A REVIEW OF CRACK CLOSURE

AD-A142 273

S. Banerjee

University of Dayton
Research Institute
300 College Park Drive
Dayton, Ohio 45469

April 1984

Interim Report for Period August 1983 Through December 1983

Approved for public release; distribution unlimited

MATERIALS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

84 06 18 061
NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Office of Public Affairs (ASD/PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

THEODORE NICHOLAS, Project Engineer
Metals Behavior Branch
Metals and Ceramics Division

JOHN P. HENDERSON, Chief
Metals Behavior Branch
Metals and Ceramics Division

FOR THE COMMANDER

LAWRENCE N. HJELM, Asst. Chief
Metals and Ceramics Division
Materials Laboratory

"If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify AFWAL/MLN, W-PAPB, OH 45433 to help us maintain a current mailing list".

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.
A comprehensive review and critique of the literature on fatigue crack closure is presented. The elements of closure: its mechanisms, experimental procedures for its determination; the phenomenological study of its dependence on different variables, and methodologies for its prediction, are all discussed in detail. Suggestions for future work and interpretations of findings of other authors are presented.
FOREWORD

This Technical Report was prepared by the Aerospace Mechanics Division of the University of Dayton Research Institute for the Metals and Ceramics Division, Materials Laboratory. Dr. Theodore Nicholas, AFWAL/MLLN is the project engineer.

The work was supported, in part, under Contracts F33615-81-C-5015 with the University of Dayton and Contract F33615-82-C-5001 with Universal Energy Systems. This review was completed during the period August 1983 through December 1983, when the author worked at the AFWAL Materials Laboratory as a Senior Visiting Scientist on leave from the Indian Institute of Technology, Bombay, India.

Some of the ideas contained in this report evolved out of the several sessions of discussion the author had with Dr. Theodore Nicholas of AFWAL Materials Laboratory and it is a pleasure to acknowledge his input, suggestions and interest in the preparation of this report. The author also acknowledges Dr. Noel Ashbaugh of the University of Dayton Research Institute for his suggestions and final editing of the manuscript.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>SECTION</th>
<th>CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
</tr>
<tr>
<td>Closure and Its Origin</td>
<td>1</td>
</tr>
<tr>
<td>Role of Closure in Crack Growth Rate Prediction</td>
<td>7</td>
</tr>
<tr>
<td>Determination of Closure</td>
<td>10</td>
</tr>
<tr>
<td>Closure in Plane Stress and Plane Strain Conditions</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>MECHANISMS OF CLOSURE</td>
</tr>
<tr>
<td>2.1</td>
<td>PLASTICITY INDUCED CLOSURE</td>
</tr>
<tr>
<td>2.2</td>
<td>ASPERITY INDUCED CLOSURE</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Single Asperity Model [8]</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Spring Clip Model [60]</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Fracture Surface Roughness Model [12]</td>
</tr>
<tr>
<td>2.3</td>
<td>OXIDE INDUCED CLOSURE</td>
</tr>
<tr>
<td>2.4</td>
<td>COMMENTS ON THE MECHANISMS OF CLOSURE</td>
</tr>
<tr>
<td>3</td>
<td>EXPERIMENTAL DETERMINATION OF CLOSURE</td>
</tr>
<tr>
<td>3.1</td>
<td>THICKNESS AVERAGED BULK AND LOCAL CLOSURE BEHAVIOUR</td>
</tr>
<tr>
<td>3.1.1</td>
<td>CMOD Gage</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Strain Gage</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Ultrasonics</td>
</tr>
<tr>
<td>3.1.4</td>
<td>Potential Difference</td>
</tr>
<tr>
<td>3.1.5</td>
<td>Special Displacement Gages</td>
</tr>
<tr>
<td>3.2</td>
<td>NEAR TIP SURFACE CLOSURE BEHAVIOUR</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Interferometric Displacement Gage (IDG)</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Direct Observation Using SEM</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Optical Interferometry</td>
</tr>
<tr>
<td>3.3</td>
<td>NEAR TIP INTERIOR CLOSURE BEHAVIOUR</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Closure Measurement Before and After the Removal of Successive Surface Layers</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Push Rod Displacement Gage</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Vacuum Infiltration Technique</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Interferometric Technique in Transparent Specimen</td>
</tr>
<tr>
<td>3.4</td>
<td>COMMENTS ON THE DETERMINATION OF CLOSURE</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (Concluded)

<table>
<thead>
<tr>
<th>SECTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>PHENOMENOLOGICAL STUDY OF CLOSURE</td>
</tr>
<tr>
<td>4.1</td>
<td>EFFECT OF $K_{\text{max}}$, $K_{\text{min}}$, and $R$ UNDER CONSTANT AMPLITUDE LOADING</td>
</tr>
<tr>
<td>4.1.1</td>
<td>The Effect of $R$ and $K_{\text{min}}$</td>
</tr>
<tr>
<td>4.1.2</td>
<td>The Effect of $K_{\text{max}}$</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Normalization of $da/dN$ Data Using $K_{op}$</td>
</tr>
<tr>
<td>4.2</td>
<td>OVERLOAD EFFECTS [27,32,36,55,65,80-94]</td>
</tr>
<tr>
<td>4.3</td>
<td>SHORT CRACK BEHAVIOUR</td>
</tr>
<tr>
<td>4.4</td>
<td>SURFACE CRACK BEHAVIOUR</td>
</tr>
<tr>
<td>4.5</td>
<td>EFFECT OF RESIDUAL STRESS</td>
</tr>
<tr>
<td>4.6</td>
<td>ENVIRONMENTAL FACTORS</td>
</tr>
<tr>
<td>4.7</td>
<td>MICROSTRUCTURAL AND FRACTOGRAPHIC FEATURES</td>
</tr>
<tr>
<td>4.8</td>
<td>MATERIAL PROPERTIES</td>
</tr>
<tr>
<td>4.9</td>
<td>EFFECT OF SIZE AND GEOMETRY</td>
</tr>
<tr>
<td>4.10</td>
<td>COMMENTS ON THE PHENOMENOLOGICAL STUDY OF CLOSURE</td>
</tr>
<tr>
<td>5</td>
<td>PREDICTION OF CLOSURE</td>
</tr>
<tr>
<td>5.1</td>
<td>ANALYTICAL PROCEDURES</td>
</tr>
<tr>
<td>5.2</td>
<td>FINITE ELEMENT BASED METHOD</td>
</tr>
<tr>
<td>5.3</td>
<td>COMMENTS ON PREDICTION OF CLOSURE</td>
</tr>
<tr>
<td>6</td>
<td>CONCLUDING REMARKS</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>152</td>
</tr>
<tr>
<td>FIGURE</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>29</td>
</tr>
<tr>
<td>10</td>
<td>31</td>
</tr>
<tr>
<td>11</td>
<td>37</td>
</tr>
<tr>
<td>12</td>
<td>42</td>
</tr>
<tr>
<td>FIGURE</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>13</td>
<td>Basis of Oxide Induced Closure - A Schematic Effect of Environment on $\Delta K_O$ and $da/dN$ [22].</td>
</tr>
<tr>
<td>14</td>
<td>Basis of Oxide Induced Closure. Typical Variation of Excess Oxide Thickness, $d$, with Crack Length and Crack Propagation Rate [22].</td>
</tr>
<tr>
<td>15</td>
<td>Identification of Closure from Load Versus Displacement (P-V) and Load Versus Offset Displacement (P-AV) Plots. $K_q$ Represents the Minimum Stress Intensity Factor at Which the Plots are Linear and $P_{ct}$ is the Stress Intensity at Which the Two Tangents on the P-V Plot Intersect.</td>
</tr>
<tr>
<td>16</td>
<td>Schematic of the Typical P-AV Test Records. Friction and Misalignment may Change the Test Records from the Type Reported in Figure 16a to that in Figure 16b.</td>
</tr>
<tr>
<td>17</td>
<td>Preferred Locations of the Strain Gage in a CT Specimen for the Determination of Closure Using Strain-Gage Technique.</td>
</tr>
<tr>
<td>18</td>
<td>Determination of Closure Using Ultrasonic Technique [41]. $P_g$ and $P_{ct}$ are Explained in Section 3.1.1.</td>
</tr>
<tr>
<td>19</td>
<td>Determination of Closure Using DC Potential Drop Technique. 1-1 and 2-2 - Potential Probe Position and 3-3 - Current Probe Position [3].</td>
</tr>
<tr>
<td>20</td>
<td>a - Variation of Closure Stress with Removal of Successive Surface Layers, b - The Effect of Specimen Thickness on Closure [2].</td>
</tr>
<tr>
<td>21</td>
<td>Push Rod Displacement Gage Technique for the Determination of Near Tip Interior Closure of an Elliptical Crack [65]. Note that the Technique can also be Used for a Through-The-Thickness Crack.</td>
</tr>
<tr>
<td>22</td>
<td>Effect of Load Ratio, $R$, on $da/dN$ [2].</td>
</tr>
<tr>
<td>23</td>
<td>Variation of $K_{op}/K_{max}$ with $a/W$ or $K_{max}$ for (a) $R=0.5$ and (b) $R=0.1$ [32].</td>
</tr>
<tr>
<td>24</td>
<td>$K_{op}$ Has No Systematic Dependence on $K_{max}$. Each Data Point From Individual CT Specimens of $B = 4$ mm, $W = 50$ mm, and Varying $a/W$ Values [31,34].</td>
</tr>
</tbody>
</table>
### LIST OF ILLUSTRATIONS (Concluded)

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Decreasing ( \frac{K_{op}}{K_{max}} ) With Increasing ( K_{max} ). The Unusual Behaviour of 2219 T87 is Attributed to Shear Lip Formation [18].</td>
</tr>
<tr>
<td>26</td>
<td>Plot of ( \frac{da}{dN} ) Versus ( \Delta K_{eff} ), Showing the Normalization of the Effect of R on ( \frac{da}{dN} ) as Reported in Figure 22 [2].</td>
</tr>
<tr>
<td>27</td>
<td>Variation of ( \frac{da}{dN} ) as a Result of Single Cycle Overload. The Crack Growth Rate at the Overload Cycle is not Represented in the Figure.</td>
</tr>
<tr>
<td>28</td>
<td>Experimental Crack Propagation Behaviour as Influenced by Overload Application - Comparison with Prediction by Various Models. The Inset Gives Details of Overload Application [84].</td>
</tr>
<tr>
<td>29</td>
<td>Crack Growth Rate Versus ( \Delta K_{eff} ) Before and During Transient Effects After a Single Overload [32].</td>
</tr>
<tr>
<td>30</td>
<td>Comparison of Experimentally Observed ( \frac{da}{dN} ) Following Overload with the ( \frac{da}{dN} ) Inferred From Closure Measurement and Constant Amplitude ( \frac{da}{dN} ) Data for a Thumbnail-Shaped Crack with ( a=)Crack Depth and ( c=)Surface Crack Length [65].</td>
</tr>
<tr>
<td>31</td>
<td>Comparison of Experimental and Predicted Crack Growth Rates for Small Cracks Emanating From a Circular Hole in Steel Specimens [101].</td>
</tr>
<tr>
<td>32</td>
<td>Comparison of ( \frac{da}{dN} ) for Non-Stress Relieved and Stress Relieved Materials. The Agreement is Good when ( \Delta K_{eff} ) is Used for Representing ( \frac{da}{dN} ) Data [110].</td>
</tr>
<tr>
<td>33</td>
<td>a - Effect of Environment on Crack Propagation in 7075-T651 Al Alloy, b - Normalization of the Data Reported in Figure 33a When ( \frac{da}{dN} ) is Plotted Against ( \Delta K_{eff} ) [39].</td>
</tr>
<tr>
<td>34</td>
<td>Effect of ( \sigma_Y ) on ( K_{op} ) in CT Specimens, ( B=4 ) mm, ( W=50 ) mm and Varying ( a/W ). Scatter in ( K_{op} ) Values Reported is Attributed to Different Precracking History [31,34].</td>
</tr>
<tr>
<td>35</td>
<td>Fracture Surface Profile as Influenced by Specimen Size and Loading (a) Symmetric Profile, (b) Non-Symmetric Profile, and (c) Out of Plane Sliding in Thin Specimens with Non-Symmetric Profile Relaxes Compressive Force [89].</td>
</tr>
<tr>
<td>36</td>
<td>Displacement and Stress Along the Crack Line of a CCP Specimen at the Maximum and the Minimum Load.</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

In a recent paper [1] titled 'Twenty Years of Reflections on Questions Concerning Fatigue Crack Growth; Historical Observations and Perspectives', Paris views 'crack closure as raising the most central question still to be resolved'. He further observes that 'crack closure is a key physical phenomenon in fatigue cracking process' particularly with reference to problem areas such as: variable amplitude load interaction effects, short cracks, threshold fatigue crack growth, and environmental effects. A review of crack closure in fatigue hardly needs a better justification.

The study of closure has been reviewed recently [2,3,4]; however, these reviews deal with only one or some of the aspects of closure.

The main purpose of this report is to review closure - its mechanisms, procedures for its determination, the phenomenological study of its dependence on the different variables, and the methodology of its prediction. As outlined below, the task is rather complex since it is difficult to provide answers to even the obvious questions concerning closure. What is attempted in this review is to define the positions and raise the critical questions concerning the above aspects of closure; hopefully, such an exercise would indicate meaningful direction of future work in closure studies.

Closure and Its Origin

The concept of crack closure was first proposed by Elber [5,6] in 1970. According to him, a fatigue crack in a body subjected to tension-tension
cyclic loading, is completely open only at high load levels; or in other words, at low load levels a part of the crack near the tip remains closed during the loading as well as the unloading phase of the cycle. The basic scheme of Elber's closure and the definition of the terms such as, $K_{op}$ and $K_{clo}$ are given in Figure 1.

The closure behaviour of a precracked body has three different aspects - $K_{op}$ value, extent of closure, and residual strain due to closure. The three aspects are interrelated but they do not always change in an identical manner in response to a change in a given fatigue loading situation.

Among these three aspects, the $K_{op}$ value is determined, studied, and calculated most often since it can be directly used for life prediction. $K_{op}$ values can be identified from a simple plot of load versus crack mouth opening displacement (see Fig. 2a), provided the closure in the precracked body is extensive. The ratio of $K_{op}/K_{max}$ can have a value anywhere between 0 and 1. A value of zero corresponds to no closure while a value of 1 indicates closure during entire fatigue cycle.

As regards the second aspect, closure is defined as extensive if a large fraction of the crack, $(\Delta a/a)$, remains closed at the start of loading, where $a$ is the physical crack length and $a_o$ is the crack length measured from compliance at the start of loading and $\Delta a = a - a_o$. In a load versus displacement plot, the difference between the slope at the start of loading and the slope at a load higher than the closure load can be related to and, therefore, is obviously a measure of the extent of closure. The ratio of $\Delta a/a$ could have values between 0 and 1.
Figure 1. Scheme of Fatigue Crack Closure. $P$ is the applied load and $K$ is stress intensity factor.

- $K_{op}$ = Minimum stress intensity at which the crack is fully open during loading.
- $K_{clo}$ = Stress intensity at which the crack starts closing during unloading.
Figure 2. Identification of $K_{op}$ and $K_{clo}$ from the Load Versus Displacement Plot (Figure 2a) and Load Versus Offset Displacement Plot (Figure 2b). $P$ = Load, $V$ = Displacement, and $\Delta V$ = Offset.
If the closure is not so extensive, that is \( \Delta a/a \) has a small value, one can use a plot of load versus offset displacement to identify \( K_{op} \) as shown in Figure 2b; note that the plot in Figure 2b is derived from the data reported in Figure 2a. Alternatively, one can use a more sensitive instrumentation or data reduction technique to determine \( K_{op} \).

The third aspect, that is the residual strain due to closure, is produced due to plastic flow at the crack tip during cyclic loading. This, in turn, leaves a strip of yielded material behind the crack tip, referred to as plastic wake. Elber proposed that the residual strain is produced by extension of the material within the plastic wake behind the crack tip. At zero load, the extended material in the plastic wake has to be accommodated by the rest of the precracked body which is elastic. As a result, residual compressive stresses are set up over the crack faces, and therefore, a part of the crack near its tip remains closed at low load levels. The usual argument is since there is no singularity at the tip of a closed crack, it cannot grow until it is fully open and this happens when \( K > K_{op} \). However, the closure has some more basic consequences. The introduction of residual compressive stresses and also displacement in the wake of the crack alters the state of stress, strain, and displacement near the tip of the growing fatigue crack. This, in turn, will decrease the size of the monotonic, as well as the reverse plastic zones to values which are significantly less than conventionally assumed.

Elber's proposed concept is referred to as plasticity induced closure. Investigations of the closure behaviours such as, extent of closure and the residual strain due to closure are rather limited. But such investigations are essential for understanding the mechanism of closure and for the prediction of closure.
Elber [5] had originally proposed that $K_{\text{clo}}$ should be greater than $K_{\text{op}}$. However, the difference between $K_{\text{clo}}$ and $K_{\text{op}}$ is less than the scatter observed in most experimental determinations of closure and for all practical purpose, one normally assumes $K_{\text{clo}} = K_{\text{op}}$ (see Figure 1). In all subsequent discussions, the symbol $K_{\text{op}}$ would be used to represent the closure behaviour.

In general, unless a distinction between $K_{\text{clo}}$ and $K_{\text{op}}$ is warranted.

Elber had initially observed [5] that at very low load levels, the precracked body has a compliance which is identical to that of an uncracked body. On the other hand, this is not always true and many experimental compliance measurements show that a part of the crack can start opening as soon as the load is applied.

In addition to Elber's mechanism of plasticity induced closure, two other mechanisms of closure have recently been proposed: asperity induced closure [7-12] and oxide induced closure [13,14]. As discussed below, both of these mechanisms are based on a concept which can be more appropriately termed 'non-closure'. Furthermore, contrary to Elber's mechanism, $K_{\text{clo}}$ should be less than $K_{\text{op}}$ according to these mechanisms.

According to the asperity induced closure mechanism, the asperities on the two mating fracture surfaces interfere and keep the crack propped open even when the load is zero. On the other hand, according to the oxide induced closure mechanism, the formation of an oxide layer just behind the crack front wedges open the crack faces and keeps the crack open, even when the load is zero. It is obvious from these physical pictures that the concept of non-closure can produce only residual tensile stresses at the crack tip when the
external load is removed. It is interesting to note that even though the
residual stress patterns near the crack tip, arising from closure and non-
closure are opposite of each other, identical experimental techniques (see
Figure 2) are used to identify the closure event, in both cases.

It is also rather surprising that even though the two different
concepts postulate opposite patterns of crack tip residual stress at closure,
the $K_{op}$ derived from these two concepts is presumed to play an identical role
in decreasing the fatigue crack growth rate. It shows how little we under-
stand the manner in which closure decreases fatigue crack growth (FCGR) and
the manner in which $\Delta K$ produces fatigue crack extension. The mechanisms of
closure are discussed in greater detail in the next chapter.

Role of Closure in Crack Growth Rate Reduction

The concept of $K_{clo}$ has been proposed mainly to achieve a more
reliable fatigue crack growth rate (FCGR). In order to predict life of a
component, several empirical laws have been proposed to characterize FCGR but
all these are either derived from or are a variation of one form or another,
of the law proposed first by Paris and Erdogan [16]. Paris-Erdogan law is
represented by the equation

$$\frac{da}{dN} = A \Delta K^n$$  \hspace{1cm} (1)

where $A$ and $n$ depend on the material and if these are known, one can
separate variables and integrate to determine $N$, the component life. Since
the above relationship is independent of crack size and specimen geometry, the constants $A$ and $n$ when determined from a particular test specimen, can be used to calculate the fatigue life of any precracked body. Equation (1) applies only to constant amplitude loading. The equation fails to characterize fatigue crack growth rate if:

1. The ratio of minimum to the maximum load, $R$, in the test specimen and the component differs.

2. The crack length in the component is very short [15].

3. The loading amplitude varies and has hi-lo load sequence leading to load interaction and crack growth retardation or acceleration.

The concept of closure has been used to account for the effect of these factors on fatigue life. Accordingly, one defines an effective $\Delta K$ which is given by

$$\Delta K_{\text{eff}} = K_{\text{max}} - K_{\text{op}}$$

Elber proposed that crack growth rates are determined by $\Delta K_{\text{eff}}$ and, accordingly, Equation (1) should be appropriately modified by substituting $\Delta K_{\text{eff}}$ for $\Delta K$. It could then be used for the calculation of life of a component.
In fact, Equation (1), when modified in a manner as described above, has been shown to account for the effect of the above three factors on the FCGR in some instances [1,2,5]. Furthermore, $K_{op}$ as influenced by residual stresses, microstructure, and environment, has also been used to account for the effect of these variables on FCGR and life. Thus, a scatter band of 2, usually observed in a log-log representation of FCGR, could probably decrease if $K_{op}$ is taken into account in representing FCGR.

