SOME UNRESOLVED ISSUES WITH FATIGUE CRACK CLOSURE - MEASUREMENT, MECHANISM AND INTERPRETATION PROBLEMS

M N JAMES

School of Manufacturing, Materials and Mechanical Engineering
University of Plymouth, Drake Circus, Plymouth
Devon PL4 8AA, England

ABSTRACT

The literature on fatigue crack closure comprises more than 1000 papers and the proceedings of several symposia. Nonetheless, the complexity of the problem and continuing unresolved issues, which are often ignored in publications, limit our ability to deal in a rational predictive way with fatigue phenomena. The present paper highlights several areas in which a more concerted and collaborative approach would give dividends, and comments on several closure controversies. It is hoped to stimulate further discussion and collaboration.

KEYWORDS

Fatigue crack closure, measurement position and resolution, plasticity-induced closure, surface cracks, controversies.

INTRODUCTION

The concept of fracture surface interaction leading to a decrease in stress intensity at the crack tip was explicitly stated as long ago as 1963 in a paper by Christensen. His study was concerned with trapping of oxide debris between the fracture surfaces of centre-cracked plates of 2024-T3 aluminium alloy. Testing was performed in tension-compression and tension-tension loading. The results demonstrated that an increase in fatigue life occurred in the presence of trapped particles in both loading cases, but was greater when part of the fatigue cycle was compressive. Interestingly, Christensen (1963) also performed photoelastic experiments showing the effect of a trapped metal particle on crack tip stresses under equal values of applied tensile and compressive loads (Fig. 1).

Although other authors may have continued research into this topic in the 1960's, details are not known to the present author and it appears that papers by Elber (1970, 1971) provided the spur (probably because they discussed the concept in terms of fracture mechanics parameters) for a vigorous research effort into the mechanisms and phenomena associated with fatigue crack closure. In contrast to the extrinsic mechanism proposed by Christensen (fretting oxidation of fracture surfaces), Elber (1971) proposed an intrinsic mechanism, wedging of the fracture surfaces due to the large plastic wake associated with plane stress crack growth. Plasticity-
induced crack closure was initially thought to be only a surface, or plane stress, phenomenon whose effect was maximised in the linear regime of the growth rate curve (Lindley and Richards 1974, Irving et al 1975). It has since been shown, however, to occur under plane strain conditions (Ewalds and Furnée 1978, Fleck and Smith 1982) and at low fatigue crack growth (fgc) rates.

In the near-threshold regime, however, additional mechanisms have been shown to become operative and closure levels usually increase markedly. Oxide-induced closure (Benoit et al 1980, Suresh et al 1981) and roughness-induced closure (Ritchie and Suresh 1982) have been invoked to account for stress ratio influences and some microstructural and environmental effects which influence crack growth rates near the threshold. Other areas in which the closure concept has been successfully applied in explaining observed crack growth behaviour are physically-short fatigue cracks (where the limited crack wake reduces closure levels relative to long cracks, e.g. James and Smith 1983a) and the transient crack growth rate effects subsequent to a change in load amplitude (primarily due to changes in plasticity-induced closure, e.g. De Koning 1981). Note that roughness-induced closure has also been proposed to contribute to crack growth retardation subsequent to an overload (Suresh 1982).

Other mechanisms which reduce the effective stress intensity factor range, such as viscous fluid-induced and phase transformation-induced closure have been observed to operate in susceptible materials and environments (Suresh and Ritchie 1984).

Despite the clarification in certain fatigue phenomena afforded by factoring out crack closure from the applied load cycle, the whole topic of closure has been a vexed issue ever since Elber's papers were published. The plane strain versus plane stress controversy was only one point of contention, and others include:


2. Interpretation of compliance traces and determination of the most appropriate parameter, and way, to characterise closure values (Allison et al 1988, Carman et al 1988, Vasudevan et al 1992, Louat et al 1993). This is influenced by:

   - the necessity of understanding the way in which crack surface contact occurs and the effects on recorded traces of, say, unzipping-type behaviour compared with that due to discrete contact points;

   - the inherent elasticity of fracture surface contact points, which implies that a unique opening or closing point will not adequately characterise the effective range of stress intensity factor at a crack tip (Chenet al 1994, 1996). The effect would become more significant for materials with a lower modulus of elasticity. This provides an explanation for the observation that the parameter ΔK_{eff} often affords a poorer correlation of fatigue data in aluminum alloys, than in steels;

   - the nonlinearity which may observed in such traces even when the crack tip is nominally open (Robin et al 1977, Garz and James 1989). This is thought to be related to the use of bulk measurement techniques. Recorded data are then affected by starter-notch and crack tip plasticity, which may vary across the specimen thickness (because of the change from plane stress at the surface to a more constrained state in the interior) leading to differential opening states (Clerivet and Bathias 1979, Vasudevan et al 1992).

