Fatigue of Short Cracks: the Limitations of Fracture Mechanics


ABSTRACT Cracks less than some critical length, defined as $l_0$, show anomalous growth behaviour which cannot be quantified by LEFM. At present, there are no reliable methods for predicting the growth rates and thresholds of these cracks. This paper presents a simple approach which has had some success in predicting the value of $l_0$ for various materials. This parameter is shown to depend on the material properties, yield strength, and grain size. The prediction of $l_0$, for a given material, enables the designer to make conservative predictions of the fatigue life of components in situations where short crack behaviour dominates.

Introduction

In recent years there have been many investigations into the behaviour of short fatigue cracks; this body of work is well summarized by three recent reviews (1)–(3). Possibly the most important point to emerge is that, where short cracks are concerned, present design procedures, whether based on S/N type data or on fatigue crack propagation data, are in danger of being non-conservative. This can be illustrated using the now-common method of displaying short crack data, i.e., the plot of threshold stress range as a function of crack length. Figure 1 shows schematically the type of results obtained; the open-circle data points represent the region in which use of either the fatigue limit or the long-crack threshold value gives too high a value for the threshold stress and is thus non-conservative.

In a paper concerned with the fatigue behaviour of an aluminium bronze alloy (4) the author defined the terms $l_0$ and $l_1$ as the values of crack length at the limits of this non-conservative region. The parameter $l_1$ is possibly only of academic interest, since there are few practical situations for which defect sizes as small as $l_1$ are important. However, the parameter $l_0$ has considerable importance in design and failure analysis work; $l_0$ values may be greater than 1 mm in some materials, so in components for which there is good control of surface condition and defect size, the initial defect sizes may be below $l_0$.

Given the practical difficulties of measuring $l_0$ values, some method of estimating $l_0$ from other material parameters is needed. The author himself has had to deal with failure analysis problems which have required the estimation of an $l_0$ value, for example, the failure of cast ships’ propellers (5). If the practical situation is such that inherent cracks larger than $l_0$ exist, then a normal
crack-tip stress distribution (8), crack deflection (9), and constraint (1). These various models will be discussed in more detail below, where it will be shown that they can be divided into two types: those based on microstructure, for which the parameter \( d \) is important, and those based on plasticity for which the appropriate parameter is the plastic zone size. The extent of the reversed plastic zone \( r_p \) in fatigue can be estimated as \( 0.04 (\Delta K/\sigma_{y}^{*})^2 \), where \( \sigma_{y}^{*} \) is the cyclic yield strength.

In what follows it will be shown that the plasticity based arguments can be reduced to the condition that, for LEFM to apply, the crack length must be greater than \( 10r_p \); hence a second estimate of \( l_2 \) is

\[
l_2 = 10r_p
\]

Equations (1) and (2) represent two independent conditions on the value of \( l_2 \); thus, for any given material, logic dictates that \( l_2 \) will be given by whichever is the larger of \( 10d \) and \( 10r_p \).

There is a fundamental point which should be borne in mind which makes this simple analysis of \( l_2 \) possible: there are many different factors which affect the behaviour of a short crack, including the five listed above, and this makes the prediction of the growth rate and threshold of a short crack very difficult. However, each of these different effects can be thought of as having an \( l_2 \) value associated with it, so the observed value of \( l_2 \) for the material will be whichever is the largest of these, assuming only that the various effects act independently.

\[
l_2 = 10 \mu m
\]

A closure argument predicts \( l_2 = 100 \mu m \) for the same material, then \( 100 \mu m \) will be the relevant value and microstructural effects can be discounted as far as \( l_2 \) is concerned.

Comparison of predictions and results

Figure 2 shows measured values of \( l_2 \) for various materials (3)(4)(10)–(17) (19)(34) plotted against \( 10d \) and \( 10r_p \). The materials covered include mild steels, high strength steels, titanium, copper, and aluminium alloys; the open circles are the \( 10d \) values and the solid circles are the \( 10r_p \) values. The letters alongside the points refer to the original publications, as shown in Table 1.

The accuracy of the determined values of \( l_2 \) is estimated to be within 20 per cent unless the error bars on Fig. 2 indicate to the contrary, likewise the accuracy of the \( 10d \) and \( 10r_p \) determinations is estimated at 20 per cent unless otherwise indicated.