Since closure depends on the formation of plastic zone and inelastic deformation in the vicinity of the crack tip, the closure could also be influenced by specimen size and geometry. As a result, the $K_{op}$ observed in a test specimen could differ from that in another cracked body even though $K$, $R$, and the uniaxial yield strength, $\sigma_y$, are identical. Since reliable FCGR prediction of a component depends on $K_{op}$, the $K_{op}$ in a precracked body must be known. Experimental determination of $K_{op}$ for all cracked body geometries would require extensive testing which is not generally feasible. However, a systematic evaluation of the effect of size and geometry on closure would be invaluable in developing a procedure for calculation of $K_{op}$ of any cracked body. Unfortunately, such evaluation has not yet been performed. In formulating such a procedure, it would also be essential to ascertain the mechanism of closure, the nature of residual stress distribution that a given mechanism produces at the crack tip and elsewhere in the specimen, and the manner in which such stress distribution decreases crack growth rate, $da/dN$. One could then propose and validate models of closure for the calculation of $K_{op}$ in an arbitrary precracked body.
A phenomenological study of the different factors which influence both $K_{op}$ and $da/dN$ can provide the appropriate basis to ascertain the above. The several factors include $K_{max}$, $K_{min}$, $R$, short crack behaviour, step overload, surface crack behaviour, thickness, $W$, $a/W$, specimen geometry, microstructural and fractographic features, environment, residual stresses, elastic and flow properties of the material, and loading history. However, the reported results of phenomenological studies on the effect of even the more important ones amongst these variables, are quite confusing. For example, an examination of Elber's plasticity induced closure mechanism indicates that $K_{op}$ should systematically depend on $K_{max}$ if all other factors are identical. On the other hand, as would be discussed later (Section 4.1), some of the investigators report that experimental $K_{op}$ increases with $K_{max}$, while others report that $K_{op}$ is independent of $K_{max}$; and yet still others report that $K_{op}$ decreases with $K_{max}$! One must interpret with caution the results of such phenomenological studies undertaken in the past, since $K_{op}$ can depend very significantly on a host of different factors including the experimental technique used for its determination.

**Determination of Closure**

As discussed earlier, Figure 2a shows that the closure load can be determined experimentally from a load versus CMOD test record. For better resolution, one can identify closure from a plot of load versus offset displacement (see Figure 2b). The offset displacement can be obtained through the use of a simple operational amplifier circuit or by processing the load-CMOD test record numerically in a microprocessor. Besides the
CMOD-based approach, there are other experimental procedures for closure determination which are based on techniques using strain gage, ultrasonics, electrical potential, laser/optical interferometry, and optical methods including replication. The agreement amongst the different methods is not always good and the reason will be evident from the following discussion.

Closure can be identified from measurements made either near the crack tip or at points relatively far away from the crack tip. The measurements made either at points far away from the crack tip or those which involve the whole specimen cross section give a thickness-averaged global $K_{op}$ of the precracked body. On the other hand, the $K_{op}$ determined through the measurement of displacement near the crack tip are influenced significantly by the complex three-dimensional residual stress pattern localized near the crack front due to either preferential through-the-thickness yielding in the surface layers and/or due to crack tunneling. As a result, the $K_{op}$ at the interior (say mid thickness) of a specimen near the crack tip can be quite different from the $K_{op}$ observed at the near tip surface. In most instances, closure determined near the crack tip is based on surface measurement. Obviously, $K_{op}$ values obtained from all three different measurements can be different.

Techniques such as CMOD gage, back face strain gage (BFS), electrical potential, and ultrasonics give a thickness-averaged global $K_{op}$ value. The Elber gage and near tip strain gage in some instances, and the interferometric technique, and the optical and SEM technique including surface replication can give near tip $K_{op}$ values at the specimen surface. Only a few investigators
(see Section 3.3) have determined near tip interior closure behaviour; techniques such as the optical interferometry in transparent specimens, the push-rod clip gage technique, and the measurement of closure before and after the removal of surface layers have been used for this purpose.

Indeed, only a few investigators make a distinction between these different $K_{op}$ values and the question as to which one of these closures should be used to arrive at $\Delta K_{eff}$ as per Equation (2), has not been properly addressed. The consequences of the lack of such distinctions are contradictory phenomenological observations and multiplicity of the proposed mechanisms of closure.

Closure in Plane Stress and Plane Strain Conditions

Plane stress and plane strain are convenient assumptions for analytical formulation of fracture problems. However, these terms have limited practical relevance in defining the state of stress in a precracked specimen which is loaded. The plane stress condition exists only at the surface of the specimen; the rest of the interior cross section is neither plane stress nor plane strain. The actual state of stress at the interior is intermediate or three-dimensional. This is especially true since a plastic zone forms at the crack tip and grows as $K/\sigma_y$ increases. The growth of the plastic zone relaxes the stresses and therefore, the state of stress becomes closer to plane stress as $K/\sigma_y$ increases. Besides this, the terms such as plane strain or plane stress are often used by investigators of closure phenomenon with the tacit assumption that increasing specimen thickness strongly decreases plastic zone sizes while the specimen width has only a
mild effect on plastic zone size. Both of these assumptions are incorrect [17]. Therefore, the interpretation, correlation, and comparison of results based on such assumptions are questionable. While examining the reported experimental results in this review, whenever possible, the use of concepts such as plane stress and plane strain will be avoided. Indeed, the different experimental closure situations can be better classified and compared in terms of extent of closure and extent of plasticity.

In the next section, the proposed mechanisms of closure are examined. In Section 3, the different approaches to closure determination are reviewed very briefly to better evaluate the significance of closure related phenomenological observations as discussed in Section 4. Section 5 briefly outlines the proposed methodology for the prediction of closure. To reiterate the important points, comments are made at the end of these respective sections. Also, the concluding remarks summarizes the major points made in the review.
2. MECHANISMS OF CLOSURE

Three mechanisms have been proposed for the closure of a fatigue crack: plasticity induced closure [5,6], asperity induced closure [7-12,60], and oxide induced closure [13,14,21]. The plasticity induced mechanism can be truly termed closure since according to such a mechanism, a fatigue crack near the tip remains closed even when the external load is tensile. The other two mechanisms can be more appropriately termed 'non-closure' since, according to the basic scheme of these mechanisms, a fatigue crack fails to close near the tip when the external load is zero, or even mildly compressive.

A study of the mechanism of closure is important due to two reasons: First, the prediction of closure in a cracked component of arbitrary size, geometry, and loading is based on a given closure model. The closure model, in its turn, is based on a proposed mechanism of closure. Thus, one must first know the mechanism in order to develop a procedure for the prediction of closure. Second, the study of closure mechanisms helps us to identify and explain the role closure plays in determining da/dN and thus helps us to develop materials with better fatigue resistance. It helps us to ascertain the validity of a given closure mechanism or a model. It also helps us to answer more basic questions. For example, it is presumed that fatigue crack cannot grow as long as it remains closed. However, could not the strain intensification ahead of a closed crack cyclically loaded to a $K_{max}$ level which even if less than $K_{op}$, be enough to promote plastic flow at the crack tip and therefore cause growth?
A closure mechanism has to be validated through experimental observations. Any proposed mechanism of closure should be consistent with all the phenomenological observation concerning the effect of the different variables on closure. Such observations are presented and discussed later in Section 4. However, reference will be made in the present section to a few of these selected observations to comment on the validity of the different mechanisms.

Presented in this section is an examination of the different mechanisms of closure and their validity and limitations.

2.1 PLASTICITY INDUCED CLOSURE

Elber [6] proposed that during fatigue crack propagation, a zone of residual tensile deformation is left in the wake of a moving crack tip (see Figure 3). The residual tensile deformation is produced due to the envelope of plastic zone (see Figure 3) which is referred to as the 'plastic wake'. Note that the plastic wake increasingly spreads in the y direction as the crack grows since $K_{\text{max}}$ increases with crack growth for most test geometries under constant amplitude loading. At zero external load, the material within the plastic wake continues to remain extended and can no longer be accommodated within the surrounding elastic field without producing a corresponding strain mismatch. Thus, in the unloaded condition, this strain mismatch will produce residual compressive stresses over the plastic wake (see Figure 3) which, in turn, transmits the compressive stresses normal to the crack surface and thus keeping the two crack surfaces pressed and closed together.
Figure 3. Plastic Wake and Residual Compressive Stresses Developed on a Growing Fatigue Crack During a Constant Amplitude Cyclic Load Control Test.
The plastic wake behind the crack front is produced due to the formation of the monotonic plastic zone ahead of the crack front during the tensile phase of the cyclic loading. This zone extends incrementally as the fatigue crack grows. However, on unloading, the monotonic plastic zone experiences compressive stresses for the reason mentioned above and this causes yielding in compression over a distance where the compressive stress exceeds the yield strength. This is referred to as reverse plasticity and the plastic zone formed as a result is referred to as cyclic or reverse plastic zone. The effects of plastic wake and the reverse plasticity can be distinguished by their respective residual compressive stress patterns in the two examples as discussed below.

When a sharp saw-cut crack (not grown by fatigue) is subjected to one cycle of loading, monotonic and cyclic plastic zones form and the residual compressive stress pattern produced is shown in Figure 4. Note that unlike in Figure 3, the residual compressive stresses behind the crack front is absent in Figure 4. As a result of Bauschinger effect, the stress increment to cause yielding during unloading or reversed plasticity, is twice the yield strength observed during monotonic loading. Since plastic zone size is inversely proportional to the square of yield strength, the reverse or cyclic plastic zone size is approximately one fourth the monotonic plastic zone. The presence of the reverse plastic zone modifies somewhat the pattern of residual compressive stress in the monotonic plastic zone.

On the other hand, in the case of a growing sharp fatigue crack, in addition to the monotonic and reverse plastic zone, a plastic wake (of the type shown in Figure 3) forms along the cracked part of the specimen. Therefore, the residual stresses in the plane of the crack in CCP and CT
Figure 4. Plastic Zone and Residual Compressive Stresses Developed on a Saw Cut Sharp Crack During a Constant Amplitude Cyclic Load Control Test.
specimens are expected to exhibit patterns as shown in Figure 5. The plasticity induced residual stress pattern in the ligament of a compact tension specimen is not known and is expected to be rather complex. However, unlike in the case of a centre cracked panel, the residual stresses at the back face of a compact tension specimen should be compressive. This is discussed further in Section 4.5. One can surmise that the magnitude and distribution of residual compressive stress, depends on the distribution of $\varepsilon_{yy}$ along the y direction at different points along the length of the wake and the distribution of displacement in the y direction at various points along the length of the crack. Obviously, if the closure were to be produced by the compressive stress over the whole length of the wake and the monotonic plastic zone, one should then take into account the residual stress distribution both ahead and behind the crack tip, as reported in Figure 5.

The two above sources of stress distribution can be examined separately. The residual compressive stress pattern produced by the plastic wake would probably influence the bulk closure behaviour and cause an extent of closure which is large and easily detectable. On the other hand, the effect of compressive residual stress within the plastic zone would be localized near the crack tip, cause an extent of closure which is small, and therefore can be detected only if the method of determination is highly sensitive. During constant amplitude loading, the residual stress distribution over the length of the wake as well as the plastic zone is expected to be smooth and continuous and, therefore, produce only one single closure load which represents the bulk as well as the local behaviour. However, one can speculate that a large single overload cycle may produce a
Figure 5. Residual Stresses Developed in the Plane of Crack in CCP and CT Specimens Due to Plastic Wake Formed on Growing Fatigue Cracks, As Postulated in Plasticity Induced Closure.
discontinuity in the residual stress pattern near the crack tip, and as a result, produce two closures - one representing the bulk and the other representing the local behaviour. This will be discussed in Section 4.2.

The residual stress pattern along the width direction as shown in Figure 5 is the primary factor which controls crack closure, and, therefore, plasticity induced crack closure models are based on such stress distribution. However, the residual stress distribution along the thickness direction can also influence closure.

The residual stress distribution along the thickness direction can be produced due to two factors: preferential yielding in the surface layers and presence of a curved crack front.

It is well known that in a relatively thick specimen, the yielding in the surface layers (see areas marked A in Figure 6) is more pronounced as compared to that at the interior. As a result, the dimension of the plastic wake in the y-direction in the surface layers is larger than that at the interior. If we assume that the residual stress pattern of the type given in Figure 5 were absent, one can then surmise that higher stretching of the yielded material along the y direction within the surface layer can introduce residual tensile stresses in the interior and correspondingly, reactive compressive stresses are introduced in the surface layer. The nature of such stress distribution is shown in Figure 7. In an actual specimen, the two stress patterns reported in Figures 5 and 7 are superimposed on each other. Thus, it is likely that the last fraction of opening during unloading or the
Figure 6. Preferential Through-The-Thickness Yielding in the Surface Layers of a CT Specimen with a Straight Crack Front.
Figure 7. Residual Stress Developed Across the Thickness (See Section 2.2 on Figure 6) Due to Preferential Yielding in the Surface Layers. The Effect of Residual Stress Reported in Figure 5 is not Taken into Account.
first fraction of closure during unloading could be influenced by the presence of the pattern of residual stress distribution along the thickness direction as reported in Figure 7. As a result, in the case of a thick specimen, the last fraction of opening in the surface layers would be difficult compared to that observed at the interior. In a relatively thinner specimen, the strains $\varepsilon_{yy}$ and $\varepsilon_{xx}$ are probably uniform across the thickness direction and the variation of residual stresses along the thickness direction may be negligible.

The presence of a curved crack front can cause preferential extension and plastic flow in the surface layers ahead of the crack front (see areas marked A in Figure 8). This may introduce a pattern of residual stress distribution along the thickness direction similar to that reported in Figure 7 and produce similar effects as discussed above. However, such an effect would be produced both in the thick and the thin specimen, as long as the depth of curved crack front is comparable to the thickness and is not negligible compared to the crack length. Obviously, a large crack length is preferable if the effect of curved crack front is to be ignored. The specimen then has to be correspondingly wide to accommodate a large crack length.

It is obvious from the above discussion that the analytical results obtained from closure models based on plasticity induced mechanism, are expected to agree with experimental closure data provided the specimen is thin and the depth of curved crack front is small. On the other hand, experimental results obtained from a thick specimen with a pronounced curved crack front, are not expected to agree with the analytical results based on
Figure 8. Preferential Through-The-Thickness Yielding in the Surface Layers of a CT Specimen With a Curved Crack Front (Compare with Figure 6).
plasticity induced closure. In fact, in the thick specimen with a curved crack front, the closure at the interior is expected to differ from that observed at the surface if plasticity induced mechanism were operative.

It is often stated [7-12] that Elber's plasticity induced closure mechanism is operative only in plane stress. A true state of plane stress does not exist anywhere in the specimen except at its surface. In that sense, the experimental condition at which the plasticity induced mechanism is applicable is not clearly defined. By plane stress, one implies that the $K$-value and the plastic zone sizes are large while the $\sigma_Y$ and thickness of the specimen are low. On the other hand, it has been shown [17] that the state of stress in a thin specimen at low $K/\sigma_Y$ value need not be plane stress. In fact, in a thin specimen, the state of stress can be quite close to plane strain if $K/\sigma_Y \sqrt{W}$ is small and the width is large [17]. In a thin specimen ($B << W$ or $a$), the whole thickness can be assumed to be approximately in the same state of stress. It has been suggested that probably $\varepsilon_{yy}$ and $\varepsilon_{xx}$ are the same across the thickness of a thin specimen [2]. Thus, the advantages of using a thin and wide specimen for closure studies is obvious.

It is often assumed without reliable experimental foundation, that $K_{op}$ produced by the plasticity induced closure mechanism is independent of $W$, $a/W$, $B$ (as long as the condition is the so called 'plane stress'), environment, microstructure, and geometry. In addition, such a mechanism would suggest that $K_{op}$ should depend on $K_{\text{max}}$, $K_{\text{min}}$, $R$, $\sigma_Y$, presence of residual stresses, variable amplitude loading which produces load-interaction and retardation, and the short and surface crack behaviour.
Several investigators have reported experimental results which are broadly in agreement with the trends as indicated by the mechanism of plasticity induced closure. However, in view of the uncertainties in the determination of closure, one must carefully scrutinize any result on closure. There are anomalies and other contradictions which cannot be ignored. For instance, often, $K_{op}$ is observed to be independent of $K_{\text{max}}$, $K_{\text{min}}$, and even $R$. The effect of $W$, $a/W$, geometry, and short crack behaviour have not been systematically investigated and the studies on the effect of $B$ on $K_{op}$ are very few. Interestingly enough, there are other investigators who report that $K_{op}$ depends on environment and microstructure (see Sections 4.6 and 4.7). Such discrepancies are often explained in terms of plane stress and plane strain. However, the plane stress and plane strain conditions in terms of experimental situations, are not clearly defined or distinguished by these investigators. It is thus obvious that even though the mechanism of plasticity induced closure is conceptually logical, several anomalies and contradictions concerning the mechanism needs to be resolved.

2.2 ASPERITY INDUCED CLOSURE

The concept of asperity induced (or roughness-induced) closure was first reported by Walker and Beevers [7] and later by McEvily and Minakawa [10] and Ritchie and Suresh [11]. Subsequently, three models based on the asperity induced closure have been proposed by Suresh and Ritchie [12,19], Mays and Baker [60], and Beevers, Carlson, Bell, and Starke [8,9].
It is proposed by these workers that the asperity induced closure mechanism is encountered in situations where the maximum plastic zone size is less than the order of the grain size [9], the reverse plastic zone becomes of the order of dominant microstructural elements [11] or where the size scale of fracture surface roughness is comparable with crack tip opening displacement [10]. In such situations, crack extension occurs along a single slip system resulting in serrated or zig-zag fracture paths with out of plane crack trajectories [9-11]. Extension of the crack along such paths results in significant mode II displacement which plays a role in promoting such closure. Normally, this type of crack extension is observed near threshold (da/dN < 10\(^{-6}\) mm per cycle) when the conditions are predominantly 'plane strain'.

The asperities on the mating fracture surfaces interfere and thus provide discrete contact points across the crack surfaces where load transfer occurs and thereby preventing the crack from completely closing. According to the asperity induced closure models, the wedging action of such interference or contact is small and is localized over very short distance behind the crack front. As a result, such wedging action is locally accommodated permitting the rest of the crack to close except a very short distance behind the crack front where the crack fails to close. Under such circumstances, the failure to close or 'non-closure' as one may term it, produces residual tensile stresses at the crack tip. In a centre crack tension panel, the residual stress pattern produced by this mechanism will be tensile throughout the ligament and in the case of a compact tension specimen, the residual stresses would be tensile near the crack tip and compressive at the back face as shown in Figure 9. Such residual stress distributions are obviously different from those reported in Figure 5.
Figure 9. Residual Stresses Developed in the Plane of Crack in CCP and CT Specimens Due to Asperity Induced Closure as Postulated. Compare with Figure 5.
The three models [8,12,60] proposed for asperity induced closure are examined below.

2.2.1 Single Asperity Model [8]

Beever et al. [8] represents the asperities through-the-thickness by an effective precompressed spring which makes line contact across the thickness. This model is termed a single asperity model and is illustrated in Figure 10, wherein the crack faces are loaded with a force \( P \) at an asperity of height \( L \), and width \( B \), located at a distance \( c \). The asperity undergoes a change in its height by 'e' due to the force \( P \). The local stress intensity factor due to force \( P \) is given by

\[
K_{local} = \left( \frac{2}{\pi c} \right) \frac{P}{B}
\]  

(3)

The corresponding displacement, \( V \), at a point \( r = c \) behind the crack tip at \( 0=180^\circ \) is obtained from the Westergaard solution

\[
V = \frac{2(1-v)K_{local}}{G} \frac{c}{2^m}
\]  

(4)

Substitution of Equation (3) and (4) gives:

\[
V = \frac{2P(1-v)}{\pi GB}
\]  

(5)
L = Magnitude of Interference by the Asperity.
P = Force Developed by an Asperity.
c = Distance of the Asperity from the Crack Tip.
b = Effective Width of the Asperity.
e = Change in Asperity Height Due to the Compression by Force P.

