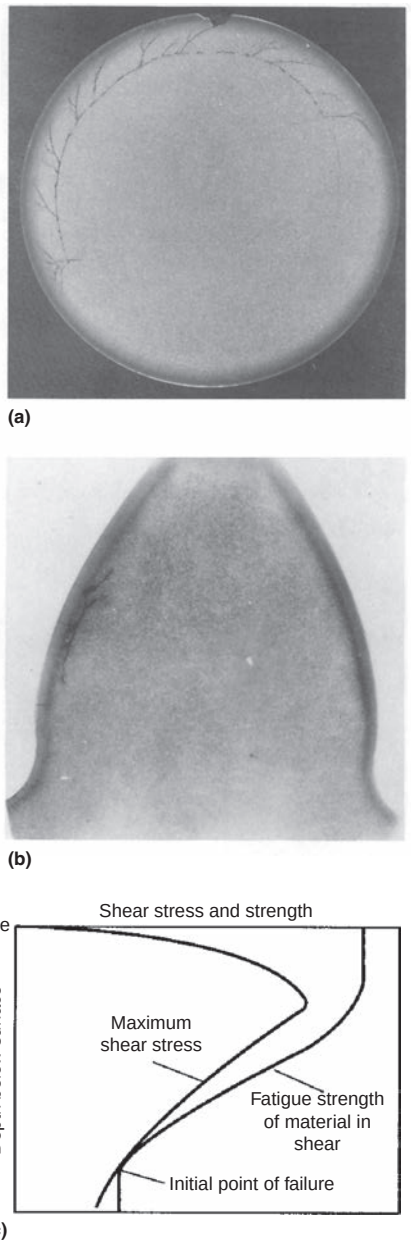


**Fig. 53** Subcase fatigue cracking. The micrographs show etched cross sections of (a) a carburized cylindrical test specimen (original magnification: 1.9×) and (b) a carburized gear tooth in which subcase fatigue cracks initiated and propagated during testing. Subcase fatigue, also known as case crushing, shows the early stages of fatigue cracking in the lower carbon portion of the carburized case. (c) The diagram indicates how the stress and strength gradients combine to cause the weakest condition at the subcase location where fatigue cracking begins.

These are an adaptation of the subcase fatigue analysis by Pederson and Rice (Ref 37). Because fatigue occurs when the applied cyclic shear stress is greater than the fatigue shear strength of the material, the different modes of contact fatigue can be described by determining where these stress profiles cross each other (Fig. 55).

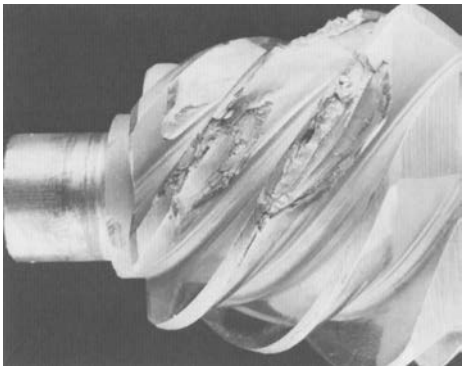

**Fig. 54** Pinion with several large cavities where metal from the surface down to the depth of the case has fallen out due to subcase fatigue. Source: Ref 36

**Table 4 List of contact-fatigue modes and their controlling factors**

| Mode of failure | Factors that control occurrence |
|---|---|
| **Subsurface** | |
| Inclusion origin | Size and density of oxide or other hard inclusions<br>Absence of other modes of failure |
| Subcase fatigue | Low core hardness<br>Thin case depth relative to radius of curvature and load |
| **Surface** | |
| Point surface origin | Low lubricant viscosity<br>Thin elastohydrodynamic (EHD) film compared with asperities in contact<br>Tangential forces and/or gross sliding |
| Geometric stress concentration | End contact geometry<br>Misalignment and deflections<br>Possible lubricant film thickness effects |
| Micropitting | Low lubricant viscosity<br>Thin EHD film compared with asperities in contact<br>Loss of EHD pressure<br>Slow operating speeds |

## Thermal Fatigue

Thermal fatigue at high operating temperatures can be distinguished into loading under cyclic mechanical stresses or strains under constant temperature (isothermal fatigue) and cyclic stresses due to cyclic temperature (thermomechanical fatigue, or TMF). At temperature ranges exceeding the recrystallization temperature, the damage mode creep becomes a major factor. The following ASTM International test standards cover the relevant procedures: ASTM E 606-12 (low-cycle fatigue), ASTM E 139-00 (creep), ASTM E 2714-13 (creep fatigue), and ASTM E 2368 -10 (thermomechanical fatigue).