Figure 2 shows that there is good predictive capacity with this model over the whole range of \( l_2 \) values and different materials. Taking the largest of \( 10d \) and \( 10r_p \) (when both are available) then of the 17 different results presented, 15 of the predictions are either correct or slightly conservative. For the 11 cases where both \( d \) and \( r_p \) values were available, there are two cases in which \( r_p > d \) and six cases where \( d \) and \( r_p \) are equal within the errors of estimation.
Another important point to note is that the values of \( d \) and \( r_p \) are always similar in magnitude here; in only one case (3) do they differ by more than a factor of three, though the data range over three orders of magnitude. This arises because we are considering near-threshold behaviour; it has been shown (7)(20)–(22) that \( d \) and \( r_p \) are approximately equal near the long-crack threshold, \( \Delta K_{th} \). The implication of these results is that, if the value of \( L_2 \) is controlled by \( r_p \), rather than by \( d \), then the value of \( L_2 \) might be expected to increase with increasing applied load. In the present paper, only near-threshold results are used, and \( L_2 \) is defined only at the long-crack threshold; few results exist at present to show how anomalous short-crack behaviour varies with applied load, though clearly this point should be pursued in the future.

**Hypotheses relating to short crack behaviour**

The above section has shown that the simple hypothesis described by equations (1) and (2) is capable of giving good predictions for \( L_2 \) values. The following section examines various proposed models of short crack behaviour in order to justify this hypothesis in the light of the various mechanisms advanced by other workers. Five distinct reasons for anomalous short-crack behaviour have been advanced:

1. microstructure (4);
2. closure (1);
3. \( K \)-estimation errors (8);
4. crack deflection (9);
5. constraint (1).

These have already received much attention and discussion elsewhere (1)–(3). Here it will be shown that the microstructure and deflection arguments can be related to the 10\( d \) prediction and that the other three arguments can be related to the 10\( r_p \) prediction.

**Microstructure**

Observations by the present author and others (4)(5)(15)(17) showed that anomalous growth rates in short cracks frequently occurred when cracks were growing in a single grain or a small number of grains. It was noted that linear elastic fracture mechanics demands a homogeneous continuum, which is unlikely to be the case until the crack front length becomes considerably greater than the grain size. Considering a typical surface crack, semi-circular in shape, a crack length of 10\( d \) would give a crack front length of about 31\( d \), which would satisfy a homogeneity requirement such as the one outlined in section (4) below. Thus equation (1); i.e., \( L_2 = 10d \) can be thought of as a sufficient condition on the value of \( L_2 \) in this case, and therefore a conservative prediction. It must be remarked, however, that \( d \) is the mean of a grain-size distribution; thus, some cracks of length 10\( d \) will pass through fewer than 31 grains along
their fronts. It can be shown (23) that there is only a very small probability of
the crack front passing through less than 10 grains at this length, given the
normal form of the grain-size distribution, and assuming no grain-shape texture
effects, such as are common in wrought alloys.

(2) Crack closure

Since crack closure is caused by residual stresses in the crack wake, it has been
proposed (e.g., (19)) that a short crack will experience less closure because there
is insufficient length behind the crack tip for the wake field to fully develop. This seems to be a very powerful argument, and it has certainly been
shown that the closure characteristics of long and short cracks differ considerably.
The picture is confused by the difficulty of accurately measuring short-crack closure; only very small changes in compliance are detected, and by the
effect of microstructural features such as grain boundaries (15)(24).

It has been proposed (1) that for a homogeneous continuum the crack length
should be greater than the reversed plastic zone size, \( r_p \), in order to establish the
full wake field and therefore ensure normal closure behaviour. This seems a
rather short length; as yet there is no reliable evidence to settle the question,
but in that case the condition of equation (2), i.e., \( l_2 = 10r_p \) will express a
sufficient condition on \( l_2 \).

(3) \( K \) estimation errors

Perhaps not enough attention has been given to the fact that the normally-used
equation for stress distribution at a sharp crack

\[
\sigma = \sigma_0 \sqrt{a/2r}
\]

where

\[
\sigma = \text{stress at a distance, } r, \text{ from the crack tip}
\]

and

\[
\sigma_0 = \text{applied nominal stress}
\]

is valid only for \( r \) very much less than \( a \) and should be replaced in other cases by the more complete form of the Westergaard equation

\[
\sigma = \frac{\sigma_0(1 + r/a)}{[2(a/r) + (r/a)^2]^{1/2}}
\]

Sinclair and Allen (6) have shown that for short cracks which show anomalous
behaviour, the difference between equations (3) and (4) is significant. However it is difficult to define a \( K \) value from equation (4) because of the lack of a simple \( r \) singularity. Using a method based on comparison of plastic zone
size, Sinclair and Allen define an effective \( K \) value, \( K_{\text{eff}} \), as

\[
K_{\text{eff}} = K(1 + r_p/a)^{1/2}
\]