Figure 10. Single Asperity Model [8].
Closure occurs when the crack surfaces make contact behind the asperities and at this point, the stress intensity factor due to externally applied load, \( K_{\text{global}} \), is zero. At this point, \( K_{\text{total}} = K_{\text{local}} = K_{\text{clo}} \) and \( \nu = L - e \) where \( e = \frac{PL}{Eb} \). Substitution of these and Equation (3) in Equation (5) and elimination of \( P \) gives

\[
K_{\text{clo}} = \frac{\sqrt{2}}{\pi c} \left( \frac{1}{Eb} + \frac{2(1-\nu)}{\pi GL} \right)^{-1}
\]

(6)

where \( E, G, \nu \) are the usual elastic constants.

\( K_{\text{op}} \) can be obtained from similar equations and represents the situation when asperity load \( P=0 \) and the compression \( e=0 \). In such a case, \( K_{\text{op}} = K_{\text{global}} = \frac{LGV}{2(1-\nu)(1-v)c} \). Obviously, according to this model, \( K_{\text{op}} \) should be higher than \( K_{\text{clo}} \). The model envisages that the formation of plastic zone of effective length \( r_y \) increases the dimension of \( c \) to \( c + r_y \) and one would then expect \( K_{\text{clo}} \) and \( K_{\text{op}} \) to decrease as \( \sigma_y \) decreases.

The dimensions \( L, b, \) and \( c \) have been measured using replicating technique and these when substituted in Equation (6) give a \( K_{\text{op}} \) value which agrees with the experimentally determined \( K_{\text{op}} \) value in the near threshold regime for a wrought nickel alloy. The typical values are \( c = 15 \) to \( 30 \) \( \mu \)m, \( L = .1 \) to \( .125 \) \( \mu \)m, and \( b = 9-11 \) \( \mu \)m.

The asperity induced closure models are important and interesting because they indicate a method for explaining and predicting the effect the dimensions of the microstructural features may have on closure. The idea
that $K_{op}$ may be determined from the fractographic features is itself very attractive. However, one should not overlook the following points while applying such a model in practice:

1. The compressive stress, $\sigma_c$, in the asperity is given by

$$\sigma_c = \frac{Ee}{L} = \frac{P}{Bb}$$  \hspace{1cm} (7)

Substitution of (3) in (7) gives

$$\sigma_c = \frac{K_I \sqrt{\pi c}}{\sqrt{2b}}$$  \hspace{1cm} (8)

Substitution of the typical values $K_I = 4.5 \text{ MNm}^{-3/2}$, $c = 30 \mu m$, and $b = 10 \mu m$ in Equation (8) yields a value of $\sigma_c = 10,000 \text{ MPa}$. This is almost 30 times the yield strength of the nickel alloy investigated. Since the asperity is in compression, the amount of stress relaxation by plastic flow of the asperity is limited. The question then is would such an asperity survive the battering experienced during the fatigue loading? For example, the growth of the crack over the next $3 \mu m$ ($10\%$ of $c$) would take about 100,000 cycles at $K_{max} = 6 \text{ MNm}^{-3/2}$, but would alter the compressive stresses only by a small amount. Would the nickel alloy asperity survive the 100,000 cycles at a compressive stress of 10,000 MPa? This question is difficult to resolve since there appears to be little scope for the manipulation of the values of $b$ and $c$ so that one can obtain a more realistic value of $\sigma_c$ from Equation (8) at which the asperity can survive 100,000 cycles.
2. According to the model, the closure behaviour is observed due to crack surface contact over dimensions of the order of 15 to 30 μm. This would correspond to an extent of closure, Δa/a ~ .001. It is difficult to detect K from the load-CMOD plot as done by the authors [9], if the extent of closure were as small as .001.

3. The model is supposed to apply to a plain-strain situation whereas all measurements of L, b, and c are made at the specimen surface. In addition, because of the characteristic variation of the fractographic features being what it is, it would be rather difficult to measure L, b, and c with confidence. This is particularly true for L which has a dimension of the order of 0.1 μm.

4. The model shows that K should increase as σ_y increases. On the other hand, the trends of experimental results on the effect of σ_y on K are just the opposite. To explain such trends, it appears that one has to use rather unrealistic values of L, b, and c.

5. The relationship for crack opening stretch width (COSW) is given by

\[ \text{CTOD} = 2 \text{COSW} \approx J/\sigma_y = \frac{K^2 (1-\nu)}{2E\sigma_y} \]  \hspace{1cm} (9)

Substitution of \( K = 4.5 \text{ MNm}^{-3/2} \) and \( \sigma_y = 350 \text{ MPa} \) and \( E = 200 \text{ GPa} \) gives COD ~ .50 μm. In fact, experimentally observed COD values are either in agreement or somewhat larger than those predicted by Equation (9). However,
it seems that the model does not take into account a crack opening of 0.5 \( \mu \text{m} \) and assumes that it plays little role in influencing the force \( P \) on the asperity whose height is as low as 0.1 \( \mu \text{m} \).

2.2.2 Spring Clip Model [60]

Mayes and Baker [60] have attempted to calculate closure induced by roughness by considering an infinite number of compression springs along the crack flank. The springs are in contact, and, therefore, transfer load only when the displacement is less than a minimum. Equating moment and force due to the springs, they obtain a relationship for load which is given by

\[
P = \frac{m e}{c} \cdot \frac{0.83 E B a l_o^2}{m^2 e^2}
\]

where \( P \) is the applied load, \( e \) is back face strain experimentally determined, \( B \) is thickness, \( E \) is Young's Modulus, \( a \) is crack length, \( C \) is compliance, \( m \) is a constant relating \( e \) and load displacement, and \( l_o \) is the effective spring length.

The relationship exhibits a \( P \) versus \( e \) relationship, similar to the one experimentally determined. Also, the residual strain, \( e = e_l \), obtained by substituting \( P = 0 \) in the above equation, obeys a relationship with crack length which is similar to the experimentally observed. These observations are the justification for the validity of the model. The model explains the observed effect of \( R \) and \( C \) on \( \Delta K \)-threshold.
In the model, the effective spring length, $l_0$, characterizes the surface roughness. Calculations show that $l_0$ changes with the material and its value ranges in the neighborhood of .01 to 0.1 µm. However, the roughness has not been measured and related to the equation as done by Beever and coworkers [8].

Spring clip and single asperity models are similar in most respects. Accordingly, some of the points made with regard to the single asperity model in Section 2.2.1 viz the items 3 and 5 are particularly relevant to the spring clip model.

2.2.3 Fracture Surface Roughness Model [12]

The concept of fatigue crack closure induced by fracture surface roughness was first outlined by Minakawa and McEvily [10,18]. Following similar scheme, Suresh and Ritchie [12,19] proposed a model which is reproduced in Figure 11. In this model, fracture surface roughness is idealized in terms of asperities assumed to be of triangular cross section of height, $h$, base, $w$, and all asperities are assumed to be equal in size. The failure to close produces a corresponding residual displacement, The final relationship is:

$$\frac{K_{clo}}{K_{max}} = \sqrt{\frac{2\gamma_X}{1+2\gamma_X}} \quad (10)$$
\[ K = K_{1\text{max}} \]

\[ \delta_{cl} = \delta_{1\text{max}} - u_{I} \]

\[ x = u_{II}/u_{I} \]

\[ \cot \psi = h/(w/2) = \delta_{cl}/u_{II} \]

\[ \delta_{cl} = \frac{2hx\delta_{1\text{max}}}{w + 2hx} \]

\[ K_{cl} = \frac{\sqrt{\delta_{cl}}}{\delta_{1\text{max}}} = \frac{\sqrt{2hx}}{\sqrt{w + 2hx}} \]

\text{or in nondimensional form, with } \gamma = h/w, \]

\[ \begin{bmatrix} K_{cl} \\ K_{1\text{max}} \end{bmatrix}_{\text{MR}} = \frac{\sqrt{2\gamma x}}{\sqrt{1 + 2\gamma x}} \]

\[ [\Delta K_{\text{eff}}]_{\text{MR}} = K_{1\text{max}} - K_{cl} = K_{1\text{max}} \left[ 1 - \sqrt{\frac{2\gamma x}{1 + 2\gamma x}} \right] \]

Figure 11. Microroughness Induced Closure Model [12].
where $\gamma$ is the roughness factor and $x$ is the ratio of Mode I to Mode II displacements. The value of $\gamma$ estimated from surface coating and profilometric studies when substituted into Equation (10) together with the experimentally determined $K_{clos}$ and $K_{\text{max}}$ values, give a value of $x$ or mode II displacement, $u_{\text{II}}$, which is consistent with the experimental examination of crack tip motion using stereomaging procedure. This is how the model is validated.

It should be noted that roughness induced closure can occur only if $R < \sqrt{h/\delta_{\text{max}}}$ and this happens when the $\delta_{\text{min}}$ exceeds the scale of roughness, $h$. The symbols $\delta_{\text{max}}$ and $\delta_{\text{min}}$ refer to the crack tip displacement at the maximum and the minimum loads, respectively.

Such a concept of closure may be useful in explaining why $\Delta K_{\text{th}}$ is higher and $da/dN$ just above the threshold regime is lower in coarser grained or lower strength material [11,12]. For such materials would give rougher surface and therefore higher $\gamma$. It can similarly explain why introduction of a soft phase in a duplex microstructure increases $\Delta K_{\text{th}}$. The soft phase causes the crack path change direction frequently producing higher roughness induced closure. It may also explain the lower $\Delta K_{\text{th}}$ and higher $da/dN$ in short crack (with crack length < grain size) as compared to those observed in the case of long cracks. For a short crack, closure is expected to be much less [11,12] since the roughness is yet to develop.

38
The roughness and the single asperity models are derived from similar considerations. Some of the difficulties of the single and the spring clip models were pointed out earlier. In the case of the roughness model, more comprehensive and direct experimental measurements are required to clearly establish the validity of the model. But it seems that most of the points made earlier with regard to the single asperity model are also valid when one examines the roughness model. Some discussion would be appropriate as regards the first point made.

In the case of the roughness model, it is reasonable to assume that the mode II displacement, $u_{II}$, is microscopic and not global and therefore, it can be accommodated by regions local to the crack tip; for if the displacement is global, it would give very large and an absurd value of the extent of closure, such as $\Delta a/a = 1$. Thus, $u_{II}$ may not be observable at distance much behind the crack tip. In that case, the distance behind the crack tip over which contact between asperities transfer compressive stresses due to the presence of Mode II displacement, becomes an important factor. Obviously, this distance must be long enough to distribute the compressive force generated by non-closure so that the stress level in the asperity are sufficiently low that they survive battering. On the other hand, this distance should be consistent with the extent of closure observed in an experiment. This is an important aspect and needs to be ascertained while verifying the proposed roughness model.
An examination of the fractured surface of a fatigue crack shows a battered appearance; the sharp edges one sees in the asperities of a monotonically loaded fracture surface are not usually observable in a fatigue fracture surface. On the other hand, one should note that large facets are present on the fracture surfaces obtained in the fatigue threshold regime. However, their presence neither guarantees that large mode II displacement have been experienced by the cracked body nor is it proof that the asperity induced mechanism of closure is operative. The mode II displacement and its distribution behind the crack length should be measured using suitable experimental technique to ascertain its role in asperity induced closure.

In a recent investigation [4], fatigue crack growth of Ti-Al alloys have been studied at room and elevated temperatures in air, vacuum, and argon in which crack closure was also determined. It was concluded that the crack closure mechanism is related to the macroscopic surface topography. The surface topographies of the different fractured specimens were examined and one can estimate that the asperity heights are in the range of hundreds of microns. Such asperity heights are three to four orders of magnitude higher than the roughness contemplated in the three asperity induced models outlined above.

Indeed, the roughness induced closure needs improved foundations of experimental and analytical work.
2.3 OXIDE INDUCED CLOSURE

Paris et al. [20] was the first to speculate that in a reactive environment, a freshly created fatigue fracture surface oxidizes and builds up corrosion product. According to them, this causes crack tip interference which effectively impedes crack growth in a manner similar to that described by Elber's crack closure mechanism. A similar argument was used by Stewart [21] to explain his experimental results on environmental effects. Following a similar approach, Ritchie, Suresh, and Moss [13] postulated the concept of oxide induced closure.

The oxide induced closure arises when the corrosion products having a thickness (typically several microns thick) comparable to the size of the crack tip opening displacement, build up near the crack tip and as a result, wedge-open the crack at \( K_i > K_{\text{min}} \). Accordingly, during the closing portion of the cycle, contact between fracture surfaces will occur earlier, thereby raising closure loads and correspondingly reducing \( \Delta K_{\text{eff}} \) [13,14]. Like asperity induced closure, oxide induced closure is observed at low growth rates \( (da/dN < 10^{-6} \text{ mm/cycle}) \) associated with near threshold condition at low load ratios under the so called plane strain condition [7-12]. Such a mechanism of closure [13,14,21] can explain some surprising observations relating to the effect of environment on \( da/dN \) and \( \Delta K_{\text{th}} \).

Figure 12 illustrates how the oxide deposits formed on freshly exposed surfaces at the crack tip in a moist environment can effectively wedge-close the crack at stress intensities above \( K_{\text{min}} \) and how such a
Figure 12. A Schematic Comparison of the Oxide Induced Closure Mechanism with the Other Mechanisms of Closure [12].
Figure 13. Basis of Oxide Filled Closure - A Schematic Effect of Environment on $\Delta K_0$ and $da/dN$ [22].
Figure 14. Basis of Oxide Induced Closure. Typical Variation of Excess Oxide Thickness, $d$, with Crack Length and Crack Propagation Rate [22].
One must note that oxide induced crack closure does not play a primary role in the near-threshold crack growth behaviour of 2021-T6 and 2024-T3 and peak-aged 7075-T6 aluminum alloys [22], where the thickness of oxide is small compared to the CTOD. This is also true for high strength steel [23] where it has been reported that the crack growth rates in regime just above the threshold are slower in hydrogen or argon as compared to in air. There are also instances such as in overaged 7075-T7 aluminum alloy where the oxide formation is significant and comparable to CTOD leading to crack arrest although $\Delta K_{th}$ is unexpectedly low. This has been attributed to the concurrent action of oxide induced closure and hydrogen embrittlement [22].

A few comments on the oxide induced closure mechanism are in order.

As is to be expected, oxide induced closure plays a major role only in specific combinations of material and environment. However, the same material when exposed to an environment which produces no oxide, does indeed exhibit finite $K_{op}$ value. Also, it has been stated [14] that unless plasticity induced closure or mode II rubbing produces the fretting contact, no oxide formation or oxide induced closure can take place. Thus, whereas plasticity and asperity induced mechanisms can produce closure in all instances, the oxide induced mechanism produces additional closure only in specific instances.

The experimental observations [13,14,21,22] reported, qualitatively support the main postulates of the oxide induced closure mechanism. However, no quantitative model interrelating $K_{op}$ with oxide thickness and its properties has been proposed. One can conceive that such a model could be
along lines similar to the single asperity model. However, the difficulties in such a model as pointed out earlier, has to be kept in mind. Nonetheless, at present, the understanding of oxide induced closure is not adequate to predict or calculate the contribution of oxide formation to $K_{op}$.

One must also note that experimental verification of the extent of closure and the residual stress distribution which are characteristic of oxide induced closure have not been investigated and verified. For instance, oxide induced closure should exhibit a residual stress distribution similar to the one reported in Figure 9, although this has not been ascertained.

2.4 COMMENTS ON THE MECHANISMS OF CLOSURE

From the above discussion, the differences between plasticity induced closure on the one hand and the asperity and oxide induced closure on the other, are quite obvious. First, the nature of the residual stress distribution near the crack tip are expected to be quite different (see Figures 5 and 9). Second, the asperity and the oxide induced closure mechanisms apply to situations where plasticity, $\Delta K$, $K_{\text{max}}$, $da/dN$, and the extent of closure that is $\Delta a/a$, are all very small. Presumably, plasticity induced closure applies to the opposite situation. However, it is interesting to recall a point made earlier: plastic zone formation is essential for the generation of closure by the oxide induced mechanism, even though the plasticity under such circumstances is extremely limited. The role of plasticity is considered to be secondary in asperity induced closure. However, a plastic wake forms even if the $K_{\text{max}}$ value is low and the plastic wake will produce the

47
characteristic compressive residual stress pattern and the crack opening stretch at the crack tip. Neither the oxide nor the asperity induced closure mechanism takes into account the presence of such residual stresses.

One of the major discrepancies in plasticity induced closure is that $K_{op}$ is observed to be practically independent of $K_{\text{max}}$ even when the plasticity is not negligible. An increasing $K_{\text{max}}$ should increase the size of the monotonic and reverse plastic zone. Therefore, it is not the reversed plasticity near the crack tip but the integrated effect of the residual displacement produced by the plastic zone and the wake, which plays the predominant role in determining closure. Similarly, contrary to expectations, $K_{op}$ is observed to be independent of $K_{\text{max}}$ even at $K_I$ levels where asperity and oxide induced mechanism are assumed to be inapplicable. Thus, a mild dependence of $K_{op}$ on $K_{\text{max}}$ is not necessarily a characteristic of oxide or asperity induced closure. On the other hand, the observation that $K_{op}$ decreases with increasing $\sigma_y$ even at very low $K_{\text{max}}/\sigma_y$ levels of loading indicates that asperity and oxide induced closure are either dependent of crack tip plasticity or, alternatively, the mechanisms play only a small role in producing crack closure.

In evaluating the effect of $K_{op}$ on $da/dN$, it is assumed that a fatigue crack cannot grow as long as it remains closed. This, in turn, is based on the assumption that fatigue crack growth rate is generated by the strength of the singularity or rather the stresses and strain around the crack tip. Since the stresses and strains near the crack tip are comprehensively represented by $K$, $K$ becomes the crack driving force for the growth of
a fatigue crack. In a cyclic loading situation, one would then expect alternating levels of $K$ or $\Delta K$ to be the crack driving force and this then would control $\frac{da}{dN}$. Since there is no detailed micromechanism interrelation of $\frac{da}{dN}$ with $\Delta K$ level and material properties, the constants $A$ and $n$ in Equation (1), which represent material behaviour, are based on empiricism and have to be obtained from experimental measurement.

It is clear from the above discussion that the exact mechanism by which $\Delta K$ controls $\frac{da}{dN}$ is not known. Therefore, the role of the different patterns of residual stress distribution obtained through the different mechanisms in influencing $\frac{da}{dN}$ is hard to analyze and predict. However, one can speculate that a residual compressive stress distribution at the crack tip produced by plasticity induced closure, effectively decreases $K_{\text{max}}$ and thus lowers $\Delta K$. On the other hand, the residual tensile stresses at the crack tip produced by theasperity or oxide induced closure increases $K_{\text{min}}$. In either case, $\Delta K_{\text{eff}}$ would be decreased even though alternative schemes of closure produce opposite patterns of residual stress distribution. Such an explanation can provide the rationale for decreased $\frac{da}{dN}$ due to $K_{\text{op}}$ corresponding to two opposite patterns of residual stress distribution.

The above rationalization appears too simplistic since it is well known that fatigue crack growth rate is a more sensitive function of $K_{\text{max}}$ than $K_{\text{min}}$. In fact, this is evident from a close examination of the usual data showing increasing $\frac{da}{dN}$ with increasing $R$ for a given value of $\Delta K$. Thus, a decrease in $K_{\text{max}}$ value due to residual compressive stress (produced by plasticity induced closure) cannot produce $\frac{da}{dN}$ which is the same as that
obtained by an increase in $K_{\text{min}}$ due to residual tensile stress (produced by asperity or oxide induced closure). Thus, the above rationalization is not consistent with the observed dependence of experimentally determined $\frac{da}{dn}$ on $K_{\text{max}}, R,$ and $K_{\text{min}}$.

There is another point with regard to the asperity and the oxide induced mechanisms which needs careful examination. If $K_{\text{op}}$ decreases $\frac{da}{dn}$ by increasing $K_{\text{min}}$ as implied by these mechanisms, one would then expect $\frac{da}{dn}$ to be independent of varying $K_{\text{min}}$ in an experimental situation where $K_{\text{op}} > K_{\text{min}}$ and $K_{\text{max}}$ is held constant. This is not observed to be true [67]. Thus, the manner in which the two opposite patterns of residual stresses as well as the two different schemes of closure decrease $\frac{da}{dn}$ is rather confusing.