Isothermal fatigue as an interaction between creep and fatigue damage can be investigated by stress- or strain-controlled fatigue tests at the relevant temperature with and without different holding (dwelling) periods. The creep influence shifts the *S-N* and LCF curves to shorter cycles to crack initiation and fracture.

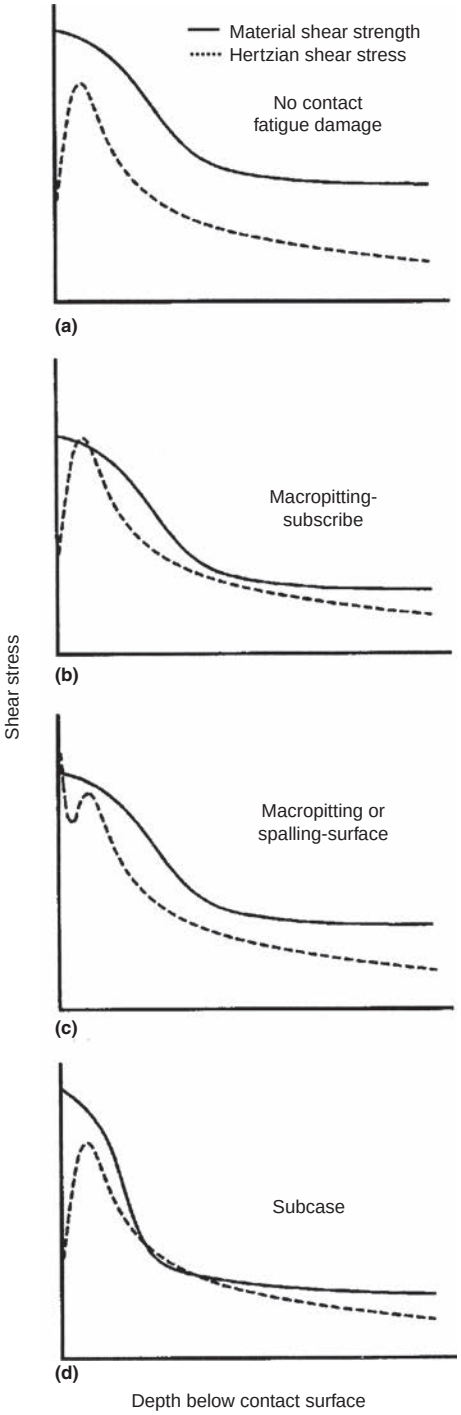


**Fig. 55** Applied shear stress and material shear strength as a function of depth representing types of fatigue damage. (a) No damage. (b) Subsurface-origin, macropitting fatigue. (c) Micropitting or surface-origin macropitting fatigue. (d) Subcase fatigue

In contrast to fatigue tests at ambient temperature, the frequency and shape of the cycle is now important, because creep damage is governed by the testing time. Therefore, with decreasing frequency and holding time, the number of cycles to crack initiation and fracture decreases.

The TMF damage is either caused by restriction of the free thermal expansion due to rigid mountings (e.g., piping systems in power plants), by combination of materials with different coefficients of expansion (e.g., stainless steel cladding on ferritic steel), or by temperature gradients due to rapid temperature changes (start/stop cycles in turbines). The thermal-induced mechanical stresses in connection with the reduced yield stress at elevated temperatures cause alternating reversed plastic flow during the hot/cold phases, which leads to crack initiation after a limited number of cycles (LCF). Thermomechanical fatigue occurs in components such as internal combustion engines (cylinder heads, pistons, valves, exhausting system), brakes, furnaces, power plant components (pipes, vessels, heat exchangers), and turbine blades.

The combined damage at high-temperature exposure is considered in the lifetime assessment by a damage sum, which has contributions from fatigue, creep, and oxidation:

$$D = D_{\text{fatigue}} + D_{\text{creep}} + D_{\text{oxidation}} \quad \text{(Eq 49)}$$

The failure will usually occur at a critical damage sum less than 1 (e.g., 0.3 to 0.5).