3. Closure values are sensitive not only to material and environmental parameters, but also to test techniques. This is particularly true for load shedding tests (Minkawara et al 1983; James and Knott 1985b) where the plastic wake is decreasing in size as the crack extends, which tends to produce closure well behind the crack tip. Meaningful comparisons between data obtained at different laboratories are then complicated.

4. Disagreement on the importance and magnitude of plasticity-induced closure between theoretical analyses based on dislocation models (Louat et al 1993) and finite element models (e.g. Blom and Holm 1992, McClung 1991).

It must be pointed out that even the name accorded to the crack surface contact phenomenon has undergone some evolution, with some authors favouring use of the term 'non-crack closure', on the basis that complete closure of the crack is prevented by the plastic wake and/or by asperity contact (Carlson and Beavers 1984). Lately, it has been recognised that closure mechanisms can be considered as crack tip shielding phenomena under the general heading of extrinsic toughening mechanisms (Ritchie and Yu 1986). In this sense, because crack closure reduces the effective crack tip driving force the material shows a greater resistance to crack growth.

A number of previous authors have addressed problems associated with fatigue crack closure (e.g. Jones et al 1978, Fleck and Smith 1982; James and Knott 1985a; Vecchio et al 1986; Hertzberg et al 1988; Allison et al 1988; Ray and Grandt 1988; Garz and James 1989). Despite these endeavours, and those of many other authors (Louat et al 1993 point out that more than 1000 papers and several symposia have considered crack closure and its implications since 1970), the complexity of the problem and continuing unresolved issues, which are often ignored in the literature, continue to limit our ability to deal in a rational predictive way with fatigue phenomena.

Even for relatively simple cases, such as the stress ratio effect in constant amplitude fatigue, no consensus exists as to the best predictive equation to use (Finney and Deirmendjian 1992). For the much more complex situation of variable amplitude loading, the waters are even more murky. Yet the usual industrial loading case involves random loading patterns; clearly, before real progress is achieved in predictive models one must be able to unambiguously separate crack wake effects from crack tip process zone influences. This will require a much better understanding of what is actually being measured when closure measurements are made, together with progress in resolving mechanistic arguments.

The purpose of the present paper is to highlight several important areas where, as noted by McClung (1991), a more concerted and collaborative research effort would yield dividends. Some comments will be made on a couple of the controversies related to closure and, as everyone who has worked in the area of crack closure has their own opinions and thoughts, it is hoped to re-stimulate some discussion on the topic.
MEASUREMENT POSITION AND RESOLUTION

As far back as 1978 (Jones et al) it was realised that the crack opening point observed on compliance traces (i.e. plots of crack opening displacement versus load), when defined in one of the usual ways (e.g. the onset of linearity is the upper part of the plot taken as corresponding to $K_{oc}$) was a function of measurement position. Data measured via a notch mouth clip gauge indicated lower closure values than obtained from near-tip gauges on the specimen surface. This is generally ascribed to differences between the surface (plane stress) and bulk (plane strain) responses to plastic deformation. If one assumes that the crack front in stable fatigue crack growth is largely self-regulating and self-similar, i.e. it cannot deviate too far from being a smooth curve and hence tends to ‘smooth’ out differences between surface and bulk effects, notch mouth clip gauge closure data should, in principle, characterise fatigue crack growth fairly well.

However, the ability of such techniques to provide ‘absolute’ values rather than merely indicate trends will depend on resolution of the technique and position in the crack wake where the major contribution to closure occurs. For instance if an artificial ‘asperity’ is introduced into a crack well behind the tip, the change in compliance caused by this can be readily detected (Garz and James 1989). If, on the other hand, a near-tip band of oxide was the major contributor to closure, the compliance technique requires a very high resolution of crack length changes to accurately monitor opening events in this crucial region. Theoretical studies of around 25 μm have been quoted for compliance measurements which incorporate offset displacement (Kikukawa et al 1976). The repeatability of such measurements is not clear, however (James and Knott 1985a - see Fig. 2), nor is the effect on the resolution of a significantly curved crack front.

With respect to measurement position, the results of a numerical finite element study of plasticity-induced closure are interesting (Blom and Holm 1992). Their work was aimed at critically assessing the ability of common compliance techniques (notch mouth clip gauge and back-face strain) to provide closure data matching those obtained from finite element analysis, i.e. where the closure point is defined as first contact of the node closest to the crack tip. The starting point of the paper was that although good agreement may be obtained between numerical results, a modified Dugdale-type analysis and experimental data, this may be largely fortuitous.