Notwithstanding the discussion above, one can formulate roughness and oxide induced closure models, wherein the additional residual displacement and the wedging action produced by roughness or oxidation, is not localized just behind the crack front but is spread evenly over the whole length of the fractured surface. Such a distribution of displacement will produce a residual compressive stress pattern over the crack faces and ahead of the crack which is identical to that produced in the case of plasticity induced mechanism and accordingly, plasticity, asperity, and oxide induced mechanisms will then produce identical effects on $\frac{da}{dn}$, thus resolving the discrepancy discussed above. Based on such a formulation of asperity and oxide induced closure, $K_{\text{op}}$ can be calculated by using methodologies normally adopted for plasticity induced closure mechanism discussed later in Section 5, wherein the microstructural roughness parameters and the volume generated by oxidation could be the additional inputs.
One can possibly use back face strain gage together with the CMOD technique to determine the residual stress pattern corresponding to a given mechanism of closure. The measurement of residual strain at the front and back of the centre crack panel specimen at zero load would indicate the nature of residual stress distribution which is characteristic of a given mechanism of closure. In the case of CT specimen, one has to measure the strain also at the intermediate points in the ligament to ascertain the pattern of residual stress distribution.

In the case of plasticity induced closure, it may be necessary to distinguish between the role and the effect of plastic wake and the monotonic plastic zone in producing closure. The former probably produces bulk closure behaviour, generates a large extent of closure and residual displacement, and is easily detectable. On the other hand, the latter probably produces local closure, generates a small extent of closure and residual displacement, and requires sensitive technique for its detection. In a constant amplitude test, the role and the effect of these two closures may not be distinguishable but in a situation where an overload is applied, the distinction between the role and effect of these two closures would be quite obvious.

Finally, one must note that asperity and oxide induced closures operate only in the near threshold regime. There are no models for oxide induced closure and the models proposed for asperity induced closure are in their preliminary stage and require validation through systematic experimental measurement. Thus, these closure models cannot be used for calculation of closure in a component even though roughness and oxide induced closures are
feasible. On the other hand, the closure observed at higher $K$ levels can be calculated only on the basis of plasticity induced closure. The approaches used for such calculation to predict closure in a precracked body are examined in Section 5.
3. EXPERIMENTAL DETERMINATION OF CLOSURE

The determination of closure is based on the measurement of one or more of the parameters such as: displacement, electrical potential, strain gage output or transmitted ultrasonic intensity from the specimen, as the load increases or decreases. The point of transition in the plot of one of these parameters against load is identified as closure. It is not easy to identify the transition point in such plots with certainty. Often the last fraction of opening during the loading phase of the cycle can be rather difficult because closure produces compressive forces which are larger near the crack tip (see Figure 5).

Some investigators consider the transition point as the minimum stress intensity factor above which the crack is fully open and therefore identify it as the lowest point at which the plot becomes linear (marked \( K_c \) in Figure 15). Others draw two tangents on the plot - one corresponding to the minimum load and the other corresponding to the fully open crack. As shown in Figure 15, the point of intersection of the two tangents, that is \( K_{ct} \), gives the load at closure. The advantage of the tangency method is that it gives consistency between the closure values obtained by two different techniques when simultaneously used in a specimen.

The transition point determined during load and unloading usually occurs at the same load but some investigators prefer to determine the transition point from the plot obtained during the unloading phase. It was pointed out in the introduction that the transition point can be identified with less difficulty if one generates a load-offset displacement plot instead of the usual load-displacement plot.
Figure 15. Identification of Closure from Load Versus Displacement (P-V) and Load Versus Offset Displacement (P-ΔV) Plots. $K_2$ represents the minimum stress intensity factor at which the plots are linear and $P_{ct}$ is the stress intensity at which the two tangents on the P-V plot intersect.
The other way to conveniently identify the closure load is to use a more sensitive instrumentation technique. It is argued that since the near tip region of the crack is last to open during loading, the displacement measured at points close to the crack tip would have a larger contribution from closure and therefore, such a measurement is expected to reveal the transition point more clearly. This has indeed been the rationale for the use of Elber type gages as well as the strain gages near and across the crack; for all these measurements can be made at points close to the crack tip.

The thickness averaged global closure behaviour has been determined by techniques such as CMOD, back face strain gage, electrical potential, and ultrasonics. But even the techniques such as the Elber gage or strain gages bonded across the crack or near the crack tip as referred to above, could give thickness averaged global closure behaviour if the distance between the location at which the displacement is measured and the crack tip is large compared to the depth of the curved crack front, the plastic zone size, and the specimen ligament.

As pointed out earlier in the introduction, the closure behaviour determined from the displacement measured at points very close to the crack tip would exhibit a trend which is quite different from the thickness averaged global closure behaviour. It was also pointed out in Section 2.1 that the residual stress may vary across the thickness near the crack tip and exhibit a complex pattern due to preferential plastic flow in the surface layers of the specimen and due to the presence of a curved crack front at the interior of a specimen. Accordingly, the closure behaviour based on the measurement at the surface and those at the interior of the specimen near the crack tip are expected to be different from each other.
The more successful techniques for the determination of closure based on the near tip surface measurements are the methods based on optical interferometry. These can measure displacements localized over small areas (say gage lengths less than 0.5 mm) with a precision better than 0.5 μm. In addition, one can use the technique of surface replication followed by SEM observation to determine near tip surface closure behaviour. If the depth of crack tunneling, the plastic zone size, and the ligament length are large, the difference between the interior and surface closure behaviours is easily detectable. In such cases, techniques such as small strain gages glued either across the crack or near the crack tip or Elber gages mounted very near the crack tip can also give closure behaviour which as a significant contribution from the near tip surface displacement. For this, the gage length of the strain gage or the clip gage should be small compared to the dimension of the plastic wake in the y-direction. The \( K_{op} \) values determined on the basis of near tip surface measurements are usually greater than thickness averaged global \( K_{op} \) values.

At present, experimental observations are too meager to define the dimension over which the near tip closure effects would be observed. Obviously, such dimension would depend on the depth of curved crack front, plastic zone size at the specimen surface, and the depth of penetration of plastic zone across the thickness and the ligament of the specimen.

As discussed earlier in Section 2.1, the residual stress in the interior near the crack tip is tensile and therefore, the closure would be observed at a lower load level in the interior as compared to at the surface. Only a few investigations have been carried out to determine the closure
at the interior and the following techniques have been used for the purpose: optical interferometry in transparent specimen, the measurement of closure before and after the removal of a surface layer, and the push rod gage technique.

The thickness averaged bulk closure behaviour would be primarily influenced by the residual stress pattern across the width of the precracked body. But this closure behaviour could also be influenced to some extent by the pattern of residual stress across the thickness near the crack tip. In fact, one can visualize that the closure of the last part of the crack could be influenced by the residual compressive stresses in the surface layers, particularly when the crack front is significantly curved. It has been observed that the closure effects determined on the basis of near tip surface measurements are higher than thickness averaged global closure. Similarly, it has also been shown that $K_{op}$ at the surface is higher than that observed at the interior. However, whether the near tip interior $K_{op}$ corresponds to the thickness averaged global closure is not conclusively established.

The above discussion concerns constant amplitude loading where the bulk and local closure behaviour could be identical as pointed out in Section 2.4. It was proposed that application of a single cycle overload could produce two closures - representing the global and the 'local' behaviours, respectively. The term 'local' refers to the average near tip closure behaviour and probably corresponds to the upper closure point reported by some investigators [32]. One can speculate that in a specimen of intermediate thickness, one would observe some difference between the 'local'
the near tip interior, and the near tip surface closure behaviours. Possibly, the offset displacement procedure technique could be developed and standardized to determine local closure behaviour. This prospect will also be examined in this section.

In this section, the different procedures for closure determination will be discussed under three sub-headings.

1. Thickness Averaged Bulk and Local Closure Behaviour

2. Near Tip Surface Closure Behaviour

3. Near Tip Interior Closure Behaviour

3.1 THICKNESS AVERAGED BULK AND LOCAL CLOSURE BEHAVIOUR

The following techniques are used for the determination of the thickness averaged bulk closure behaviour.

1. CMOD Gage \([3,24-32]\)

2. Strain Gage \([3,27,28,33-37]\)

3. Ultrasonics \([12,38-41,52]\)

4. Potential Difference \([3,4,28,30,42-46,60,65]\)
5. Special Displacement Gage [2,3,6,28].

While using the special displacement gages such as Elber and Nowack gage or strain gages bonded across or near the crack tip to determine thickness averaged bulk closure behaviour, one must remember that the location of such gages has to be far behind the crack tip so that the gage outputs are not influenced by the presence of curved crack front or through-the-thickness yielding at the surface layers or surface strains near the crack tip.

The thickness averaged local closure behaviour could be determined from CMOD gage output using offset displacement technique.

3.1.1 CMOD Gage

The displacement is measured from a clip gage mounted across the notch mouth and located either at the load-line or at the cracked edge (in case of single edge-cracked specimen). A plot of the type shown in Figure 15 is obtained from which the transition point can be identified. The standard CMOD gage, the fixtures used for locating the gage into the specimen, and other details of the related instrumentation are described in the standards (ASTM E399 or E647). As pointed out earlier, the transition point can be more clearly identified from a load versus differential displacement plot (see Figure 2b). Such a technique has been used by several investigators [3,31,32,68]. The basic idea of the measurement of the offset displacement is discussed below.
When the specimen is fully open, the displacement, \( V \), is related to load, \( P \), linearly through

\[ V = CP \] (11)

where \( C \) is the compliance of the specimen. However, the \( P-V \) plots have nonlinearity which can be represented by the offset displacement as given by

\[ AV = g(V - CP) \] (12)

where '\( g \)' is the gain of the amplification and \( AV \) gives the nonlinear part of the displacement or the offset displacement. Thus, at loads higher than the closure load, \( AV = 0 \) and a plot of \( P-AV \) is a vertical line (see Figure 2b). At loads less than the closure load, the \( P-AV \) plot is nonlinear. The point of transition from linear to nonlinear gives the closure load. Accordingly, the closure load can be identified with a higher sensitivity and less subjective error from such a plot. Offset displacement can be obtained using analog circuitry or digital data in a microprocessor.

It cannot be overemphasized that the accuracy of the \( P-AV \) plot depends primarily on the accuracy of the measurement of the load and the displacement. Accordingly, the errors in displacement measurement due to the misalignment and friction in the loading fixture or clevis and the clip gage support has to be minimized. One can use clevis with ball bearings [32,141], and special support point for the clip gage [3] to achieve this.
In addition, one must note that the identification of the closure load has some inherent uncertainty. The minimum stress intensity factor at which the P-V plot is linear if identified as the closure point (marked $K_{op}$ in Figure 15) would be somewhat higher than the closure point identified by the tangency technique (marked $K_{ct}$ in Figure 15). One can use a regression analysis to determine $K_{op}$ from the P-V plot. But a small amount of inherent nonlinearity in the upper part will introduce uncertainty in determination of the true $K_{op}$ from P-V test record. Because of higher sensitivity of P-ΔV test records, $K_{op}$ determined from such a record could be higher than that obtained from the P-V test record (see Figure 15).

If one uses P-ΔV plot, the identification of closure is easier, but even this procedure is not free from some degree of uncertainty. Figure 15 gives an idealized representation of the P-ΔV plot. The actual P-ΔV test records tend to develop a loop area since the signal ΔV is amplified in such records. To generate P-ΔV plots, one normally uses a value of $C$ in Equation (12) corresponding to that experimentally observed at loads which are higher than $P_{op}$ but which are somewhat less than $P_{max}$. As a result, the P-ΔV plots are curved as illustrated in Figure 16. The identification of $K_{op}$ from such test records would be unambiguous, only if the upper part of the P-ΔV record is a straight line. But the upper part of the P-ΔV test record always tends to curve. The origin and interpretation of this curvature are discussed next.
Figure 16. Schematic of the Typical P-M Test Records. Friction and Misalignment may Change the Test Records from the Type Reported in Figure 16a to that in Figure 16b.
It is obvious from Figure 16 that at low load levels, that is between the points marked A and B, the bulk of the crack opens up as the load increases and the curve bends clockwise. At intermediate load levels, that is between B and C, the curve is nearly vertical and this facilitates identification of $K_{op}$ as shown in Figure 16a. However, uncertainties in the identification of $K_{op}$ from the plots of the type reported in Figure 16a can arise if the amplification is high or if the plasticity is large for they tend to decrease the interval between B and C and increase loop area. As the load increases further, the P-ΔV test record again bends clockwise between C and D due to one or more of the following reasons.

1. Incremental growth of plastic zone

2. Crack extension

3. 'Local' opening of the crack near the tip

4. Misalignment and out of plane bending

If crack extension is absent and incremental growth of plastic zone is negligible, as happens after a single cycle overload, and the misalignment and out of plane bending effects are negligible, the clockwise bend in the P-ΔV test record in the region C to D would indicate a local opening of the crack. Thus, an identification of the local or an upper [32] closure behaviour from the P-ΔV test record is possible. However, in the case of constant amplitude loading, such identification would be erroneous.
On the other hand, a P-ΔV test record of the type shown in Figure 16b shows that the curve in the region between C to D bends counterclockwise. Such a test record is encountered quite often, both in constant and variable amplitude loadings. The counterclockwise bend indicates that the crack is closing as load increases. Since a crack cannot close with increasing load, a counterclockwise bend in the upper part of the P-ΔV test record (that is between B and D in Figure 16b) occurs due to friction and misalignment as stated earlier.

Even if one eliminates misalignment and friction, the vertical part of P-ΔV test plot tends to bend in a clockwise manner due to the growth of plastic zone at large $K/\sigma_\text{Y}\sqrt{W}$ values or due to further opening of the crack at the 'local' opening point. The determination of $K_{\text{op}}$ is difficult if the curve bends either in a clockwise or a counterclockwise manner. Finally, at higher amplification or higher $K/\sigma_\text{Y}\sqrt{W}$ value, the loop areas are excessive and this tends to further complicate the identification of $K_{\text{op}}$. Thus, some amount of arbitrariness is unavoidable if one uses P-V or even the P-ΔV plots, for the identification of closure.

$K_{\text{op}}$, as identified from the P-ΔV test record, need not correspond to $K_\text{p}$ or $K_\text{ct}$ obtained from the P-V plots. And, the question as to which of these — that is $K_\text{p}$ and $K_\text{ct}$ — more appropriately represents closure is, therefore, not significant. If the specimen has a curved crack front, then the crack tip at the interior may be open at a stress intensity factor less than $K_{\text{op}}$. One the other hand, there is every likelihood that at $K = K_\text{ct}$, the last fraction of the crack might not have opened up, even in a constant amplitude loading situation. In a hi-lo load sequence, in addition, one has to also consider the possibility of a local closure.
CMOD gage is an attractive, convenient, and popular technique for the determination of thickness averaged $K_{op}$ since most often one has to determine compliance of the specimen during fatigue crack growth studies. Even though using such a technique together with the offset displacement procedure enables one to achieve a highly sensitive detection of the last fraction of crack opening, the techniques need further improvement and standardization with regard to specimen alignment and minimization of friction at the clevis pins and at the clip gage support. In addition, further investigation is required to properly interpret the various features of a load versus offset displacement plot so that the occurrence and the identification of an upper closure point [32] can be properly ascertained.

3.1.2 Strain Gage

In order to measure thickness averaged bulk closure behaviour, strain gages can be bonded to the specimen at various locations, A, B, C, and D in Figure 17. A commonly used location is on the back face of a specimen, marked A in Figure 17. The strain gage at D straddles the crack and only the ends of the gage length of the strain gage are bonded onto the lower and upper part of the specimen across the crack. The middle part of this strain gage is unbonded. In all cases, the signal from the strain gage is recorded against load. One can then use either the criterion of deviation from linearity or the intersection of tangents to determine the closure load.
Figure 17. Preferred Locations of the Strain Gage in a CT Specimen for the Determination of Closure Using Strain Gage Technique.
The back face strain gage technique usually gives results which agree well with those obtained from CMOD gage unless the crack length is very short. Besides, the back face strain gage shows less hysteresis. Therefore, back face strain gage technique is widely used for the determination of closure.

The strain gage mounted at location B, C, or D (see Figure 17) can also give thickness averaged bulk $K_{op}$ if the strain gage is large and is not located very close to the crack tip. In fact, $K_{op}$ results obtained from such measurements are also in excellent agreement with those obtained from CMOD.

There are stress-free regions in a precracked specimen, particularly at points which are behind the crack tip but just above and below the crack. Thus, a strain gage located at such points may not experience any strain, therefore, it gives either no erroneous closure [3,34]. It was observed that strain gage of 1x2 mm located with its centre at $a/W = 0.5$ in a $B = 4$ mm and $W = 50$ mm CT specimen at a point 3 mm above the crack faces can sense the closure over the whole range of $0.4 < a/W < 0.65$ [34]. There could be other locations also in the plane of the specimen.

The strain gage, particularly at the back face, has advantages over the CMOD clip gage for the determination of closure. If proper strain gage procedure is followed, one can dispense with the problems of friction at the clip gage support. Besides, at high frequencies and high temperatures, strain gage technique has some obvious advantages in determining closure.
3.1.3 Ultrasonics

As the crack closes, the acoustic resistance of a specimen changes and, therefore, the intensity of the ultrasonic signal reflected from or transmitted through a fatigue crack changes as closure takes place. The basic scheme of the technique is illustrated in Figure 18a. A schematic plot of the load versus transmitted ultrasonic intensity is shown in Figure 18b. In this figure, $P_{\text{op}}$ based on deviation from linearity and $P_{\text{ct}}$ based on tangency method, give two different closure loads.

However, ultrasonic methods do not always give closure loads which agree with those obtained from CMOD or strain gage techniques. It is reasoned that the asperities on mating crack surfaces slide past one another and the contact between the asperities keeps the crack surfaces acoustically closed even though it is mechanically open [28,46]. The sliding is introduced due to the presence of mode 11 which, in turn, is produced either due to specimen misalignment or crack branching. However, ultrasonic techniques, unlike the electrical potential technique, can be used even if the specimen surfaces are oxidized [12].

The ultrasonic technique has been successfully and consistently used by one group of workers [38-41]. However, the experience of some others with regard to this technique is not so satisfactory [52].
Figure 18. Determination of Closure Using Ultrasonics Technique [41]. $P_z$ and $P_{ct}$ are explained in section 3.1.1.
3.1.4 Potential Difference

The electrical resistance of a specimen changes as the crack closes. This principle as applied in the potential difference technique is illustrated in Figure 19. A constant current source supply or a constant voltage source together with a resistor in series with the specimen, can be used to feed the current probe. The resultant signal across the potential probe is amplified and recorded. One can use the offset procedure to obtain offset output and examine the load-offset output plot for a more precise identification of closure [3].

Since the output signal from the specimen is small, necessary precaution against thermal drift is essential. Also, it is reported [3] that potential could be measured across the potential probe position 1 or 2 (see Figure 19). The tangential technique could be used to identify closure point.

However, one must note that potential difference technique gives results which do not agree with those obtained from CMOD or back face-strain gage, not only for titanium alloys [28,65], but also for steel [3,60]. This discrepancy could arise due to different reasons [28].

First, a tenacious and insulating oxide layer forms when the specimen is cyclically loaded in air and this prevents electrical contact and closure detection even when there is mechanical contact and the load is transferred across the crack surfaces. Second, but an opposite reason, is that when a specimen is cyclically loaded in vacuum, the asperities on the
Figure 19. Determination of Closure Using DC Potential Drop Technique. 1-1 and 2-2 - Potential Probe Position and 3-3 - Current Probe Position [3].
mating fracture surfaces may slide and make electrical contact without effectively transferring load across the crack surfaces. Finally, plasticity at the crack tip and varying contact resistance at the specimen/pin contact can contribute to the potential drop measured. In fact, it has been stated that the results of potential drop technique have to be interpreted with extreme caution [2].

3.1.5 Special Displacement Gages

Special displacement gages such as single cantilever Elber gage [6], Nowack gage reported in Reference [28], and twin cantilever gages [3] have been developed for the measurement of displacement or small gage lengths across the crack on the specimen surface and at points very close to the crack tip. These gages have high sensitivity, exhibit low hysteresis, and give P-V plots with higher slope change at the transition point for the identification of closure. Thus, in the initial stages of closure studies, they were widely used for the determination of closure. One can also use load-offset displacement procedure [3] for a more precise identification of the closure point from the outputs of these gages.

As pointed out earlier, such displacement gages can be used for the determination of bulk closure. But in situations where the depth of crack front, plastic zone size, thickness, and width of the specimen are relatively large, such displacement gages can be used for the determination of near tip surface closure.
In case of ordinary clip gages located near the crack mouth, the displacement signals are large and therefore, closure which changes the compliance of the specimen contributes only a small fraction of it. However, if the displacements were measured over small gage lengths close to the crack tip, the total displacement would be small and closure would make a large contribution to the total displacement measured. As a result, the transition point can be identified more easily [2,3].