Equation (5) implies that for any value of \( r_p \), \( K_{\text{eff}} \) will be greater than \( K \).
However, if \( r_p/a \) is small, \( K_{\text{eff}} \) will approach \( K \) in magnitude. If \( K_{\text{eff}}/K \)
is arbitrarily set to 1.05 (i.e., taking a 5 per cent difference between \( K_{\text{eff}} \) and \( K \)),
this will lead to a value for \( R_p/a \) of approximately 0.1, which would amount to a
restatement of the condition in equation (2). The choice of the figure of 5 per
cent is arbitrary, but represents the amount of scatter usually observed on
threshold data measurement. The use of 10 or 2 per cent would lead to
estimates of \( l_2 \) which would be of the same order of magnitude.

(4) Crack deflection

Figure 3 shows schematically the difference between an idealized straight crack
and a real crack. The crooked, deflected crack path tends to lower the effective
\( K \) value, as has been considered very thoroughly by Suresh (9) and Kitagawa
et al. (25). Suresh also showed qualitatively that a short crack, because it has only
a few deflections in it, would be expected to have a \( K \) value which varied from
one crack to another, whereas in a long crack these effects would average out,
giving a consistent value for \( K \). Suresh was unable to use this deflection
argument to explain the increased growth rates of short cracks, and indeed this
does not seem to be possible unless one postulates a very strong dependence on
the \( K_\text{II} \) value.

---

Fig 3  Schematic of idealized and real cracks

Considering the definition of \( l_2 \), this deflection condition amounts to another
microstructural effect, since it has been shown that for cracks growing under
near-threshold conditions a faceted or ‘structure sensitive’ growth mode occurs
(21) in which the facets are equal to the grain size.

Two effects can be considered here: the effect of crack length and the effect
of crack front length. Considering crack front length, a deflected crack will be
modelled here as shown in Fig. 4, i.e., a crack with a straight section of length
\( a \) and an end portion of length \( d \) deflected through an angle \( \theta \). Using (27) it is
possible to estimate the reduction in effective \( K \) value resulting from deflecting
the end portion. For example, if \( \theta = 45 \) degrees the \( k \) value is reduced from its
nominal value by about 20 per cent. This figure is only slightly dependent on the
values of \( a \) and \( d \), because an increase in the ratio \( d/a \) tends to decrease \( K_\text{II} \) but
'to increase \( K_\text{II} \) for a given \( \theta \).
FATIGUE OF SHORT CRACKS: THE LIMITATIONS OF FRACTURE MECHANICS

This analysis implies that, as far as this condition is concerned, a crack with a short front may grow faster or slower than a crack with a long front. Taking again the case of a semi-circular surface crack, if the crack length is 10d then the crack front length will be 31d, so the condition \( l_2 \approx 10d \) will again express a sufficient condition for \( l_2 \).

The effect of deflections along the crack length is more difficult to analyse. Suresh (9) considered the singly and doubly deflected cracks and developed a method to deal with a long crack containing many deflections, but as yet the critical range of lengths around \( a = 10d \) cannot be treated.

The other postulated effect of crack deflection, as advanced by Beavers and others, is that it induces premature closure, the roughened crack surfaces coming into mutual contact at a higher applied \( K \) value in the cycle than expected. If we note again that the degree of surface roughness is proportional to the grain size at near-threshold growth, a similar microstructural condition, i.e., equation (1), would be expected.

(3) Constrain

It has been pointed out (1) that a small crack is essentially contained in the surface plane stress field until it penetrates some distance into the body of the material, where part of the crack front begins to experience plane strain conditions. Unfortunately it is by no means clear whether a crack in a plane stress field will grow faster or slower than a plane strain crack; experimental results are conflicting (e.g., compare (26) and (27)).

Following Knott (28), it can be assumed that the plane stress field extends into the material from the surface for a distance approximately equal to the plastic zone size. This is shown schematically in Fig. 6 for a semi-circular crack.