The rationale behind this statement derives from assumptions which are often implicit in numerical models, i.e. fracture surfaces which experience unzipping contact and have no microstructurally derived surface roughness. Plasticity-induced closure in such a planar crack can be calculated for a position corresponding to contact of the first node behind the crack tip. Experimental compliance techniques are unlikely to have a high enough resolution to detect such closure so close to the crack tip. Hence experimental results should generally be lower than the numerically calculated closure values. If agreement exists between data from numerical models of plasticity-induced closure and experimental results, Blom and Holm (1992) argue that this is a result of additional closure arising from other mechanisms. As the reality of crack growth is likely to be somewhat different to the assumptions made in a numerical analysis, this argument may well be valid under certain circumstances, and measured closure values may therefore reflect a component from surface roughness and discrete asperity contacts.

Blom and Holm obtained compliance curves numerically, at various node positions in their model (e.g. around the crack tip position, at the back-face of the specimen and at the notch mouth). Their results showed the following:

- The sensitivity of direct compliance curves is much less than that obtained using the offset (or reduced) compliance technique, and closure points are hard to determine from them.
- Compliance curves obtained from either a notch mouth clip gauge or back-face strain, are not sensitive enough to characterise pure plasticity-induced closure (see Fig. 3). Data should be measured close to the crack tip to obtain accurate assessments of closure. For the particular case of a compact tension specimen modelled by them, compliance data corresponding to the clip gauge position was found to be more accurate than that from back-face strain.

In contrast to these results which indicate that plasticity-induced closure is best measured close to the crack tip, Vasudevan et al (1992) suggest, in relation to an analytical dislocation-based model of closure, that displacements should be measured at the loading line on the specimen, i.e. via notch mouth clip gauge. This is to avoid spurious local closure indications due, for example, to asperity contact.

INTERPRETATION OF COMPLIANCE CURVES

The paper of Blom and Holm (1992) also makes the point that since there is no obvious definition of what constitutes the closure, or opening, point in a compliance plot (i.e. that point that characterises the full effective stress intensity range) the complete curve should be shown in any published work. The question of identifying that portion of the applied stress range below the nominal opening point which contributes to crack growth, has been considered by other authors (Shih and Wei 1974; Vecchio et al 1986, Vasudevan et al 1992; Chen et al 1994). Blom and Holm also presented data showing the increase in contact load across the fracture surfaces as a function of applied load level. This contact load may cause yielding of parts of the fracture surface (e.g. asperities or the near-tip region), and/or the unloading subsequent to fracture surface contact may allow additional reverse plasticity to occur in the crack tip process zone (Vasudevan et al 1992). In such a case the effective range of stress intensity factor would be greater than indicated by conventional $K_{oc}$ or $K_x$ values.

It seems to the present author that this question of defining $\Delta K_{oc}$ can best be resolved by further numerical analyses that explicitly consider the relationship between crack flank contact forces and the range/sign of stresses and strain in the crack tip plastic zone. Some experimental studies have addressed this problem by making the assumption that stress ratio effects on fatigue crack growth rate in a particular alloy, are solely a function of crack closure (e.g. Seetharam and Dash 1992). The effective range of stress intensity factor can then be found by collapsing growth rates curves obtained at low stress ratios onto the curve obtained at a high R value (which is taken as being closure-free). Closure levels can then be inferred from the resulting $\Delta K_{oc}$ values.

This technique suffers from two serious drawbacks, firstly crack growth rates are well known to be subject to inherent scatter (a factor of two is often quoted as being reasonable - Clark and Hudak 1975) and, secondly, fatigue crack growth rates in a number of materials are sensitive to stress ratio effects through $K_{max}$ in the fatigue cycle, as well as through crack closure (e.g. Li Wenfeng and James 1997). Indeed, it is likely to be generally true that crack growth rates, and the mechanism of cracking, will be influenced by both $\Delta K$ and $K_{max}$ in the applied cycle. In this respect, Vasudevan et al (1993) have claimed that near-threshold growth rate phenomena can be
Some Unresolved Issues with Fatigue Crack Closure

near-threshold at low R values). Closure values in vacuum were, however, of similar magnitude to those measured in air in the linear regime of crack growth (0.15 - 0.20). The higher 'intrinsic' threshold value, together with the different slope in the crack growth rate data observed in vacuum, were ascribed to increased slip reversibility, compared to that occurring in an air environment.