It has been pointed out [62] that surface strains behind the crack tip may influence the displacement values measured by special displacement gages of small gage length and located very close to the crack tip and this can affect the closure value somewhat. Thus, such special displacement gages should be used with caution. On the other hand, it has also been reported [3] that special displacement gages located near the crack tip give closure values which are identical to those obtained by CRND or back face strain gage technique.

3.2 NEAR TIP SURFACE CLOSURE BEHAVIOUR

It has been pointed out earlier that the special displacement gages or strain gages located very close to the crack tip can measure near tip surface closure behaviour in certain circumstances. However, the more reliable method for the measurement of near tip surface closure behaviour are the following:

1. Interferometric displacement gage [49-54]
2. Direct observation using SEM [3,7,8,9,47,56,62]

3. Optical interferometry [52,53,63,89].

3.2.1 Interferometric Displacement Gage (IDG)

IDG is a laser based technique which measures the relative displacement between two shallow reflecting indentations located .05 to 1 mm apart across the crack. The indentations are produced by microhardness indenter. The two interference fringe patterns form due to overlapping diffracted laser beams. The motion of the fringe pattern is a measure of the displacement. The resolution of the technique is 0.25 μm for a range of 400 μm; the corresponding error band is 1 percent. Since the gage length of IDG could be as small as 50 μm, it can measure displacement within a very small region around the crack tip. Such measurements are bound to be influenced by preferential yielding through the thickness in the surface layer and the presence of curved crack front and shear lips. This can indeed produce a much stronger closure effect than exhibited by the typical bulk measurements such as CMOD or back face strain gage techniques. Indeed, closure measured [52,53] using IDG technique show the above trend. However, IDG measurements can also be influenced by surface strains near the crack tip [62].

IDG technique is powerful since it is a non-contact method so it has all the advantages of avoiding friction, high temperature, or environment. One can numerically process the fringe data and obtain real time measurement. Its use at a faction far behind the crack tip can give reliable through-the-thickness bulk closure behaviour.
3.2.2 Direct Observation Using SEM

One of the convincing techniques for determining near tip surface closure is direct observation of a two stage replica taken of the crack tip region of the specimen. The two stage replication consists of acetate tape, evaporation of gold on it, and the final support of the replica using electrodeposited copper [8,9]. Almost a similar procedure but different somewhat in the details, was used by another investigator [3]. Since the acetate has a resolution of .01 μm, displacements as small as .01 μm can be measured.

Replicas have to be taken at fixed loads; thus, several replicas are required to determine the closure load. However, the use of replicas is a direct method of closure observation and if properly carried out, is free from ambiguity of closure determination when other techniques are used.

A variation of this method consists of producing a fine scratch mark running along the specimen surface at an angle to the crack [36]. As the crack opens, the scratch mark splits at the crack line and therefore, shifts a certain distance along the x direction. This shift can be related to the crack opening. One can probably further modify this method to determine the near tip mode II displacement produced by the $K_{II}$ component, if the scratches are marked at fixed intervals of say 50 μm or so. It must be noted that at a higher magnification, the crack changes path frequently and appears to branch. Such change of path of crack propagation has been observed even when $\Delta K$ was significantly higher than threshold [3]. In fact, at higher $\Delta K$, 75
the branches could be longer. It is doubtful if such extensive branching occurs at the interior during constant amplitude loading. What, however, is interesting is that a direct observation using SEM gives closure values which agree very well with the CMOD or back face strain gage \([3,62]\) which measure bulk \(K_{op}\) values. On the other hand, one \([3]\) of these two investigators observed that direct observation gives closure values which agree very well with the closure value obtained from special displacement gage located very near the crack tip whereas the other \([62]\) reported just the opposite.

3.2.3 Optical Interferometry

The optical interferometry technique can be used to measure the transverse displacement at points close to the crack tip. A collimated beam of monochromatic light is directed through an optically flat quartz plate positioned directly on the polished specimen surface. With load, the specimen surface separates from the quartz plate and produces an optical interference pattern. Knowing the fringe order, the transverse displacement can be measured with a resolution of 0.25 \(\mu m\) at a given point at various times during loading. A combination of the load-time and transverse displacement time data gives load versus transverse displacement record from which one can determine closure.

The optical interferometry gives results which agree very well with those obtained by IDG technique. This is indeed interesting for optical interferometry measurements can yield a displacement ahead of the crack whereas IDG measures a displacement behind the crack tip. This supports the viewpoint that there is probably a fair degree of continuity in the
distribution of residual displacement and compressive stresses across the crack front. On the other hand, these $K_{op}$ measurements do not agree with $K_{op}$ obtained from CMOD gages. This is to be expected. However, what appears confusing is that the replication of crack tip followed by direct observation under SEM give $K_{op}$ which agree well with CMOD gage [3].

The use of optical interferometry for the determination of closure is limited but could be useful if one were to investigate the transverse displacement together with closure.

3.3 NEAR TIP INTERIOR CLOSURE BEHAVIOUR

The investigation of near tip interior closure behaviour is very limited. The following four techniques have been used:

1. Closure measurement before and after the removal of successive surface layers [2]

2. Push rod displacement gage technique [65]

3. Vacuum infiltration technique [64]

4. Interferometric technique in transparent specimen [58].
3.3.1 Closure Measurement Before and After the Removal of Successive Surface Layers

The closure was determined in a 10 mm thick 2024-T3 Al-alloy specimen ($W = 100$ mm, $a = 12.7$ mm) after successive removal of 1 mm surface layers from each side [2]. As can be seen from Figure 20a, $K_{op}$ decreased drastically when the first layer was removed. The lowering was much less during subsequent removal but the lowering of $K_{op}$ continued even to a reduced thickness level of 4 mm when it reached 50% of its original value. On the other hand, it can be noted from Figure 20b that closure measured on specimens of different thicknesses prepared from the same material show that increasing thickness decreases $K_{op}$ [2]. Thus, the major part of the closure probably originates from the preferential through-the-thickness deformation in the surface layer. The presence of a curved crack front and the formation of large shear lips could also contribute to such a marked thickness dependence. Even though it is an interesting and straightforward technique, no similar investigations have been reported.

Since electrical potential technique had been used to detect closure in the investigation and the electrical potential technique is not entirely satisfactory in determining closure, one can suspect the results reported in Figure 20. But these results tend to agree with the result obtained from interferometric technique in a transparent specimen (see Section 3.3.4) and also by others on metallic materials [59].
Figure 20. a – Variation of Closure Stress with Removal of Successive Surface Layers
b – The Effect of Specimen Thickness on Closure [2].
3.3.2 Push Rod Displacement Gage

This technique has been developed and used to determine closure at the interior of a thick specimen as well as at the interior of the part through elliptical crack [43,65] in a plate. Figure 21 illustrates the push rod displacement gage technique. The technique shows that closure at the interior is less than that measured at the surface using a near tip strain gage. These are interesting and important measurements and need careful reconfirmation by other techniques such as removal of the surface layer followed by a recheck of the closure behaviour at the first and subsequent cycles of loading.

3.3.3 Vacuum Infiltration Technique

The use of vacuum infiltration technique [64] in a 2024-T3 Aluminum alloys has shown that the primary closure contact points are in the center and within 1 mm of the crack tip, even though substantial shear lips are present.

Based on the observed sharpness of fractographic features in the shear lip and flat fraction region, it has been contended that closure does not take place at the interior at all and is confined only at the shear lips [62] of a very high strength steel ($\sigma_y = 1677$ MPa). These observations contradict the results of vacuum infiltration technique. Probably, the difference is related to the yield strength of the materials investigated.
Figure 21. Push Rod Displacement Gage Technique for the Determination of Near Tip Interior Closure of an Elliptical Crack [65]. Note that the Technique can also be Used for a Through-The-Thickness Crack.
3.3.4 Interferometric Technique in Transparent Specimen

Monochromatic light interference fringe patterns produced at the fatigue crack during the loading of a transparent polymethylmethacrylate CT specimen were examined [58] to provide three-dimensional measurement of crack surface displacement with a resolution of 0.25 μm. In the unloaded condition, the crack is closed at the surface of the specimen and also along the curved front within the surface layers. But the crack is open at the interior. The results suggest that the closure effect or $K_{op}$ observed on the specimen surface would be higher than those observed in the interior of the specimen.

Whereas the results support the expected trend that $K_{op}$ observed near the crack tip at the surface would be higher than those observed at the interior of a specimen with a curved crack front, it must be noted that a material such as polymethylmethacrylate undergoes viscous flow. Thus, one may question the applicability of these results to an elastic-plastic material such as a metal. Unfortunately, only one such investigation could be located and such investigation cannot be carried out with metal.

3.4 COMMENTS ON THE DETERMINATION OF CLOSURE

Apart from the techniques discussed above, some other techniques -- such as photoelasticity [75], enlargement of photographs of the crack tip [74], direct observation of the crack tip under optical microscope [36] and SEM [55], fatigue crack growth rate at different R-values [48,76], and electron fractography [94] -- have also been used to determine closure.
The more mature and established methods for the determination of bulk closure appear to be based on the CMOD, back face strain gage, and the use of a special displacement gage such as Elber gage. The value of $K_{op}$ determined by CMOD and back face strain gage technique tend to agree with each other in most instances. However, with uncertainties and contradictions in $K_{op}$ determination being what it is, it is preferable to determine bulk $K_{op}$, using simultaneously two different techniques. In fact, important aspects of closure behaviour should be confirmed by simultaneously using two different techniques.

The offset procedure can be used to identify $K_{op}$ from the signals obtained from these three techniques. The offset displacement technique is very sensitive and therefore, it can detect the last fraction of closure. However, the determination of a thickness averaged 'local' closure behaviour from the offset technique needs refinements of procedure. Even if such refinements are accomplished, it is not clear if the local $K_{op}$ determined from the offset technique would differ from the near tip interior and surface $K_{op}$ values.

In the case of plasticity induced closure, the difference in the roles of the plastic wake and the crack tip plastic zone in producing closure should be characterized. The offset displacement procedure can be improved and developed to identify the distinction between the closures produced by these two factors. The plastic wake probably produces bulk closure - whereas crack tip plastic zone and reverse plasticity could have a predominant influence on the local closure. The effect of reverse plasticity is important
and should be visible in experiments where a high-low load sequence leads to crack growth retardation; on the other hand, the effect of plastic wake is probably relevant in the case of constant amplitude loading where the effect on R on the da/dN has to be taken into account. In order to distinguish between the two closures, one could machine the material in the plastic wake by EDM and study the effect of its removal on closure detected by the offset procedure on da/dN. This could also help ascertain the origin of the upper and the lower closure points [32] wherein the former has been used to explain overload retardation effects.

The offset displacement procedure using either a differential amplifier or a microprocessor is sensitive and can easily identify the transition point. However, some workers [3] tend to identify the closure load at some point higher than the transition point. No definite procedure is reported to identify such closure points and it appears arbitrary. The offset procedure needs refinement and standardization to minimize the effect of friction, misalignment, and out-of-plane bending. Besides, the characterization of the load versus offset displacement plots and loops needs to be standardized. In the absence of these, even a sensitive technique such as offset procedure is not free from uncertainties.

The more attractive and reliable methods for near tip surface closure measurements appear to be the techniques such as interferometric displacement gage and replication followed by SEM measurements. Surface closure is expected to be higher than the closure at the interior and this has been confirmed by different investigators. However, to clearly
distinguish between the two, the more reliable technique is the progressive removal of surface layers from the specimen and the study of its effect on closure. Such a method, even though laborious and slow, is quite worthwhile in view of the contradictions in the closure behaviours reported by the different investigators.

The push rod gage technique [43,65] appears to be an interesting method of determining near tip interior closure behaviour, however, one has to shift gage location as the crack grows and drill more holes in the specimen. As pointed out above, the most certain method of determining the role of surface versus near tip interior bulk closure is to machine surface layers and measure closure by the offset procedure. Such an experiment can be used to distinguish between the bulk, the local, the near tip interior, and the near tip surface closure behaviours. In order to isolate the effects of a curved crack front and that of the thickness, one could produce specimens with a straight crack front using special notches [17].

Most of the closure determination studies report $K_{op}$. Yet, to understand, characterize, and predict closure, it is essential to measure the extent of closure and the residual displacement due to closure. Only a few studies report the extent of closure and residual displacement due to closure.

The use of terms such as plane stress or plane strain is confusing. In relation to closure studies, the experimental situations which correspond to two distinct groups of behaviour need to be clearly ascertained. Thus,
the effects of geometry, thickness, width and crack length on $K_{op}$, extent of closure, and residual strain (or displacement) require a careful investigation. During these investigations, the effect of machining of surface layers as well as the material from the wake in a few selected instances would ascertain and distinguish the closure at the surface and at the interior.
4. PHENOMENOLOGICAL STUDY OF CLOSURE

In this section, some observations concerning the effect of different variables on closure will be presented and discussed. Most of these facts are experimentally determined and only a few are numerically obtained. Most investigations report the effect of the different variables on $K_{op}$. Since the effect of the variables on the two aspects of closure, that is, the extent of closure and residual displacement produced by closure are rarely reported, the effects of these variables on $K_{op}$ alone will be discussed. The effects of some of these variables are extensively studied, yet some of them are hardly investigated. The effect of the following variables on $K_{op}$ are examined in this section:

1. $K_{max}$, $K_{min}$, and $R$ Under Constant Amplitude Loading

2. Overload

3. Short Crack Behaviour

4. Surface Crack Behaviour

5. Residual Stresses

6. Environmental Factors

7. Microstructural and Fractographic Features
Elber [6] proposed that crack growth rates are determined by $\Delta K_{\text{eff}}$ rather than by $\Delta K$. Accordingly, various investigators have shown that the effects of the stress ratio, hi-lo load sequence, the short crack, the residual stresses, and environment on fatigue crack growth rate can be suitably accounted for by substituting $\Delta K_{\text{eff}}$ for $\Delta K$ in the Paris-Erdogan law - see Equations (1) and (2). The applicability of these equations and their ability to account for the effects of these variables will also be briefly examined at the end of the respective sections.

4.1 EFFECT OF $K_{\text{max}}$, $K_{\text{min}}$, AND R UNDER CONSTANT AMPLITUDE LOADING

In any fatigue loading situation, the three variables, $K_{\text{max}}$, $K_{\text{min}}$, and $R$ are interrelated through $R = K_{\text{min}}/K_{\text{max}}$. However, contrary to expectation, even in a relatively simple loading situation such as constant amplitude loading, the three variables exert independent influences on $da/dN$. This shows the inherent complexity of fatigue. A careful examination of fatigue crack growth rate data shows that amongst these three variables, $K_{\text{max}}$ has a very strong and positive influence on $da/dN$ in all types of constant amplitude tests. On the other hand, the influence of $K_{\text{min}}$ on $da/dN$ ranges between zero to positive and that of $R$ between zero to negative, depending on the type of constant amplitude test. It is likely then that these three variables should also influence $K_{op}$ independently; but the influence of an individual variable on $K_{op}$ need not be identical to that on $da/dN$. 

88
The effect of $R$ on $da/dN$ obtained under constant amplitude loading is shown in Figure 22. For a given $\Delta K$, as $R$ increases, $da/dN$ increases. Life prediction would be obviously simple if all the $da/dN$ data for different $R$ values as reported in Figure 22 were to collapse to a single line. Accordingly, several approaches have been used to achieve this. The important ones amongst these are the empirical equations (13) and (14) given below.

Forman Equation [78]

$$\frac{da}{dN} = \frac{C \Delta K^n}{(1-R)} \frac{\Delta K}{c - \Delta K}$$

(13)

Walker Equation [2,79]

$$\frac{da}{dN} = f\left(\frac{\Delta K^n}{(1-R)^m}\right)$$

(14)

These equations have been used with some degree of success. However, in recent times, the closure based concept is increasingly being used to achieve the desired collapse of data. The closure based concept is more popular since the concept has a physical basis and can account for the effect of many other variables on $da/dN$.

Since the three variables are interrelated, the determination of $K_{op}$ could be undertaken in three different types of constant amplitude test situations, as given below:
Figure 22. Effect of Load Ratio, $R$, on $da/dN$ [2].
1. R is constant, $K_{\text{max}}$ and $K_{\text{min}}$ vary

2. $K_{\text{max}}$ is constant, $K_{\text{min}}$ and R vary

3. $K_{\text{min}}$ is constant, $K_{\text{max}}$ and R vary

Most of the investigations belong to the first type; only a few investigations which belong to the second or the third type have been undertaken so far. In fact, it appears that a comprehensive investigation of the effect of all these three different types of tests on $K_{\text{op}}$ has not been carried out.

One must also note that most of the closure studies are carried out under load control test conditions. In a load control test, $P_{\text{max}}$ and $P_{\text{min}}$ are held constant and, therefore, $K_{\text{max}}$ and $K_{\text{min}}$ increase as the crack length increases. The $K_{\text{op}}$ values obtained from such tests are usually plotted against $K_{\text{max}}$ presuming that $a/W$ has no effect on $K_{\text{op}}$ at constant $P$. In order to ascertain that crack length has no effect on $K_{\text{op}}$, it is necessary to perform $K$ control tests. This will be discussed later.

In the subsequent paragraphs, the effect of $R$, $K_{\text{min}}$, and $K_{\text{max}}$ on experimentally observed $K_{\text{op}}$ during the three different test situations, are discussed successively.
4.1.1 The Effect of \( R \) and \( K_{\text{min}} \)

The effect of \( R \) on \( K_{\text{op}} \) has been studied by several investigators \([6,24,32,44,66,67,68,70]\) first amongst these being Elber [6] who proposed the relationship

\[
U = \frac{K_{\text{max}} - K_{\text{op}}}{K_{\text{max}} - K_{\text{min}}} = 0.5 + 0.4 R \quad (15)
\]

Substitution of \( \Delta K_{\text{eff}} = U \cdot \Delta K \) and the Equation (15) in Equation (1) enables us to evaluate the effect of \( R \) on \( \text{da/dN} \). Based on their respective experimental results, Equation (15) has been modified by other investigators to forms such as \( U = 0.68 + 0.91 R \) [67] and \( U = 0.707 + 0.408 R \) [70]. Also, it has been reported that the value of \( U \) ranges from as low as 0.4, to as high as 0.85 for plane strain and 0.75 for plane stress [66,67]. Values of \( U \) as high as 2 have been reported [70,73] in the literature. However, it has been pointed out [125] that \( U \) values greater than 1 represent a contact free crack surface and, therefore, have no physical significance in fatigue crack growth.

Experimental data shows [63,69,70] above, that at a certain value of \( R > R_c \), further increase in \( R \) has no effect on \( \text{da/dN} \). It is also observed that at \( R > R_c \), \( K_{\text{op}} < K_{\text{min}} \). To represent fatigue crack growth rate using Equations (1), (2), and (15) at \( R > R_c \), \( K_{\text{op}} \) is assumed to be equal to \( K_{\text{min}} \) and, therefore, \( U = 1 \), and accordingly, Equations (1), (2), and (15) give a \( \text{da/dN} \) which is independent of \( R \) as observed experimentally. However,

92
it has also been argued [73] that by assuming $U = 1$ in situations where $K_{op} < K_{min}$, one uses the same $da/dN$ versus $\Delta K_{eff}$ relationship even though the crack tip opening ranges are quite different. And it has also been reported [36] that at $R > R_c$, $da/dN$ increases with increasing $R$. Thus, the effect of $R$ on $U$ and on $da/dN$ is unclear.

Substitution of $K_{min} = R \cdot K_{max}$ in Equation (15) gives

$$\frac{K_{op}}{K_{max}} = 0.5 + 0.1 \cdot R + 0.4 \cdot R^2$$

(16)

Equation (16) should be independent of $K_{max}$ or $K_{min}$; or, in other words, Equation (16) should hold true for all the three different types of tests. However, as will be discussed in Section 4.1.2, $K_{op}/K_{max}$ is observed to depend on $K_{max}$ and a plot of $K_{op}/K_{max}$ versus $K_{max}$ (or $a/W$) shows a transition [29,32].