The damaging effect of TMF cycles can be evaluated on the basis of isothermal LCF curves or (preferably) on real TMF tests under cyclic temperature. The TMF material behavior is based on tests conducted under a uniaxial stress state on small, unnotched specimens subjected to temperature cycles generated by inductive coils. The expansion of the specimen is controlled by gripping the specimen in a servohydraulic testing machine (Fig. 56), which simulates different amounts of constraint by stiffness variation of the clamping and in application of superimposed mechanical loads, possibly with holding periods at the peak temperatures.

The induced-damage-relevant mechanical strain, $\varepsilon_{\text{mech}}$, in the specimen is the difference of the total strain, $\varepsilon_{\text{tot}}$, and the thermal strain, $\varepsilon_{\text{therm}}$:

$$\varepsilon_{\text{mech}} = \varepsilon_{\text{tot}} - \varepsilon_{\text{therm}} \quad \text{(Eq 50)}$$

For the unixaxial stress state, the thermal strain is calculated from the linear expansion coefficient, $\alpha$, and the temperature difference, $\Delta T$:

$$\varepsilon_{\text{therm}} = \alpha \cdot \Delta T \quad \text{(Eq 51)}$$

For totally restricted expansion ($\varepsilon_{\text{tot}} = 0$), the mechanical strain according to Eq 50 and 51 is $-\alpha \cdot \Delta T$; for free unrestricted expansion ($\varepsilon_{\text{tot}} = \alpha \cdot \Delta T$), the mechanical strain is 0.

According to the temperature-time cycle relative to the mechanical strain-time cycle, the TMF tests are classified into in-phase (IP) or out-of-phase (OP). The OP cycle, in which the mechanical compressive strain coincides with the maximum temperature, is generally more damaging compared to the IP cycle

(Ref 38). The OP test simulates the behavior of a hot surface during a cooling cycle (e.g., edges of a turbine blade).

### Identification

Identifying features of low-cycle thermal-fatigue failures are multiple initiation sites that join randomly by edge sliding to form the main crack, gaping cracks, oxide wedges filling the crack, a blunted crack tip, and a transgranular or intergranular fracture, sometimes with branching of the cracks (Fig. 57). The cracks extend perpendicular to the principal stress direction and often show a crack-arrest condition.

Intergranular fractures are caused by creep-dominated stress-rupture phenomena (long periods at elevated temperature under high static tensile load). The primary failure mechanism involved in stress rupture (creep) is grain-boundary sliding. In thermal-fatigue cracks, slip processes and cleavage operate much as they do in failure at normal temperature, but the evidence is often destroyed by oxide formation, flame polishing, and melting processes.

Under certain circumstances, thermal-fatigue and creep failures blend into each other. Thermal fatigue is the basic mechanism in failures that occur because of numerous short heating and cooling cycles. Stress rupture becomes an important consideration as the cycle time increases and therefore is primarily a long-term process. Most thermal-fatigue fractures are of the low-cycle, high-strain type; the fracture surfaces are rough and faceted at or near the initiation sites and are more fibrous and with shear lips at 45° angles in the final fracture area.

## ACKNOWLEDGMENT

This article was revised from "Fatigue Failures," *Failure Analysis and Prevention*, Volume 11, *ASM Handbook*, ASM International, 2002, p 700–727.

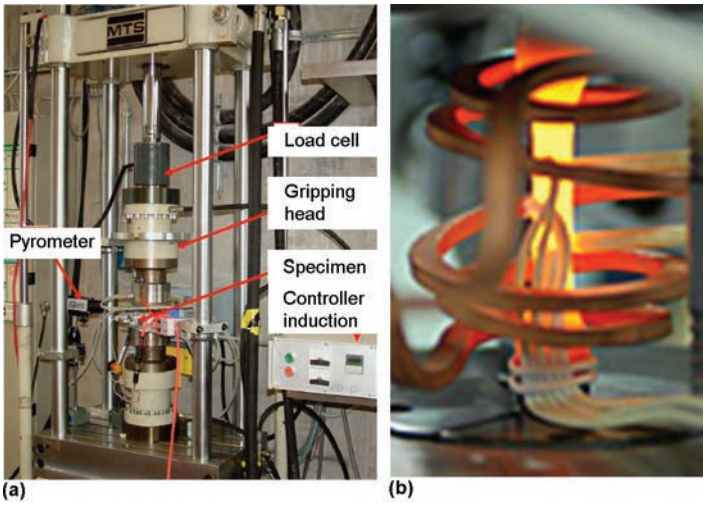


**Fig. 56** Example of a thermomechanical fatigue (TMF) test rig. (a) MTS servohydraulic testing machine (100 kN) equipped for TMF testing. (b) Induction-heated round specimen