PART-THROUGH OR SURFACE FLAWS

As should now be clear, the closure behaviour of through-thickness cracks remains generally incompletely understood, although it is well known that there is a difference in plasticity-induced closure from the surface to the interior of the specimen. This three-dimensional aspect of closure behaviour has been considered for through-thickness cracks in a number of finite element studies, some of which are mentioned by Chermahini et al (1989).

The situation regarding closure information for part-through, or surface (thumbnail) cracks is even worse, as similar studies are scarce in the literature, despite the undoubted importance of such cracks in engineering structures and components. One reason may be that a major part of the appeal of fracture mechanics lies in the removal of geometry as a factor in fatigue crack growth, i.e. provided that an appropriate stress intensity calibration can be found, tests on through-thickness cracks (which are experimentally straightforward) should apply to other crack geometries. However, this assumes a similar influence of, and trends in, crack closure, which may not be the case. Chermahini et al (1993) have presented a three-dimensional finite element analysis of closure behaviour in semicircular and semi-elliptic cracks. This indicated that, for both crack shapes, the crack tip located on the surface plane closed first and opened last, although semi-elliptic defects showed more complex behaviour than semicircular. Closure levels were higher at the surface tip, by about 50%, than in the depth of the specimen.

Experimental data regarding closure of surface cracks is also relatively scarce, and most of the available information relates to small fatigue cracks (typically < 1-2 mm in depth). Thus measurement of closure and interpretation of what has been measured is even less certain for surface cracks than is the case for through-thickness cracks. The question of what is actually measured by compliance-based systems, such as laser interferometry across two indentsations at the centre of the crack (e.g. James and Sharpe, 1989), is yet to be answered. It is not clear whether the indicated closure level (bearing in mind that the most appropriate parameter for this point on a compliance trace is, as yet, unresolved) reflects an average of crack depth and surface opening, or is a direct reflection of opening at the maximum depth position.

This question has been addressed experimentally, albeit somewhat inconclusively, by James and Smith (1983b). In essence, surface fatigue cracks were grown naturally in four-point bend specimens to a surface length of about 3 mm (maximum depth of about 450 μm). The specimens were then thinned to a depth of 3.5 mm and loaded in a jig on the stage of an SEM. Displacements at the centre of the crack were photographed as the specimen was deflected by the same amount it experienced in the fatigue cycle. A relatively consistent change in compliance was noted at a deflection ratio (d/d_max) of about 0.32. Specimens were then infiltrated with heated epoxy resin under vacuum, at deflection ratios of 0.28, 0.37 and 0.46. These values were chosen to be 0.05 either side of the nominal opening point (to allow for scatter in the data), and well above the opening point. The cracks were then sectioned at their midpoint and examined in an SEM in an effort to determine when the tip opened. Although progressive opening of the crack was observed (Fig. 6), it was not clear whether the crack tip

explained in terms of a relationship between these two parameters, exclusive of any crack closure contribution. Further modelling work, linked with carefully designed experimental programmes, is required to determine the relative importance of ∆K, K_max and crack closure to fatigue crack growth in particular alloy systems.

RELEVANCE OF PLASTICITY-INDUCED CLOSURE

Analytical modelling of crack closure is clearly of importance to understanding the basic phenomenon. The model of Vasudevan et al (1992) is useful in providing insight into the shape of compliance curves (Fig. 4), and is a valuable starting point for discussion of the relevance of plasticity-induced closure to fatigue crack growth rates. They used displacation modelling to consider the cases both of a fatigue crack closing on an asperity, and that of closure arising as a consequence of plastic wake-induced residual stresses (Louat et al 1993). One important conclusion of their work is that displacements of crack surfaces, arising from plastic deformation at the crack tip, are incapable of causing closure under non-vanishing applied loads.

This view is refuted by a simple experiment performed on a thin sheet of polycarbonate by James and Knott (1985b). Referring to Fig. 5, the procedure was as follows: A notched compact tension specimen was subjected to a single loading cycle. A saw cut was then made from the notch root. A second loading and cutting sequence was then performed. The regions of plasticity which were induced under the applied load at the notch root, and at the tip of the first saw cut, are visible in Fig. 5 as a whitening of the specimen. Both cuts were made with the same saw, and it is clear that a very considerable reduction in width of cut has occurred in the plastic regions. Clearly, had either the width of the cut been smaller or the surface been rougher, crack closure would have been observed. Fatigue cracks seldom have planar fracture surfaces, and this inherent roughness is likely to influence plasticity-induced closure, even in the linear regime of crack growth, as well as giving rise to surface roughness-induced closure per se at near-threshold growth rates. This point is often overlooked in discussions of crack closure.