According to Equation (16), at constant $K_{max}$ and positive $R$ values, $K_{op}$ should increase with $R$ and the minimum value of $K_{op}/K_{max}$ should be 0.5. On the other hand, $K_{op}/K_{max}$ has been reported [24,31,32,43,52,65-68] to be ranging in values from 0.15 to almost 1, even at positive $R$ ratios. Yet contrary to all these, in a recent investigation, it has been suggested that $K_{op}/K_{max}$ is independent of $R$ [32].
Elbers original tests were performed with CCP specimens prepared from aluminium alloys. It is hard to infer from these results to what extent specimen size, material, the type of test and test control parameters, and the method of $K_{op}$ determination have influenced the trends of $K_{op}$ results reported above. Indeed, such discrepancies in the experimental observations on $K_{op}$ discredit a plasticity induced closure mechanism. However, as will be discussed shortly, the discrepancies could have alternative explanations.

In a recent investigation [67] where the tests were of the second type, the effect of varying R and $K_{min}$ on $K_{op}$ has been investigated at $K_{max} =$ constant. An examination of the data shows that a sevenfold increase in $K_{min}$ and R, raises $K_{op}$ only by 40%.

4.1.2 The Effect of $K_{max}$

As regards the effect of $K_{max}$ on $K_{op}$, the bulk of the investigations [3,6,29-34,44,48,66-68,70,71,73] have been carried out using type I test, where R is constant and both $K_{max}$ and $K_{min}$ vary. This is achieved by a load control test where $K_{max}$ and $K_{min}$ increase as the crack length increases. Such test results, discussed later, show apparently three different patterns: (a) $K_{op}/K_{max}$ is almost independent of $K_{max}$ [3,6,32,70] as shown in Figure 23a, (b) $K_{op}$ is independent of $K_{max}$ at constant R, even when $K_{max}$ is in the Paris regime as reported in Figure 24 [4,31,34,48], and (c) $K_{op}/K_{max}$ decreases systematically as $K_{max}$ increases as shown in Figure 25. Such results have been obtained by several investigators [10,18,30,47,66,71,72], particularly when $K_{max}$ values are in the threshold region.
Figure 23. Variation of $K_{op}/K_{max}$ with $a/w$ or $K_{max}$ for (a) $R=0.5$ and (b) $R=0.05$ [32].
Figure 24. $K_{op}$ has no systematic dependence on $K_{max}$. Each data point from individual CT specimens of $B = 4$ mm, $W = 50$ mm, and varying $a/W$ values [31,34].
AISI 1018
SAE 4135
2219 T87
P/MMAS7 T6
R = 0.05
f = 30 to 50 Hz
The observation that $\frac{K_{op}}{K_{\text{max}}}$ is independent of $K_{\text{max}}$ at constant $R$ and high $K_{\text{max}}$ values (see Figure 23a) implies that $K_{op}$ increases with $K_{\text{max}}$. Such an observation is consistent with the mechanism of plasticity induced closure irrespective of whether $K_{op}$ originates from the plastic wake or the plastic zone ahead of the crack tip. However, as reported above, the bulk $K_{op}$ has also been observed to be independent of $K_{\text{max}}$ at constant $R$ and high $K_{\text{max}}$ values [4,31], particularly in a situation where the $K_{\text{max}}$ is changed after the crack has grown with a certain history of loading. At low $R$ values, sometimes it undergoes transition [32] to a lower $\frac{K_{op}}{K_{\text{max}}}$ value as shown in Figure 23b. Such a transition is also reported by others [29,32,66]. Usually, such results are obtained in a load control test when $K_{\text{max}}$ is in the Paris regime.

The history of loading probably has a strong influence on $K_{op}$ but this has not been systematically investigated. If the prior history of loading were to exert a major influence on the bulk $K_{op}$ value, it is more likely that the main contribution to bulk $K_{op}$ originates from the plastic wake rather than the plastic zone ahead of the crack.

It was pointed out while discussing the first pattern of results that in the Paris regime tests, $\frac{K_{op}}{K_{\text{max}}}$ undergoes a transition at higher $a/W$ values (see Figure 23b). Such results are obtained under load control and at higher $K_{\text{max}}$ values with $R$ = constant. If the plastic wake is important in producing closure, then obviously, the crack length vis-a-vis the length of the wake in relation to the spread of the wake in the $y$ direction should play a significant role in governing the $K_{op}$ value. The effect of even a fairly wide plastic wake could decrease if the wake is
located at a far off point from the load line of the specimen. The observed transition in the $K_{op}/K_{max}$ value is, therefore, not inconsistent with plasticity induced closure. In view of the above, it is necessary to isolate the effects of $K_{max}$ and $a/W$ in producing closure. Such isolation is possible if closure is determined at different $a/W$ values, in a test where $K_{max}$ and $R$ are held constant. Such an investigation can also help to evaluate the effect of the plastic wake on $K_{op}$, a point which is particularly important in view of the observed dependence of plastic zone on $a/W$, recently reported by several investigators [1/121-124].

On the other hand, a systematic decrease in $K_{op}/K_{max}$ with increasing $K_{max}$ (see Figure 25) is considered inconsistent with Equation (16). In fact, this unusual pattern of closure behaviour is cited as important evidence that asperity or oxide induced closure is operative instead of plasticity induced closure at low $\Delta K$ regions close to the threshold value.

While this logic appears reasonable, one should not overlook the following three points.

First, a close examination of the pattern of decreasing $K_{op}/K_{max}$ with increasing $K_{max}$ of the type shown in Figure 25 would indicate that $K_{op}$ is either independent or is only mildly dependent on $K_{max}$. In fact, this trend agrees with the recent observation that $K_{op}$ is practically independent of $K_{max}$. In one of these studies [31,34], $K_{max}$ or $\Delta K$ was increased by 100 to 200% and the crack was grown for considerable length. Even then the change in bulk $K_{op}$ was insignificant (see Figure 24). The results reported in other studies [4,48] confirm the above. Obviously, a change in...
$K_{\max}$ or $\Delta K$ changes the plasticity only at the crack tip; apparently, this change has only a minor influence on $K_{op}$. Even though plasticity induced mechanism is discounted [4] since $K_{op}$ is observed to be independent of $K_{\max}$ (see Figure 24), one should not ignore the possibility that bulk $K_{op}$ probably originates from the plastic wake, and the wake previously developed is loading history dependent and such history effect on bulk $K_{op}$ can override the effect produced by a change in a local crack tip plasticity during constant amplitude loading for some duration of crack extension.

The second point concerns the very high $K_{op}/K_{\max}$ values observed in the threshold regime. The $K_{op}/K_{\max}$ value is nearly equal to 1. To determine threshold, one normally precracks at a higher load and once the crack has grown outside the notch field, the load-shedding is started in order to determine $\Delta K_{th}$. Thus, the threshold test as usually run, produces a characteristic plastic wake. The plastic wake established during precracking loading history could continue to determine $K_{op}$ over a significant length of subsequent crack extension during a threshold test. Under such circumstances, if $K_{op}$ remains constant, $K_{op}/K_{\max}$ could be nearly equal to 1 as $K_{\max}$ decreases during load shedding. If $K_{op}$ is as significantly history dependent as speculated here, could there be a minimum $K_{op}$ corresponding to the $K_{th}$, as postulated? Carefully conducted experiments may answer such a question.

The third point concerns the explanation of the observed pattern that $K_{op}/K_{\max}$ decreases with $K_{\max}$. As discussed above, most threshold tests are conducted following a systematic load-shedding pattern as crack length increases. After the threshold is determined, if load is
kept constant, the $K_{\text{max}}$ value increases; but the initial plastic wake produced at low $K_{\text{max}}$ value during threshold tests, may continue to determine $K_{\text{op}}$ value for quite some distance, thus producing a low $K_{\text{op}}/K_{\text{max}}$ at higher $K_{\text{max}}$.

Since the relative roles of plastic wake and monotonic plastic zone in producing the closure at different $a/W$ values is not known, all the above discussion is speculative. But the main point is that one should not ignore the effect of loading history, that is, the change in $\Delta K$ or $K_{\text{max}}$ with increasing crack length to suitably explain the observed facts through plasticity induced closure phenomenon.

The reported differences in the effects of $K_{\text{max}}$, $K_{\text{min}}$, and $R$ on $K_{\text{op}}$ by the various investigators exhibit no systematic pattern. A comprehensive experimental programme can sort out the contradictions and anomalies which partly originate from the method of determination, partly from the type of tests, and partly from the material, specimen size, and geometry.

4.1.3 Normalization of $da/dN$ Data Using $K_{\text{op}}$

Several investigators [2,6,66,68,70] have reported that the effect of $R$ on $da/dN$ can be normalized if $da/dN$ is plotted against $\Delta K_{\text{eff}} = K_{\text{max}} - K_{\text{op}}$ instead of $\Delta K$. One such result is reported in Figure 26. Yet there are some others [26,44] who report that crack closure cannot fully account for the effect of $R$ on $da/dN$. 

101
Figure 26. Plot of \( \frac{da}{dN} \) Versus \( \Delta K_{\text{eff}} \), Showing the Normalization of the Effect of \( R \) on \( \frac{da}{dN} \) as Reported in Figure 22 [2].
Indeed, it is intriguing that so many investigators find such good correlation between $\frac{da}{dN}$ and $\Delta K_{\text{eff}}$. When one considers the wide divergence in the observed effects of $K_{\text{max}}$, $R$, and $K_{\text{min}}$ on $K_{\text{op}}$, the inherent uncertainties in the determination of $K_{\text{op}}$, our ignorance as to which $K_{\text{op}}$ (bulk, local, interior, or surface) really controls $\frac{da}{dN}$, and the contradictions as to the precise mechanism by which closure decreases $\frac{da}{dN}$, $\Delta K_{\text{eff}}$ appears to be a rather powerful parameter. But to what extent a log-log plot with a permissible scatterband of 2 conceals any discrepancies that would have been otherwise observed, is hard to say. On the other hand, it is clear that the values of the constant $A$ and $m$ should change significantly if one were to change the basis and approach to closure determination. Under these circumstances, $A$ and $m$ can have hardly any physical significance even though considerable research is done to relate them to the basic material behaviour.

Instead of the anticipated straightline, a plot of $\frac{da}{dN}$ versus $\Delta K$ is sometimes reported to be curved - sometimes concave [2,66] and sometimes convex with sharp bends [69]. Thus, in such cases, obviously Equations (1) and (2) fail to represent the $\frac{da}{dN}$ data. The question is, do these effects manifest because closure is ignored in the representation of such data. On the other hand, the opposite question is equally valid. If for a given set of data, a plot of $\frac{da}{dN}$ versus $\Delta K$ yields a straightline, would a plot of $\frac{da}{dN}$ versus $\Delta K_{\text{eff}}$ also produce a straightline?
4.2 OVERLOAD EFFECTS [27,32,36,55,65,80-94]

The constant amplitude loading fatigue crack growth studies discussed in Section 4.1 are useful in understanding and characterizing fatigue crack growth. However, what is encountered in service is variable amplitude loading often with an ordered sequence such as a hi-lo sequence producing the so-called load-interaction. This is referred to as an overload effect and it causes significant crack growth retardation. As a result of such effects, life prediction based on the crack growth data obtained through constant amplitude loading can be overly conservative. Thus, even though the characterization of overload effects are most important for a precise life prediction, it is a complex phenomenon and is not well-characterized. For the sake of simplicity, in this section, the discussion will primarily concern the overload effects produced by a single cycle step-load even though other loading patterns such as block overloads, compressive loads in compression-tension load cycle, or compressive loads in tension-compressive load also produce significant load-interaction effects.

The overload interaction effect is experimentally observed when a load excursion in the form of a single overload is applied during constant amplitude fatigue crack growth at a baseline $\Delta K$ level. The effect of the single cycle overload on $\frac{da}{dN}$ versus $N$ and $\frac{da}{dN}$ versus $a$ at constant $\Delta K$ and $R$ is shown in Figure 27. Before the overload, the $\frac{da}{dN}$ has a constant value at region A. The overload has the following effects on $\frac{da}{dN}$: (1) accelerated $\frac{da}{dN}$ due to overload from the baseline level as in A (this acceleration of $\frac{da}{dN}$ is not shown in Figure 27, since $\frac{da}{dN}$ for a single
The crack growth rate at the overload cycle is not represented in the figure.

Figure 27. Variation of da/dN as a Result of Single Cycle Overload. The Crack Growth Rate at the Overload Cycle is not Represented in the Figure.
cycle is difficult to determine), (2) a short lived initial acceleration of $da/dN$ at the baseline $\Delta K$ immediately after the overload, as in B, (3) a decrease in $da/dN$ to a minimum as in C which is followed by an increase in $da/dN$ to the baseline $\Delta K$ level as in C through D, and (4) continued growth with a $da/dN$ value identical to that before the overload, as in D. Also indicated in Figure 27 is Z, the delayed retarded zone and $a^*$, the overload retardation zone. Whereas several investigators [80,83,85,94] show that $a^*$ is nearly equal to the overload plastic zone size, sometimes in plane strain and sometimes in plane stress, many report [65,90] that $a^*$ is four to five times larger than even the plane stress plastic zone size.

Several empirical models [95-98] have been proposed to explain the overload effect. A comparison of the experimental data and the results of calculations based on the models proposed by the various investigators are shown in Figure 28. The agreement between the models and experiments is not good. Part of this originates from the scatter in $da/dN$ data. But basically, the factors which cause the observed decrease in $da/dN$ are not properly understood.

Attempts have been made to explain the observed effects of the overload using the following concepts: (1) crack tip strain hardening [99,100], (2) crack branching or deflection [19,88], (3) fracture surface microroughness [19], (4) residual compressive stresses introduced due to overload ahead of the crack tip [6,96], and (5) crack closure [2,27,32,65,85,92,94].

106
Figure 28. Experimental Crack Propagation Behaviour as Influenced by Overload Application — Comparison with Prediction by Various Models. The Inset Gives Details of Overload Application [84].
Even though crack tip strain hardening can explain some of the observed overload effects, it has also been shown [80] that the phenomenology of retardation behaviour is independent of whether the material is strain hardening or softening during cyclic loading. Recently, crack branching has been proposed to explain part of the observed retardation [19,88]. But the extent of such retardation could depend significantly on the geometry of loading, such as the presence of a bending moment. The role of fracture surface microroughness in producing closure and decreasing da/dN had been examined earlier in Section 2.2. While microroughness can be expected to play some role in producing closure at low ΔK values near the threshold region, the specific combination of $K_{\text{max}}$, $\alpha\gamma$, size, and geometry at which the role of plasticity becomes negligible and the roughness plays a predominant role in producing closure and retarding da/dN is not clear. Besides this, the position is somewhat confusing. Because, whereas one group of investigators [84] show that a rougher fracture surface produced by intergranular fracture would decrease $K_{\text{op}}$ and accelerate da/dN following overload, others [7-12] contend that roughness could produce more closure and decrease da/dN.

Residual compressive stresses ahead of the crack tip due to overload plastic zone and behind the crack front due to the plastic wake are produced by a mechanism discussed in Section 2.1. It is necessary to distinguish between them conceptually, such as considered earlier in Sections 2.1 and 3.4 in terms of 'bulk' and 'local' closure. Experimentally also, one can possibly distinguish between the two. The 'local' closure could be produced mainly by the residual compressive stresses ahead of the crack and therefore, this may play a predominant role in explaining overload retardation effects. On
the other hand, under constant amplitude loading, the distinction between the bulk and local $K_{op}$ may disappear and the bulk $K_{op}$ produced by the residual compressive stresses in the plastic wake behind the crack front is as good in explaining the observed effect of $R$, $K_{max}$, and $K_{min}$ on $da/dN$.

An examination of Figure 27 would obviously show that in order to explain the observed pattern of the change of $da/dN$ following the application of overload, $\Delta K_{eff}$ should increase at first, and $K_{op}$ should decrease to account for the short-lived initial acceleration of $da/dN$. The short-lived initial acceleration marked B in Figure 27, is not observed sometimes. However, what always follows is a drop in $da/dN$ to a minimum (see C in Figure 27) followed by a continuous increase in $da/dN$ till it reaches its original value. These changes in $da/dN$ should correspond to an increase in $K_{op}$ to a maximum followed by a gradual decrease to the original $K_{op}$ prior to overload. Since plastic wake does not change with overload, it is difficult to conceive that closure produced by the plastic wake can give rise to such initial sharp decrease and then a sharp increase in $K_{op}$ followed by its steady decrease. In fact, the bulk $K_{op}$ is not observed to change when $K_{max}$ is changed during overload [31], implying that bulk $K_{op}$ probably originates from the plastic wake. One, therefore, has to take into account the role of overload plastic zone formed ahead of the crack in modifying the residual stresses and displacement. The overload plastic zone can generate residual compressive stresses which have no correspondence with the compressive stresses in the plastic wake. It can, therefore, give rise to compressive stress humps at the peak load spot as postulated [125,126,138] and directly supported by the experimental measurements of near tip displacements [84] and indirectly supported by the identification of an upper closure point [32].
Even though overload retardation has been extensively investigated, only a few investigators [80-94] have used the closure concept to explain the observed effects. Only a few [27,32,65] have measured the $K_{op}$ values before and after overload and correlated the resultant $\Delta K_{eff}$ with $da/dN$. Figure 29 reports one such result where the agreement is good. More impressive are the predicted versus observed $da/dN$ preceding and following an overload for a thumbnail shaped part through crack on the surface of a specimen which, as discussed later, has a pair of closure and $da/dN$ values, one corresponding to the surface and the other corresponding to the maximum depth point in the interior. This is reported in Figure 30. Notwithstanding the above, some have reported that $K_{op}$ cannot account for the observed overload effects [44, 43,144].

There is an even more important contradiction. Some investigators [85,94] have produced good correlation between $\Delta K_{eff}$ obtained from Equation (15), and $da/dN$ during overload retardation. As discussed in Section 4.1, Equation (15) has doubtful validity. The others have used bulk measurements such as ultrasonics [27] and obtained excellent correlation between $\Delta K_{eff}$ and $da/dN$. On the other hand, recent $K_{op}$ measurements which correlated $\Delta K_{eff}$ with $da/dN$ very nicely [32,65] are all based on near tip or local closure behaviour of the crack tip measured using highly sensitive instrumentation. It was discussed in Section 3 that local and bulk behaviour in most test situations would give different $K_{op}$ values. Yet a plot of $\Delta K_{eff}$ versus $da/dN$, even in the situation where the estimation or determination of $K_{op}$ has doubtful basis, yield excellent correlation! It seems a log-log plot with the permissible scatterband of 2 can accommodate all such differences. If gross measurements and estimation can as well yield the normalized $da/dN$ data as desired, what is the need for precision measurements?
Figure 29. Crack Growth Rate Versus $\Delta K_{\text{eff}}$ before and during transient effects after a single overload [32].
Figure 30. Comparison of Experimentally Observed $\frac{da}{dn}$ Following Overload with the $\frac{da}{dn}$ Inferred From Closure Measurement and Constant Amplitude $\frac{da}{dn}$ Data for a Thumbnail-Shaped Crack with $a=$Crack Depth and $c=$Surface Crack Length [65].
If the $K_{op}$ originating from the overload plastic zone were to be responsible for the observed overload effects, the determination of $K_{op}$ produced by such small effects at regions close to the crack tip require precision measurements. As discussed in Section 3, such measurements as well as the interpretation of test records, require standardization.

4.3 SHORT CRACK BEHAVIOUR

A component spends a major part of its life when the crack present in it, is short. It is well established that a short crack grows at a rate much faster than that of a long crack [11]. Life prediction could be significantly non-conservative if this is ignored. Therefore, reliable life prediction requires that short crack behaviour be properly understood and characterized.

The fast growth of short cracks is attributed to several factors: first, continuum requirements are no longer satisfied, second, $K$ or LEFM is no longer applicable and finally, closure is much less, in the case of a short as against a long crack [11]. A short crack is hard to define for its length depends on the material and $K$ level at which the crack is grown. However, most experiments report short crack behaviour in the range of 25 $\mu$m to 2 $\text{mm}$ [11].

In order to understand and characterize the fast growth of short crack in terms of closure, some numerical and experimental investigations have been carried out. Newman [101,102] has used Dugdale's strip yield plasticity
analysis to devise a 'ligament model' which permits the calculation of closure in CCT specimen geometry using a finite element technique. The closure has been calculated for a short crack emerging out of a hole in a plate specimen and also a CCT specimen without a hole [101]. Newman did not measure closure but compared the experimental crack growth rate with the crack growth rate predicted for a short crack, taking into account the closure effect as obtained from the finite element calculations. Such comparison for the cases both with and without the hole shows a reasonable agreement [101] as reported in Figure 31. The implication is that the crack closure effect in a short crack is less than that of a long crack.