Applying the condition of equation (2), i.e., putting \( a = 10r_p \), simple geometry shows that approximately 94 per cent of the crack front will be out of the plane stress region; 6 per cent experiencing plane stress conditions. This

---

**Fig 4** Model deflected crack to be used in analysis

Taking a crack front only one grain in length, and allowing \( \theta \) to vary at random from +45 degrees to -45 degrees gives a distribution of possible \( K \) values; expressed as the ratio \( K/K_{\text{nominal}} \) these vary from 0.99 to 0.87 if one takes two standard deviations from the mean of the distribution. Thus a very small crack front such as this would be expected to show about 13 per cent variation in its \( K \) value if a large number of cracks were examined. If the crack front length is now increased, using a model crack front, as shown in Fig. 5, then the distribution of possible \( K \) values tends to narrow, as the effects from separate grains tend to cancel each other out. Figure 5 shows the results of a simple computer analysis of this model; here \( K/K_{\text{nominal}} \) is plotted as a function of crack front length, expressed as a number of grain diameters. It can be seen that for a crack front longer than 10d the distribution becomes narrow and roughly constant at about 2 per cent of the mean.

**Fig 5** Results of analysis on the effect of crack front length on \( K \). The scatter bands are placed at two standard deviations from the mean of the data

---

**Fig 6** Schematic illustrating the amount of the crack front length which is contained within the surface plane stress region
would seem enough to ensure that the plane stress region was unimportant; hence, equation (2) is again shown to be a sufficient condition on the value of $l_2$.

**Summary of the above discussion**

It has been possible to show that the initial hypothesis used in this paper, that $l_2$ can be estimated as the larger of $10d$ and $10r_p$, constitutes a sufficient condition for each of the five effects considered. Since we are seeking a conservative prediction, a sufficient condition is all that is required. The experimental results shown in Fig. 2, however, suggest that the condition used here is generally either correct or slightly conservative only, so it seems that no severe overestimate of $l_2$ is likely.

**Methods of production of short cracks**

When examining experimental data on this subject, it is important to consider the method used by the experimenters to produce the short cracks, as this may affect their behaviour.

In most experimental studies of short cracks, the cracks are produced by initiation from plain specimens, the threshold values being achieved either by load shedding or by load increase after stress relief heat treatment. The process of stress relief presents some problems (29) but the two methods seem to give roughly comparable results.

However, some workers (e.g., (18)(30)(32)) produce short cracks by machining material from specimens containing long through-cracks. James and Knott (30) point out the possible nature of stress history effects and show that, in an alloy steel, even if stress relief is attempted, a much higher value of $l_2$ would be deduced from these short through-cracks than from surface cracks of more common form. Incomplete stress relief, leaving residual closure stresses, may be involved. Similar results were obtained by Breat et al. (18) and in an aluminium alloy by Zeghoul and Petit (32).

**A conservative design approach**

The emphasis throughout this paper has been to formulate and justify a simple method of predicting an approximate and conservative value of $l_2$. It is believed that this addresses a growing need for designers and failure analysts attempting to use fracture mechanics in fatigue situations.

It is to be hoped that in the near future, analytical methods will be developed to predict the growth behaviour of cracks shorter than $l_2$; some progress has been made in that direction using a statistical approach (37), thus extending our predictive capacity into this difficult area. Until this has been achieved, a conservative approach must be used, such as setting the defect size to $l_2$ in the case where the real defect size is smaller than $l_2$.

Finally, it should be recognized that the number of situations in practice where small crack behaviour dominates is small. The most widely quoted example of the short crack problem is the jet engine turbine blade, a situation in which the extreme demands placed on the material require a very small inherent defect size. Indeed this is probably the only commonly occurring case of a crack of length less than $l_2$ being subjected to stresses well in excess of its propagation threshold. However a number of other cases arise which involve short cracks stressed close to their thresholds; one example would be precision machine parts such as engine pistons, which require very good surface finish and may be made from materials with quite large $l_2$ values. It is unlikely that NDT crack inspection techniques can be used to help monitor short cracks in service, as the necessary detection accuracy cannot be achieved by any NDT method presently in use.

**Conclusions**

1. Cracks less than some critical length, denoted $l_2$, show anomalous growth behaviour which cannot be described by linear elastic fracture mechanics.
2. Difficulties in the experimental measurement of $l_2$ call for some method of predicting this parameter for any required material. The predicted value of $l_2$ should be conservative, i.e., an overestimation, in order to be useful for design purposes.
3. It is proposed that $l_2$ can be estimated to be the larger of $10d$ and $10r_p$, $d$ being the effective grain size and $r_p$ the cyclic plastic zone size.
4. It is shown by comparison with experimental data that this hypothesis gives a reasonable prediction of $l_2$ for many different materials.
5. The use of this simple predictive model is justified in detail by examining the various current theories on short crack behaviour.

**References**

THE BEHAVIOUR OF SHORT FATIGUE CRACKS