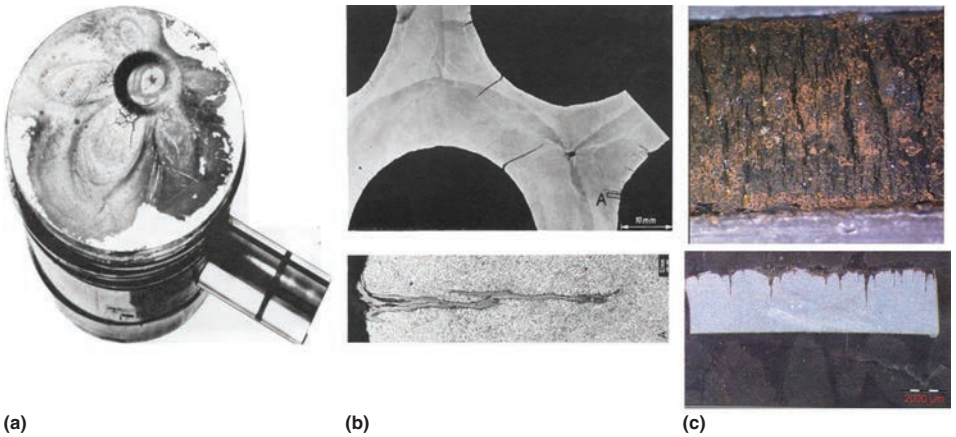


**Fig. 57** Examples of thermomechanical fatigue cracks in structures. (a) Piston of a diesel engine. (b) Heat exchanger. (c) Cooling channel

## REFERENCES

1. Forschungskuratorium Maschinenbau (FKM), *FKM Guideline: Analytical Strength Assessment*, 6th ed., VDMA Verlag GmbH, Frankfurt/Main, 2013
2. B.P. Haigh, Report on Alternating Stress Tests of a Sample of Mild Steel, *Rep. Br. Assoc.*, Vol 85, 1915, p 163–170
3. J. Goodman, *Mechanics Applied in Engineering*, Longmans, Green & Co., London, 1899
4. W. Gerber, Bestimmung der zulässigen Spannungen in Eisenkonstruktionen, *Z. d. Bayer, Architekten u. Ingenieurvereins*, Vol 6 (No. 6), 1874, p 101–110
5. W. Schütz, Über eine Beziehung zwischen der Lebensdauer bei konstanter und veränderlicher Beanspruchungsamplitude und ihre Anwendbarkeit auf die Bewertung von Flugzeugbauteilen, *Z. Flugwiss.*, Vol 15, 1967, p 407–419
6. A. Palmgren, Die Lebensdauer von Kugellagern, *VDI-Z.*, Vol 68 (No. 14), 1924, p 339–341
7. M.A. Miner, Cumulative Damage in Fatigue, *Trans. ASME*, Vol 67, 1945, p A159–A164
8. M.R. Mitchell, Fundamentals of Modern Fatigue Analysis for Design, *Fatigue and Fracture*, Vol 19, *ASM Handbook*, ASM International, 1996, p 227–249
9. S.S. Manson, Fatigue: A Complex Subject—Some Simple Approximations, *Exp. Mech.*, Vol 5 (No. 7), 1965, p 193–226
10. L.F. Coffin, Low Cycle Fatigue—A Review, *Appl. Mater. Res.*, Vol 1 (No. 3), 1962, p 126
11. M.R. Mitchell and R.W. Landgraf, *Advances in Fatigue Lifetime Predictive Techniques*, STP 1122, American Society for Testing and Materials, 1992
12. J.C. Newman and I.S. Raju, An Empirical Stress Intensity Factor Equation for the Surface Crack, *Eng. Fract. Mech.*, Vol 15, 1981, p 185–192
13. R.O. Ritchie, Near Threshold Fatigue Crack Propagation in Steels, *Int. Met. Rev.*, Vol 24 (No. 5, 6), 1979
14. F. Erdogan and M. Ratwani, Fatigue and Fracture of Cylindrical Shells Containing a Circumferential Crack, *Int. J. Fract. Mech.*, Vol 6, 1970, p 379
15. P.C. Paris, "*The Growth of Cracks due to Variation in Load*," Ph.D. thesis, Lehigh University, Bethlehem, PA, 1960
16. S. Pearson, Initiation of Fatigue Cracks in Commercial Aluminum Alloys and the Subsequent Propagation of Very Short Cracks, *Eng. Fract. Mech.*, Vol 7 (No. 2), 1975, p 235–247
17. R.C. McClung, K.S. Chan, S.J. Hudak, Jr., and D.L. Davidson, Behavior of Small Fatigue Cracks, *Fatigue and Fracture*, Vol 19, *ASM Handbook*, ASM International, 1996, p 153–158