On the other hand, the experimental determination of $K_{op}$ shows a different trend. The experimental investigations of closure in short cracks are few [56,104]. One of the investigators [56] reports results which indicate that the closure behaviour of a short crack is not radically different from that of a long crack; for example, for a .36 mm crack, at R = 0, $K_{op}/K_{max} \sim 0.6$. In fact, if $K_{op}$ is as high as reported, the closure effect in short cracks cannot explain the high crack growth rates exhibited by a short crack. Accordingly, the applicability of $K$-based representation of fatigue crack growth rate is questionable.

It is difficult to characterize closure for a short crack via the plasticity induced mechanism unless the respective roles of plastic wake and plastic zone as dependent on crack length are properly established. Ritchie has proposed [11,12] that a short crack grows fast since the roughness does not develop in a short crack and consequently, closure is less
Figure 31. Comparison of Experimental and Predicted Crack Growth Rates for Small Cracks Emanating From a Circular Hole in Steel Specimens [101].
severe. Such a proposal is inconsistent with the observation that a short crack when viewed under a microscope shows a zig-zag path and has a roughness parameter and asperity height (see Figure 11), which is quite large compared to the crack length. Similarly, crack deflection [88] cannot explain the fast growth rates observed in short crack since a short crack deflects significantly quite often. And in all probability, the closure of a short crack may not be negligible. The reason for rapid growth of short cracks becomes even more difficult to explain due to the proposed effects of crack closure, roughness, or crack deflection!

4.4 SURFACE CRACK BEHAVIOUR

Most engineering structures fail in fatigue by the growth of a part-through, thumbnail shaped fatigue crack. The shape of the crack changes as it grows and the growth rate of the crack is not the same along the surface and at the maximum depth point in the interior. At the surface, the state of stress is plane stress whereas at the interior, the constraint to plastic flow is high and the conditions are closer to plane strain. If plasticity induced mechanisms were operative, the closure where the crack meets the surface would be stronger than the closure at the maximum depth point. Even in the case of a through-the-thickness crack, a curved crack front is produced at the mid-thickness of a fracture mechanics specimen. Thus, the basic observation on closure in the case of a thumbnail shaped crack may apply to the closure observed in a specimen with a curved crack front.
Recently, two investigators [65,67] have reported results of closure studies on thumbnail shaped cracks. One of them [65] has measured the closure at the interior using a push-rod gage technique and at the surface using a strain gage technique. It is observed that $U_{\text{interior}}/U_{\text{surface}} \sim 1.13$ where $U$ is defined in Equation (15). Thus, the closure at the surface is stronger than that at the interior. The other investigator [67] estimated $U_{\text{interior}}/U_{\text{surface}}$ from the crack growth rate, the observed variation of aspect ratio of the crack as it grew and the experimentally determined $K_{\text{op}}$ at the surface. The agreement between the $U_{\text{interior}}/U_{\text{surface}}$ value obtained by the two investigators is good. In fact, the results obtained have been used to predict the retardation and subsequent growth of a crack following a single cycle overload of the thumbnail shaped crack [65] as reported earlier in Figure 30.

If one assumes that such results are applicable also to the growth of a through-the-thickness crack with a curved front, it has interesting implications. The crack front curvature does not change much during the growth of such cracks and if one were to assume that the closure concept is valid, the implication is that FCGR is faster at the surface than at the interior of a thick specimen. Probably, the difference in the state of stress at the surface and at the interior produces this difference in FCGR.

4.5 EFFECT OF RESIDUAL STRESS

The effect of residual stress on $K_{\text{op}}$ and $da/dN$ is important since closure and its effect are manifested in the residual stress it produces.
The nature and origin of the residual stresses produced by closure are briefly discussed at first. Later, the effect of the residual stresses produced by processing and specimen fabrication on $K_{op}$ and $da/dN$ is considered.

The nature of residual stress due to closure of a CCT specimen was experimentally determined [5] by cutting sections across the specimen and measuring forces to reverse displacement. The finite element formulation for the same geometry reports [136] a stress distribution which is given in Figure 5.

On the other hand, the residual stress distribution in the ligament of a CT specimen has not been comprehensively investigated even though this specimen is widely used by many investigators for closure studies. However, one can speculate the stress distribution pattern in a CT specimen by combining the results of the various investigators as discussed below.

The residual stresses near the crack tip of a CT specimen have been determined by X-ray [105]. The X-ray beam size is rather large and X-ray measures the stresses only at the surface. This, naturally, introduces significant uncertainties in the reported stress pattern as it exists across the thickness and very close to the crack tip. However, at high values of overload, the residual stress pattern produced near the crack tip is as reported in Figure 5. The general pattern of residual stresses as measured in the ligament of the CT specimen prepared from a photoelastic material [106] is similar to that shown in Figure 5. Similarly, the residual back face strain value experimentally determined by the back face strain gage
technique [107,108] is compressive as represented in Figure 5. This is further confirmed by the fact that at a given \( K_{op} \) value, the magnitude of the residual back face strain increases as crack length increases [107]. Thus, Figure 5 is a fair representation of the contact and the residual stresses in the plane of the crack of a CT specimen.

It has been shown [106] by progressive removal of the plastic wake by machining, that in a photoelastic material, the residual compressive stress originates from the plastic wake and this causes closure. Also, a dislocation model is proposed which agrees with the decrease of such residual stress when the length of the plastic wake was progressively decreased by machining [106]. Such results emphasize the importance of the compressive stresses and the fact that the whole length of the plastic wake plays a role in producing closure.

In addition to the residual stresses originating from closure as discussed above, one must consider the effect of the residual stresses introduced in a material during processing and fabrication such as welding, forging, or extrusion. The fatigue crack growth rate in the specimen prepared from such a material could exhibit acceleration or deceleration of crack growth rates depending upon the pattern of residual stresses present.

One investigator has accounted for the effect of residual stress on \( da/dN \) by the modification of the stress intensity factor relationship through suitable formulas [109]. Such a modification changes \( K_{max}, K_{min} \), and \( R \) and one can then use suitable empirical relationship to normalize \( da/dN \) data. Alternatively, one can account for such effects by considering
the effect the residual stresses have on experimentally determined $K_{op}$. It has been observed [110] that $K_{op}$ decreases if the residual stress is tensile and $K_{op}$ increases if the residual stress is compressive over the crack faces. Such effects are obviously similar to that produced by the application of compressive load and tensile overloads, respectively.

In view of the above, as well as the point made earlier that closure produces a given pattern of residual stresses, one can expect that the effect of the residual stress on $da/dN$ can be appropriately represented through its effect on $K_{op}$. In fact, the use of $\Delta K_{eff}$ instead of $\Delta K$ in the case of a specimen which contains residual stress, has been shown [110] to produce FCGR data which agrees quite well with the FCGR data obtained after the material is stress relieved [110]. This is shown in Figure 32. Unfortunately, such investigations are few.

It has also been noted [110] that residual stresses when present across the thickness with tension at the interior and compression at the surface layers, enhances crack front curvature. This is synonymous with lower $K_{op}$ values at the surface layers and higher values at the interior. Interestingly enough, such results are consistent with the results obtained for a thumbnail shaped surface crack as discussed in Section 4.4.

An examination of the effect the residual stress pattern has on the $K_{op}$, can help us to examine the validity of the various mechanisms proposed for closure. For as pointed out in Section 2, the residual stress produced by the plasticity induced mechanism is compressive ahead of the crack tip.
Figure 32. Comparison of $\Delta a/dN$ for Non-Stress Relieved and Stress Relieved Materials. The Agreement is Good when $\Delta K_{\text{eff}}$ is Used for Representing $\Delta a/dN$ Data [110].
whereas it is tensile in the case of oxide and asperity induced closure.
The asperity and oxide induced closure mechanisms appear inconsistent, when
this observation is taken into account with the fact that residual tensile
stress decreases $K_{op}$ and thereby increases $da/dN$.

4.6 ENVIRONMENTAL FACTORS

The effect of the environment on fatigue is very complex and
important. It is therefore extensively studied. The environmental effects
are examined below in general terms as they concern closure. The effects
are different in the near threshold and in the Paris regime.

In the Paris regime, that is, at $\Delta K$ values substantially higher
than the threshold value, the fatigue crack growth rate in steel is highest
in hydrogen sulphide and hydrogen and decreases as the environment is changed
successively to moist air, to dry gaseous environment, and to vacuum [23,28,
39,41,111]. A similar trend with regard to the effect of humidity and vacuum
environment is also obeyed in the case of aluminum and titanium alloys.
These differences have been explained traditionally in terms of hydrogen
embrittlement and accordingly, models based on the appropriate anodic or
cathodic processes at the crack tip and at the flank of the crack have been
proposed. Several mechanisms of hydrogen embrittlement have been advanced to
explain the effect of frequency, etc. But the precise mechanism of hydrogen
embrittlement is not clearly known. On the other hand, attempts have been
made to characterize the observed difference in fatigue crack growth rates in
terms of the $K_{op}$ determined in various environments [39]. Unfortunately,
such investigations are rather few.
$K_{op}$ of aluminum alloys in the Paris regime, has been determined [39,41] in vacuum, dry oxygen (or nitrogen) and also in air with different percentages of relative humidity. It is observed that $K_{op}$ is highest in vacuum and decreases successively as the environment changes to dry oxygen (or nitrogen), to air with 50% to air with 100%, relative humidity, as shown in Figure 33. Correspondingly, the difference in $da/dN$ is accounted for in terms of closure, since a plot of $da/dN$ versus $\Delta K_{eff}$ yields a straight line which is independent of all environments. Contrary to such observations, it has also been reported that in case of titanium alloys, $K_{op}$ does not change even if the environment is changed from vacuum to one atmosphere [4,28]. Thus, the results of attempts to incorporate environmental effects is not conclusive. In fact, it has also been shown that a change from dry to humid environment significantly changes the crack tip strains and crack tip opening displacements in an aluminum alloy [112]. Thus, apart from the $K_{op}$ effects, one should consider local crack tip microplasticity to explain the observed difference in $da/dN$ due to changes in environment.

Regarding the environmental effects in the Paris regime as discussed above, in the threshold and near-threshold regime, the effects are quite different and depend on the $R$ value [13,14,21,22,23]. In the near-threshold regime at low $R$ value, $da/dN$ is highest (and correspondingly, $\Delta K_{th}$ is lowest) in all dry environments and is the same for all gases, be it argon or hydrogen. Thus, $da/dN$ in steel decreases as the environment is changed to moist air [13,14,21]. Since hydrogen and argon exhibit identical behaviour, hydrogen embrittlement cannot explain the observed difference [13,14,21]. The more interesting point is that at high $R$ values, dry hydrogen, dry argon,
Figure 33.  

30a  
Figure 33a.  Effect of Environment on Crack Propagation Rate in 7075-T651 Al Alloy.  

30b  
Figure 33b.  Normalization of the Data Reported in Figure 33a When $da/dN$ is Plotted Against $\Delta K_{\text{eff}}$ [39].
and moist air all exhibit some da/dN and threshold behaviour. This difference at the high and low R values has been explained through the proposed oxide induced closure mechanism. The limitations of this mechanism have been discussed earlier in Section 2.3. Additionally, there is another important aspect which apparently cannot be explained through this mechanism. For instance, in vacuum, the da/dN is slower than in all other environments, particularly at high R-values [21,22,114], in case of steels and aluminum alloys. Obviously, oxide induced closure cannot explain the observed difference in the da/dN for a vacuum and for a dry hydrogen environment since the difference in the oxidation behaviour in these two environments would be rather small.

The oxide induced closure mechanism has been proposed recently. The mechanism should be verified through K_{op} measurements in the different environments in the threshold regime. In fact, no verification in terms of da/dN versus environmentally modified ΔK_{eff} has been attempted in this regime.

The effect of the environment is much too complex by itself. When one takes closure into account, the observed phenomenon becomes too complex to understand and explain. For a better characterization of the problem, further experimental work is necessary.

4.7 MICROSTRUCTURAL AND FRACTOGRAPHIC FEATURES

In the past, it has been generally contended that microstructure has a small influence on the da/dN in the Paris regime and has a significant
influence on da/dN and $\Delta K_{th}$ only in the threshold regime. However, experimental results reported recently show that microstructural features, such as grain size, lamellar spacing, dispersion of phases, solute level, and the inclusion size and shape distribution, can significantly influence da/dN in the Paris regime.

In the Paris regime, the da/dN decreases as the prior austenitic grain size in steel [113,116] or $\alpha$-grain size or the dispersion of phases in Ti-alloy [117,118] increase or the dispersion of $\alpha$ and $\alpha'$ in dual phase steel [88] changes. In fact, an interesting result on the effect of grain size is $\Delta K_T = 5.5\sigma_y\sqrt{d}$ where $d$ = prior austenitic grain diameter [113]. Similarly, an increase in interlamellar spacing [115,116] of pearlite in steel increases da/dN and decreases $\Delta K_{th}$. In the threshold regime, microstructural features influence da/dN for all materials. In general, all microstructural features which promote coarse, planar, heterogeneous, and reversible slip with increased slip length decrease da/dN, particularly at low $\Delta K$ levels. Conversely, features which promote homogeneous, wavy slip produce increased da/dN.

Fractographic features, in general, originate from the microstructure, and one of the features which decrease da/dN is increased crack branching or crack path deflection producing increased roughness and secondary cracking. The crack deflects since the crack propagates along specific crystallographic planes in the grain and therefore, the larger the grain size, the larger the deflections which in turn produce slower da/dN and higher $\Delta K_{th}$. Such
increased slip planarity could also be produced by changing precipitate morphology [118]. It is contended that $K_{op}$ constitutes the bulk of $\Delta K_{th}$. Thus, the microstructural features which increase $\Delta K_{th}$ and decrease $da/dN$ at the threshold regime as discussed above, will correspondingly increase $K_{op}$.

In fact, $K_{op}$ measured in different Ti-alloys has been shown to relate to the roughness [4]. However, the scales of roughness considered in this investigation are in the range of 200 $\mu$m which is three orders larger than the roughness contemplated in the models of roughness induced closure [8,12,60]. As $K_{op}$ increases, the roughness also increases and since such increase has no systematic relationship with yield strengths of the titanium alloys investigated, the plasticity induced mechanism is discounted. However, one should also consider that plasticity depends not only on yield strength, but also on $K_{max}$. The three titanium alloys which had widely differing $\Delta K_{th}$ and therefore, $K_{op}$ in these alloys were determined in different ranges of $K_{max}$ values and, therefore, the three alloys experienced three different loading histories. Thus, $K_{op}$ or roughness could differ even if the $\sigma_Y$ values of these alloys were the same. Thus, the possibility of plasticity induced closure cannot be discounted in these experiments.

Recently, it has been suggested [88] that crack deflection over distances of the order 30-100 $\mu$m at angles ranging from 30 to 70° can retard crack growth rate substantially. Such deflection decreases $K_I$ or even the effective $K$ of \(^{1}\) branched crack and produces the observed decrease in $da/dN$. It is shown that a change in microstructure can produce crack deflection.
and accordingly cause a decrease in \( \text{da/dN} \). Apparently, such a change in microstructure also increases \( K_{op} \). However, according to the author [88], the increased \( K_{op} \) can only partly account for the decrease in \( \text{da/dN} \); and part of the decrease in \( \text{da/dN} \) is accounted for by crack deflection.

Measurement of fractographic features is difficult. For example, a change in striation spacing determined through fractographic observation [77] has been used to identify the closure load. In fact, such confidence in striation spacing measurement is reinforced by a recent investigation where an excellent correspondence between striation spacing and macroscopic crack growth rates was obtained [69]. In fact, these investigators also report a good correlation between striation spacing and \( \Delta K_{\text{eff}} = K_{\text{max}} - K_{op} \) for different \( R \) values [69]. However, others report that the observed correspondence between macroscopic growth rate and striation spacing is poor [7,94].

A change in the slope of a \( \text{da/dN} \) versus \( \Delta K \) plot often corresponds to a change in micromechanism such as faceted to striated growth or transgranular to intergranular [69]. What is interesting, however, is that a plot of \( \text{da/dN} \) versus \( \Delta K_{\text{eff}} \) with closure taken into account appears relatively smooth [69]. It is not clear whether this is the result of \( K_{op} \), appropriately accounting for change of micromechanism or of changed format of data representation.

Microstructural and fractographic features are difficult to quantify and determine. To add to this, the determination of closure has its inherent uncertainties and unanswered questions. Thus, significant phenomenological
study is required before \( K_{op} \) can be related to the microstructural and fractographic features. Such studies are important, for they alone can provide the important clues, for instance, to develop materials where \( K_{op} \) always equals \( K_{max} \), arresting all fatigue crack growth!

4.8 MATERIAL PROPERTIES

The important properties which may influence \( K_{op} \) are Young's modulus, uniaxial yield strength, the strain hardening (both monotonic and cyclic), and strain softening (cyclic). It has been observed [125] that \( K_{op} \) should be independent of Young's modulus. However, this has not been investigated. Similarly, the effect of strain hardening or softening on \( K_{op} \) has not been investigated. Since cyclic strain hardening and strain softening has little effect [80] on the pattern of overload retardation - a phenomenon directly related to closure, it is likely that cyclic hardening or softening has only a secondary effect on \( K_{op} \).

In one of the investigations, the effect of \( \sigma_y \) on bulk \( K_{op} \) has been studied in a test situation where other factors are identical. The result is reported in Figure 34. It appears that \( K_{op} \) decreases as the yield strength increases suggesting obviously that a plasticity induced mechanism is probably operative. One can, of course, argue that the differences in the microstructural features in materials with different yield strengths can produce different fracture surface roughness and, therefore, the corresponding differences in \( K_{op} \) [4]. This has been discussed earlier in Section 4.1.
Figure 34. Effect of $\sigma_y$ on $K_{op}$ in CT Specimens, $B=4$ mm, $W=50$ mm and Varying $a/W$. Scatter in $K_{op}$ Values Reported is Attributed to Different Precracking History [31,34].
4.9 EFFECT OF SIZE AND GEOMETRY

The effect of size and geometry on $K_{op}$ must be evaluated for reliable life prediction. However, little experimental and only limited numerical work has been carried out for a comprehensive evaluation of the problem.

Size refers to dimensions such as the thickness (dimension along $z$ direction) the crack length and the width (dimensions along $x$ direction) and geometry refers to the shape, in general, and in particular, to the ratio of dimensions along $x$ and $y$ direction, to the loading configuration and also to the nature of the far field stress distribution in the body.

In numerical experiments, Newman [103] has varied the effect of the nature of the far field distribution by considering CCT specimens without a hole but with a crack in one case and with crack emerging from a hole in the other case. With closures calculated numerically and $da/dN$ determined experimentally in both the cases, Newman shows that the difference in closure accounts for the difference in $da/dN$. Since $da/dN$ in the case of a crack emerging from a hole would be faster, $K_{op}$, in the case of a specimen with a hole would be lower than $K_{op}$ of the specimen without a hole. However, in order to use such numerical analysis for specimen configurations other than CCT, one requires closed form relationships for $K$ and displacement with a Dugdale plastic zone in the specimen. Unfortunately, such results are not available for all configurations. Besides, such an approach gives results for plane stress or plane strain plasticity only. Since the state of stress

131
in a precracked body is intermediate, some systematic experimental study would be useful. Experimental investigations of the effect of geometry on $K_{op}$ are scarce. However, there is indirect evidence that $K_{op}$ could depend on the geometry. For instance, the overload retardation behaviour of a CCT specimen is reported to be substantially different from that of CT specimen [82] implying that their $K_{op}$ values of a CCT are higher than that of a CT specimen. And, the fatigue crack growth data when plotted as $da/dN$ versus $\Delta K$ often show layering where for a low constraint specimen such as a CCT, $da/dN$ values are higher than that of a CT specimen [120]. Similarly, side-grooving can lower $da/dN$ values. If one takes all these facts into account, it would seem that plots of $da/dN$ versus $\Delta K_{eff}$ for CCP and CT specimens may show increased differences and layering. An investigation of the $K_{op}$ value of CT and CCP specimens should be interesting.

Recent investigations [17,121-124,139-147] have shown that the plastic zone, even when it is small, depends systematically on $W$, $a/W$, and specimen geometry and that the small scale yielding assumption which has been invoked in the analysis of fatigue and fracture quite freely in the past, should be used with caution. One can thus expect $K_{op}$ to depend on $W$ and $a/W$ to some extent. In fact, even if plasticity were to be independent of $W$ and $a/W$, and we assume that $K_{op}$ originates from the plastic wake, one can still expect $K_{op}$ to depend on $a/W$ because the wedging action of the plastic wake for instance, in a CT specimen, would have less of an effect at the load-line as the length of the wake increases. It is likely that the increased plasticity at higher $a/W$ probably compensates the decreased effect of a long plastic wake producing negligible effect on closure unless the crack lengths
are much different or the plastic wake is too small. But these speculations require further investigation. The effect of width on $K_{op}$ has not been reported but using similar arguments, one would expect an effect of width on $K_{op}$. Since plasticity decreases with width [121-124], $K_{op}$ should decrease with increasing $W$ at a given $a/W$. However, this too is a matter of investigation.