Vasudevan et al (1992) and Louat et al (1993) support their conclusion regarding the insignificance of plasticity-induced closure by considering fatigue crack growth in vacuum, where the oxidation contributions to closure would be absent. They argue that the contribution from plasticity-induced closure should then be more apparent, especially at high stress ratios. Their reasoning for this is that plasticity-induced closure (i.e. K_pl) should be a function of K_max, which increases nonlinearly with stress ratio. As observed threshold values of ∆K in vacuum are relatively constant over a range of R from 0.1 - 0.7, they deduce that crack closure due to plasticity is either non-existent or insignificant. However, this argument is not conclusive, as although plasticity-induced closure may scale approximately with K_pl (James and Knott 1985b), crack tip opening displacement (CTOD) scales with K_max. Equally, closure will only influence fatigue crack growth if K_pl > K_max in the applied loading; thus if K_max is large (as it would be at high R values), K_pl can be quite high but still not have any influence on crack growth rates.

For example, James and Knott (1985c) considered near-threshold crack growth in air and vacuum (~ 0.1 mPa). Closure values were monitored via back-face strain and notch mouth clip gauge, on four-point bend specimens of QIN (IY 80) steel. Growth rates in air were influenced markedly by stress ratio (0.2 ≤ R ≤ 0.5), while those in vacuum were little influenced. Closure values in vacuum were generally less than K_pl and did not show the same increase at near-threshold values of ∆K as was observed for the air data at R = 0.2. This was explained in terms of oxide-induced closure (which is maximised in an air environment when CTOD is small, i.e.
was completely open at $d/d_{\max} = 0.37$. At $d/d_{\max} = 0.48$, however, the tip was fully open.

As surface tip opening levels would be expected to be higher than those in the depth, these results were interpreted as indicating that the measured change in compliance did correspond fairly well to opening of the maximum depth position of the crack. Further work on this question would be of considerable value, particularly if numerical modelling was linked with a programme of experimental work.

CONCLUSIONS

From this discussion of some aspects of fatigue crack closure and, particularly, the analyses of Vasudevan et al. (1992) and Blom and Holm (1992), the net conclusion is that there is still considerable disagreement and confusion is to the nature, magnitude, measurement and interpretation of fatigue crack closure. Part of the problem, no doubt, arises because the models do not provide an accurate enough reflection of the physical reality of crack growth processes, and the effects of these on the fracture surface. Another factor must be the relatively poor communication amongst those members of the fatigue community who are interested in closure modelling and experimentation. The final factor, however, is that experimental work is seldom structured (for reasons of funding and time constraints) to consider the whole range of influences that may be relevant to a particular question.

The time would appear right for an e-mail round robin on the whole issue of fatigue crack closure. A carefully structured programme of collaborative numerical and experimental work could then be set-up which would be likely to shed light on a number of issues in fatigue crack closure that, currently, remain obscure.

REFERENCES


McClusky, R. C. (1991). The influence of applied stress, crack length, and stress intensity factor...


Fig. 1  Photoelastic experiment of Christensen (1963), demonstrating the effect of a trapped oxide particle. In this case, under identical compressive loading of the notch, the particle reduced the tangential boundary stress from approximately 25 MPa to 8.4 MPa.

Fig. 2  Illustration of resolution problem with compliance-based closure measurements. $K_{cm}/K_{max}$ values from three identical tests are shown as A, B, C. Fractography indicated identical thickened oxide layers near the crack tip, and crack growth only recommenced at $\Delta K > \Delta K_{cm}$. In two tests this effect has been detected as a higher closure value, whilst in the third test it has not.

Fig. 3  Offset, or reduced, compliance curves from the numerical model of Blom and Holm (1992). Nodes 49, 107 and 165 are close to the crack tip, while node 196 corresponds to the back-face of the specimen.

Fig. 4  Illustration from Vasudevan et al (1992) showing asperity induced closure with the effects of forward and reverse plasticity.
Fig. 5  Demonstration of the possibility of fatigue crack closure occurring from notch or crack tip plasticity. The specimen is polycarbonate, and both cuts were made with the same saw. The whitered regions correspond to plastic deformation.

a) $2c = 1012 \, \mu m$

b) $2c = 1100 \, \mu m$

c) $2c = 1063 \, \mu m$

200 x

Fig. 6  SEM photographs of similar size surface cracks, sectioned at the maximum depth after infiltration with heated epoxy resin at deflection ratios ($d/d_{max}$) of:

a) 0.28  
b) 0.37  
c) 0.46

Compliance data indicated an opening deflection ratio of approximately 0.32.