18. A.J. McEvily, *Fatigue Crack Thresholds, Fatigue and Fracture*, Vol 19, *ASM Handbook*, ASM International, 1996, p 134–152
19. S.T. Rolfe and J.M. Barsom, *Fracture and Fatigue Control in Structures—Application of Fracture Mechanics*, 3rd ed., American Society for Testing and Materials, 1999
20. D.A. Ryder, "*The Elements of Fractography*," AGARDograph No. AGARD-AG 155-71, Advisory Group for Aerospace Research and Development of NATO, 1971
21. M. Hayes, Fatigue of Springs, *Fatigue and Fracture*, Vol 19, *ASM Handbook*, ASM International, 1996, p 363–369
22. R.M.N. Pelloux, *Met. Q.*, Nov 1965, p 34
23. H. Neuber, *Theory of Notch Stresses: Principles for Exact Calculation of Strength with Reference to Structural Form and Material*, AEC-Tr-4547, Springer Publishing, 1958; available through NTIS, U.S. Department of Commerce
24. R.E. Peterson, *Stress Concentration Design Factors*, John Wiley & Sons, 1974
25. W. Pilkey, *Peterson's Stress Concentration Factors*, 2nd ed., John Wiley & Sons, 1997
26. W.C. Young, *Roark's Formulas for Stress and Strain*, 6th ed., McGraw-Hill, 1989
27. A. Blake, *Practical Stress Analysis in Engineering Design*, 2nd ed., Marcel Dekker, 1990
28. W.N. Findley, Modified Theories of Fatigue Failure under Combined Stress, *Proc. SESA*, Vol 14 (No. 1), 1956, p 35–46
29. A. Fatemie and D.T. Socie, Critical Plane Approach to Multiaxial Fatigue Damage Including Out-of-Phase Loading, *Fatigue Fract. Eng. Mater. Struct.*, Vol 11 (No. 3), 1988, p 149–165
30. D.L. McDowell, Multiaxial Fatigue Strength, *Fatigue and Fracture*, Vol 19, *ASM Handbook*, ASM International, 1996, p 263–273
31. J. Hickling, Korrosionsschäden bei zusätzlicher mechanischer Beanspruchung, *Systematische Beurteilung technischer Schadensfälle*, 5th ed., G. Lange, Ed., Auflage, DGM, Wiley-VCH, 2001
32. C.A. Moyer, "Comparing Surface Failure Modes in Bearings and Gears: Appearances versus Mechanisms," AGMA Technical Paper 91 FTM 6, 1991
33. R.S. Hyde, Contact Fatigue of Hardened Steel, *Fatigue and Fracture*, Vol 19, *ASM Handbook*, ASM International, 1996, p 691–703
34. T.E. Tallian, *Failure Atlas for Hertz Contact Machine Elements*, American Society of Mechanical Engineers, 1992
35. D.G. McPherson and S. Rao, Mechanical Testing of Gears, *Mechanical Testing and Evaluation*, Vol 8, *ASM Handbook*, ASM International, 2000, p 861
36. D.J. Wulpi, *Understanding How Components Fail*, 2nd ed., ASM International, 1999, p 197
37. R. Pederson and S.L. Rice, Case Crushing of Carburized and Hardened Gears, *SAE Trans.*, Vol 68, 1960, p 187
38. H. Sehitoglu, Method, *Advances in Fatigue Life Prediction Technique*, M. Mitchell and R. Landgraf, Ed., American Society for Testing and Materials, 1992, p 47–76