The effect of thickness on $K_{op}$ requires three-dimensional elastic-plastic analysis, for, as pointed out in Section 3, preferential yielding in the surface layer can substantially alter the closure response of a precracked body. The evidences for these are several: progressive removal of surface layers progressively decreases $K_{op}$ [59], overload retardation is less in a thick as compared to a thin specimen [83], the introduction of side grooves dramatically decreases the number of delay cycles [119], machining of surface layers reduces the number of delay cycles [142], and the closure at the interior is less than at the surface [43,65,67]. And even though in most fatigue crack analysis, the crack front is idealized in the manner shown in Figure 6, the actual crack front is very complicated - more complicated than the curved crack front reported in Figure 8. The true nature of three-dimensional crack front is evident from Figure 35 as reported in [63,89]. As a result of such a crack front, some out of plane bending and modes II and II are invariably produced at the crack front in any specimen. To what extent, this has produced the contradictions, anomalies, and the usual scatter observed in fatigue data is hard to say. But, when one takes into account all these observations, the problem of the fatigue crack growth and particularly of the closure, indeed, appears too complex to characterize and formulate. Thus, it seems that the first step in resolution of the problem is to generate FCGR and closure data in experiments where modes II and II are practically negligible.
Figure 35. Fracture Surface Profile as Influenced by Specimen Size and Loading (a) Symmetric Profile, (b) Non-Symmetric Profile, and (c) Out of Plane Sliding in Thin Specimens with Non-Symmetric Profile relaxes Compressive Force [89].
(a) MAXIMUM STRESS

(b) MINIMUM STRESS
The above discussion confirms that closure in practical situations is a three-dimensional problem. And it is interesting that in spite of the complex nature of the problem as discussed above, \( K_{op} \) determined from the two-dimensional analysis or from the experiments which are based on two-dimensional considerations can produce such good correlation with \( da/dN \).

4.10 COMMENTS ON THE PHENOMENOLOGICAL STUDY OF CLOSURE

The phenomenological studies concerning closure reveal numerous contradictions, discrepancies, and unexplained facts. However, a few general observations concerning the different aspects of closure as discussed in this section are summarized below.

The basic experimental facts concerning the dependence of \( K_{op} \) on \( K_{max} \), \( K_{min} \), and \( R \) for a given constant amplitude test situation are not clearly established. Unless this is accomplished, it is difficult to formulate closure phenomena and to interpret the results of closure studies. Most studies make little distinction between 'bulk', 'local', near tip surface, and near tip closure behaviours. One wonders if such distinction is relevant since we do not know which one of these closures really control \( da/dN \). This adds to the general confusion.

In spite of this, the use of \( \Delta K_{eff} \) instead of \( \Delta K \), normalizes \( da/dN \) data for wide variations of \( K_{max} \), \( K_{min} \), \( k \), and accounts for the change in the growth rate produced by step overloads, surface crack behaviour, residual stress effects, and the effect of environment in many instances. This is indeed intriguing when one considers (1) the wide divergence that is observed in the results of studies conducted to evaluate the effects of the different
variables, (2) the uncertainties in the determination of $K_{op}$, and (3) our ignorance as to which $K_{op}$ really controls $da/dN$. In fact, the nonlinear plots of $da/dN$ versus $\Delta K$ in the Paris regime are sometimes curved convex and sometimes concave. If closure were to have a systematic relationship with $K_{max}$, it is unlikely that use of the closure concept would remove the nonlinearities in all instances.

$K_{op}$ at the interior is less than the $K_{op}$ at the surface of a thumbnail crack. This needs further investigation. The behaviour of short cracks remains unexplained and requires systematic experimental work for its characterization.

A study of the effect of residual stress on $K_{op}$ and $da/dN$ indicates that the residual compressive stresses introduced at the crack tip during closure decrease $da/dN$. In fact, the introduction of a controlled residual stress pattern and its effect on $K_{op}$ and $da/dN$ can be examined to check the validity of the different mechanisms of closure that have been proposed.

Correlation of microstructural and fractographic features with $K_{op}$ and $da/dN$ is an important area of research, both from the standpoint of developing fatigue resistant materials and in understanding the origin of closure.

The asperity and oxide induced closure mechanisms need to be carefully studied in terms of the residual stress pattern that these mechanisms produce. Such studies can provide a firm physical foundation for a more realistic formulation of the models from which $K_{op}$ can be calculated.
The plasticity induced closure mechanism has some basic discrepancies such as: $K_{cp}$ is observed to increase, remain constant, or even decrease with increasing $K_{max}$ in regimes where admittedly, asperity and oxide induced closure is inapplicable. These and other discrepancies probably arise since the history dependence of closure extends over dimensions which are orders of magnitude larger than the plastic zone size. However, if the history dependence is established through experimental work, other questions must be raised about the customary assumptions made in closure and fatigue studies.

Even though the effects of the material properties and the size and geometry on closure are most important, they have not been studied. Careful experimental work needs to be conducted because recent work has shown that the assumption of small scale yielding is inconsistent with experimental observations. In fact, such studies may indicate new directions in the development and application of more reliable and economic life prediction technology.
5. PREDICTION OF CLOSURE

A reliable calculation of closure is key to reliable life prediction. Based on a given mechanism and model of closure, it should be possible to calculate $K_{op}$ in any precracked body of arbitrary size and geometry made from a given material and subjected to a given history of loading. The various mechanisms of closure were discussed in Section 2 and it was pointed out that the asperity and oxide induced mechanisms require further development in order to use them for prediction of closure. In this section, various methods based on the plasticity induced mechanism for the calculation of $K_{op}$ are briefly examined. As pointed out earlier, the residual displacement in the wake and the residual stresses ahead of the crack would modify the state of stress, strain, and displacement at the crack tip. These modifications have to be taken into account to calculate plasticity and closure and also to understand how $K_{op}$ changes $\Delta K_{eff}$ and $da/dN$.

To facilitate subsequent discussion, the methods of the calculation of $K_{op}$ are classified into two broad categories: (1) Finite element based methods, and (2) Analytical procedures not based on finite elements. The analytical procedures give a better physical insight into the closure phenomenon but can be used only for simple geometries. On the other hand, the finite element method can, in principle, be used for complex loading and geometry but the results may depend upon the element mesh chosen. The analytical procedures are discussed at first.
5.1 ANALYTICAL PROCEDURES

Several groups of workers have used analytical procedures to calculate $K_{op}$. The works of Fuhring and Seeger [125,126], Dill and Saff [127], Budiansky and Hutchinson [128], and Paris and Hermann [32] are discussed very briefly in this section.

Fuhring and Seeger [125] were the first (their work was originally published in German) to use Dugdale's model for the analysis of crack closure which takes into account the stress and displacement variation for every consecutive half-cycle. They examined the details of the residual stress and strain field at the wake of a growing fatigue crack which produce contact stresses and which, in turn, affect the entire state of stress, strain, and displacement in the vicinity of the crack tip. The contact stress calculation is based on the assumption of a trapezoidal distribution of residual displacement along the length of the crack. The contact stresses have a large gradient and this is taken into account in discretization of the crack axis into nodes and line elements. The stresses and displacement are solved from Dugdale's results using an iteration procedure.

The results show that under constant amplitude loading with constant load, $K_{op}/K_{max}$ is almost independent of $K_{max}$ for aspect ratios less than 0.7. The value of $U$ increases with $R$. The analysis also evaluates the extent of closure and the effect of a reversed cycle ($R < 0$) on closure.
The results obtained by Fuhring and Seeger are for plane stress. Even though the analysis evaluates the effect of aspect ratio and periodic spacing, 2b, the results apply to a remotely stressed infinite sheet with an array of colinear cracks with length 2a and periodic spacing 2b. The applicability of such results to finite specimens with geometries such as CT or three point bend is yet to be examined.

These workers have used a similar approach to predict crack growth rates under variable amplitude or sequence loading which are presumed to be governed by memory rules [126]. The main trend of the results of Fuhring and Seeger have been verified [84] by displacement measurements in the vicinity of the crack tip using a grid technique.

Dill and Saff [127] have used a simple contact stress model of closure. The approach is similar to the one used by Fuhring and Seeger [125]. They evaluated the stress intensity caused by crack surface contact which, in turn, produces an interference between the mating crack surfaces. The interference is determined from an analysis of the elastic displacement during loading, elastic displacement during unloading, and the permanent deformation left in the wake of a growing crack. A Dugdale model is used to make an integral equation formulation of the closure condition. The interference is treated as a wedge acting behind the crack tip and the contact stresses created by this wedge are calculated by idealizing the wedge as 25 constant stress elements which experience only compressive contact stress which have a maximum value of δγ. An influence coefficient matrix is developed for the stress-displacement relationship between elements using a weight function.
approach. The stresses give contact stress intensity at the minimum load and therefore, the $\Delta K_{\text{eff}}$. The results of the model show good agreement with Elber's experimental results as given in Equation (15). The predicted crack growth from this contact stress model compares quite well with the experimentally determined crack growth rates obtained under spectrum loading. It was pointed out in Section 4.1 that Equation (15) does not agree with experimental results of several other investigators. To this extent, the general applicability of the model is doubtful.

The work by Budianisky and Hutchinson [128] calculates crack opening load, crack closing load, and residual plastic stretch at the crack tip by a complex function method under the assumption of small scale yielding according to the ideally-plastic Dugdale-Barenblatt model. The results are reported for a range of load ratios. It is observed that in contrast to the stationary crack which suffers reverse plastic flow over a distance which is $k$ the monotonic plastic zone size, the corresponding reverse plastic zone size for a growing fatigue crack is $1/10$ of the monotonic plastic zone. The $K_{\text{clo}}/K_{\text{max}}$ is $0.48$ and $K_{\text{op}}/K_{\text{max}} = 0.56$. The analysis provides justification for the adoption of an effective stress intensity range to represent fatigue crack growth rate. However, the analysis is appropriate only for plane stress; on the other hand, the plane-strain condition applies over most of the crack tip region. Thus, even though the results given some valuable insight into closure and guidelines for future work, the results reported here are not expected to agree with experiments.
Paris and Hermann [32] have reported an approximate relationship of $K_{op}$ which incorporates some of the features of the closure phenomenon. Using the results of this relationship, they predict the variation of $K_{op}/K_{max}$ following overload which agree very well with the experimentally determined variation of upper closure point. The relationship indicates that $K_{op}/K_{max}$ should be independent of $R$. Since the nature of this relationship is quadratic, Paris and Hermann have hinted that the transition in $K_{op}/K_{max}$ experimentally observed at $a/W \sim 0.55$ to 0.6 during a constant amplitude test is consistent with the two $K_{op}/K_{max}$ values that would be obtained from such a quadratic equation. However, whether the transition is a phenomenon inherent to closure or is merely a result of the history dependence of $K_{op}$, is a matter of experimental investigation.

5.2 FINITE ELEMENT BASED METHOD

Finite element based methods of the calculation of $K_{op}$ of three different groups of workers are discussed below: Ogura, Ohji, and Ohkibu [129,130], Nakagaki and Atluri [131], and Newman and his coworkers [132-137].

Ohji, Ogura, and Ohkubo [129] and Ogura and Ohji [130] have used incremental finite element elastic-plastic analysis with a fine constant strain triangle mesh. Results were obtained for a double edged notched specimen which explained the experimentally observed non-propagating fatigue cracks ahead of a notch using the concept of closure. These workers have also investigated the effect of overload on closure and residual stresses. It was observed that the closure stress at first decreases and then increases.
to a maximum and then drops again. Such a variation of closure stress obviously can explain the observed variation of crack growth rate following retardation as discussed in Section 4.2. Overload produces residual compressive stresses and the region of such stresses increases as the overload increases. They also show that compressive load just following an overload decreases $K_{op}$ which corresponds to the usually observed increase in the experimental $da/dN$ in such cases. The analysis was done for a double edge notched specimen under plane stress condition. It is not clear as to how the contact stresses generated in the plastic wake were taken into account in their analysis.

Nakagaki and Atluri [131] have used circular-sector shaped hybrid elements centered at the crack tip with the HRR stress and strain singularities embedded in the special elements near the crack tip. The procedure is computationally inexpensive due to (1) the use of such elements which permit adoption of rather coarse mesh and (2) the use of a procedure wherein the elastic part of the structure is isolated and its stiffness remains unchanged during loading. In this analysis, the stresses at opening and closure are determined for a CCP specimen and they agree quite well with each other. $\Delta K_{eff}$ is calculated from $\sigma_{cl}$ obtained for constant amplitude and other sequence loadings. In all cases, $\Delta K_{eff}$ is observed to follow a pattern which is commensurate with the observed patterns of $da/dN$ during such load sequencing. The crack surface deformation profiles are different for different load sequencing and this controls the resultant residual stresses and possibly the growth retardation. Obviously, the approach has computational advantages. It also takes into account blunting, which is experimentally observed during overloading. However, it is not clear as to how the contact stresses in the plastic wake are taken into account in their analysis.
The most extensive and comprehensive finite element based analysis of closure has been done by Newman and coworkers [132-137]. First, to investigate closure using finite elements [132], Newman [136,137] has developed and verified a crack closure model that simulates plane stress and plane strain conditions by using a 'constraint factor'. The analysis is based on a Dugdale model which was modified to leave plastically deformed material in the wake of the advancing crack.

A schematic of the crack surface displacement reported in Figure 36 shows a plastic region of length \( p \) and a residual plastic deformation along the crack surfaces. These regions are composed of rigid perfectly plastic (constant stress) bar elements with a flow stress \( \sigma_0 \). The bar elements are broken in the region of residual plastic deformation and carry compressive load only. The physical crack is of half length \( c \) which on unloading, produces a reverse plastic zone, \( w \). The governing equations for the crack closure model were set up using the stress intensity factor and crack surface displacements for a CCP specimen and then solved for the contact stresses and the stresses in the plastic zone. The opening stress, \( S_o \), was calculated by equating the stress intensity factor due to an applied stress increment \( (S_o - S_{\text{min}}) \) to the stress intensity factor due to the contact stresses. The calculations were made both for plane stress and plane strain conditions, that is with varying constraint factor. A constraint factor of 2.3 was chosen for further calculation of life prediction since it gave a good correlation under constant amplitude loading.
The model gives results wherein $K_{op}/K_{max}$ is in the range of 0.25 to 0.35 and is independent of crack length. The plane stress values $K_{op}/K_{max}$ were higher, that is in the neighbourhood of 0.5 to 0.6. With increasing stress level, the plane stress $K_{op}/K_{max}$ decreased.

The model was used to correlate crack growth rates under constant amplitude loading and to predict crack growth under aircraft spectrum loading on 2219-T851 aluminum alloy plate material. The predicted crack growth lines agreed well with experimental data obtained from 80 crack growth tests subjected to various load histories. The ratio of predicted to experimental lives ranged from 0.5 to 1.8.

Even though the choice of a constraint factor of 2.3 has no fundamental basis, the predictive capability of the model is quite encouraging. The applicability of this model to other geometries and size needs further investigation.

5.3 COMMENTS ON PREDICTION OF CLOSURE

It is obvious from the discussion above that analysis of closure is not sufficiently developed, to enable prediction of closure in a precracked body of arbitrary size and geometry. Since the plasticity in the surface layers plays an important role in determining fatigue crack growth following overload, the problem is obviously three-dimensional. Thus, even though two-dimensional analysis of various sizes and geometry would help us to achieve a better understanding and characterization of closure, the development
of a satisfactory methodology of prediction of closure would probably await the development of three-dimensional elastic-plastic analysis of precracked body. However, it should be emphasized that a systematic experimental study can produce a better comprehension of the problem and this can eventually help us to simplify the three-dimensional problem to something which is manageable.
6. CONCLUDING REMARKS

1. The asperity and oxide induced closure mechanisms as proposed produce residual tensile stresses near the crack tip whereas the crack tip residual stresses as experimentally observed and as considered in plasticity induced mechanism are compressive. These different residual stress patterns cannot give the same \( \frac{da}{dN} \) even though the resultant \( \Delta K \)'s may be equal. This contradiction can be resolved if the models based on asperity and oxide induced closure mechanisms are reformulated to produce compressive stresses over the whole wake of the crack. The asperity and oxide induced closure mechanisms require further experimental work and development before they can be used for the analysis and prediction of closure.

2. One should distinguish between bulk, near tip surface, near tip interior, and possibly local closure behaviours during the experimental determination of closure. It is necessary to ascertain which of these closures actually controls \( \frac{da}{dN} \) in different experimental situations.

3. Crack mouth opening displacement gage, back face strain gage, and Elber clip gage are the most widely used for the experimental determination of bulk \( K_{OP} \). The use of the offset procedure either through the use of a differential amplifier or a microprocessor can increase the magnification and sensitivity of closure determination very significantly. However, the advantages gained from such a procedure cannot be realized unless there is a corresponding improvement in minimizing friction, misalignment, and out of
plane bending during loading of the specimen and in the measurement of
displacement. In addition, the procedure for identification of the closure
load from the P-ΔV plots needs to be standardized. This may help us to
identify the difference between bulk and local $K_{op}$ wherever such a difference
exists.

4. The more attractive and reliable methods for the determination of near
tip surface closure are the interferometric displacement gage and the repli-
cation followed by measurement with a scanning electron microscope.

5. The push rod gage technique appears to be an interesting method of
determination of near tip closure at the interior. However, the most certain
method of ascertaining the difference in the closure at the interior and
surface is to determine closure by offset procedure after progressive removal
of the surface layers by machining. An analysis of the closure behaviours
so determined on specimens of different initial thickness can provide
important insight into the role and origin of closure. Similar investigation
of the variation of $K_{op}$ with progressive machining of the plastic wake along
the crack plane can help us to distinguish between the different closures and
the roles they play in decreasing da/dN.

6. The use of terms such as plane stress and plane strain is confusing
since most experimental situations correspond to a state of stress which is
intermediate. Besides, the experimental conditions of closure studies which
correspond to these two extreme behaviours need to be clearly defined and
established.
7. In order to understand and characterize closure, it is necessary to study and report the extent of closure and the residual displacement due to closure in addition to the $K_{op}$ values usually reported in most investigations.

8. The basic experimental facts concerning the dependence of $K_{op}$ on $K_{max}$, $K_{min}$, and $R$ are not clearly established. Besides, there are several contradictions and discrepancies. Even then, the use of $\Delta K_{eff}$ instead of $\Delta K$ accounts for the change in growth rate produced not only by wide variations of $K_{max}$, $K_{min}$, and $R$, but also by the other factors such as step overload, surface crack behaviour, residual stress effects, and the effect of environment. To what extent a log-log plot with a permissible scatter band of two conceals any discrepancies that would have been otherwise observed is hard to say.

9. Residual compressive stresses increase $K_{op}$ and decrease $da/dN$ while the effect of residual tensile stress is the opposite. This should be taken into account in formulating any mechanism of closure.

10. Correlation of microstructural and fractographic features with $K_{op}$ and $da/dN$ is an important area of research both from the standpoint of developing fatigue resistant materials and in understanding the origin of closure.
11. The plasticity induced closure mechanism has some basic discrepancies such as: \( K_{op} \) is observed to increase, remain constant, or even decrease with increasing \( K_{\text{max}} \) in regimes where asperity and oxide closures, as postulated, are inapplicable. These and other discrepancies probably arise since, in most experiments, the effect of the history dependence of closure which could extend over dimensions that are orders of magnitude larger than the plastic zone size is ignored.

12. Systematic experimental investigation is required to evaluate the effect of material properties, size, and geometry on closure. Machining of surface layers after overload application or sidegrooving increases post overload crack growth rates very significantly. Thus, the surface layer and thickness influence \( K_{op} \). Similarly, recent investigations show that plastic zone can depend on \( W \) and \( a/W \). These results underline the importance of such an investigation.

13. Analysis of closure is not sufficiently developed to predict closure in a precracked body of arbitrary size and geometry. Satisfactory prediction of the real world problem requires three-dimensional elastic-plastic analysis. However, two-dimensional elastic-plastic analysis and systematic experimental work can produce a better comprehension of the problem and eventually help us to simplify the real world problem to something which is manageable.
REFERENCES


