Service load fatigue damage — a historical perspective

Paul C. Paris a,* , Hiroshi Tada a , J. Keith Donald b

a Washington University, St. Louis, MO 63130, USA
b Fracture Technology Associates, Bethlehem, PA, USA

Abstract

Although a general understanding of many aspects of fatigue crack growth behavior was established in the early 1960s, a specific 'accumulation of damage model' for computation of growth under a wide variety of service loads was lacking. The control of growth rates by $K$, the crack tip stress intensity factor and its reversing plastic zone, was well understood but somehow crack closure was overlooked until W. Elber's astute observations in the late 1960s. Various 'service load damage models' were proposed in the 1970s which at best successfully augmented extensive testing under simulated service load conditions. More recently, the crack closure/finite element model of J. Newman has shown the greatest promise. However, the precise cyclic load crack growth data of J.K. Donald has shown limitations of the 'crack opening based $D_K$' as the most appropriate damage criterion. A simple partial crack closure model will be explored herein to attempt to better understand the near-threshold effects observed. Although this new model, as an augmentation of Newman's model, shows great promise, both of these models require measurements of opening loads which are avoided by Donald's 'adjusted compliance ratio' and other approaches. Further possibilities for modeling will be discussed in this historical context to attempt to anticipate future developments. Certainly, some simplification through physical understanding is still required. The recent development of a simplified view of cyclic crack growth behavior by R.W. Hertzberg will be cited as an example of ultimate relevant simplification. Optimism that better understanding will lead to yet better and simpler 'accumulation of service load damage models' will be emphasized. © 1999 Published by Elsevier Science Ltd. All rights reserved.

Keywords: Fatigue; Damage tolerance; Crack growth; History

1. Historical introduction

Prior to the 1960s fatigue crack growth rates were not well analyzed for purposes of safety issues in structural situations such as aircraft. However, it was recognized that advantages could be gained if methods of analysis could be developed to predict the rates of crack growth. An early model analysis was attempted by Head [1], but it did not supply a useful method of prediction for applications. Frost [2] analyzed test data of limited extent and concluded that growth rates were controlled by the parameter $\sigma'$ which was neither precisely correct nor useful for a wide variety of structural crack configurations. Much of their difficulty can be understood by noting that test data were extremely limited by the lack of versatile testing equipment and that convenient crack stress analysis methods had not been developed. McEvilly [3] by the late 1960s provided a wide range of crack growth data on two aluminum alloys and correlated his data using a parameter based on Neuber's notch analysis which was somewhat awkward for adaptation to crack analysis. However, despite the fact that it was the first successful correlation, he has not been given sufficient credit in the literature for that work. Rapid progress did not begin until the work of Paris [4] on applying the Irwin [5] crack stress analysis method to fatigue crack growth and the development of attainable versatile servo-controlled hydraulic testing systems.

In 1957, as a faculty summer associate at Boeing-Seattle, Paris suggested that fatigue crack growth rates could be correlated using the elastic crack tip stress intensity parameter, $K$ [6], and that data so represented could be related through this parameter to predict growth rates in structural cracks from laboratory data for the material and environment of interest. However, lacking test equipment to try the method, it was not until 1959 that data became available to verify that this method did work for a wide range of crack growth rates (i.e. $10^{-7}$ to $10^{-2}$ inches per cycle) from three independent sources.

* Corresponding author.
on two materials. The paper written on that work at that time was not published until 1960 [4], since it was delayed by rejection by three journals (ASME, AIAA, and Phil. Mag.). Though that method is widely accepted today, in the late 1960s at Boeing it was rejected by an outside review panel for federal supersonic transport exploratory studies as ‘it simply won’t work’. Moreover, the federal agency funding the most extensive fatigue studies on multiple occasions stated ‘no interest’ in such work, although since 1970 they have funded more work than any other source. It was a study of rejection by authority with preconceived notions and blind self-interest, with a total reversal after more than 10 years. This was an interesting personal and historical lesson on ‘radical’ discoveries.

Within the Boeing Transport Division the relevance of being able to analyze fatigue cracking rates of structures was immediately recognized up to the top levels of engineering management and was given every encouragement and priority status for development. A small group under the management of W.E. Anderson (e.g. [7]) devoted a majority of their time to the task. Well before the mid-1960s much was known by this group. Load ratio effects had been well documented, the correlation of data from various metal alloy bases by $\Delta K/E$ was noted, environmental influences had been initially explored, overload delay effects were recognized, etc. and it was also recognized that an accumulation of damage model or method was very much needed. Lacking an accumulation of damage model for computational purposes, it was recognized that programmed load tests could be substituted and still correlated to structural applications through the crack tip stress intensity factor methods. During the early 1960s variable band width random load fatigue crack growth rate testing was even performed by S.H. Smith [8] at Boeing.

However, the development of a predictive accumulation of damage method was elusive. The trial model presented in Paris’s dissertation [9] can only be acknowledged as embarrassingly naive, if of interest at all. At least it served to indicate recognition of the need, which has yet to be fully resolved some 36 years later! Much of the further discussion here is devoted to that need.

2. Further early progress

During 1965 at Lehigh, B. Lindner’s M.S. thesis [10] (also [15]) presented experimental data on 7075 aluminum alloy down to growth rates approaching $10^{-10}$ inches per cycle, showing a leveling of the curve below a rate of about $10^{-8}$, which implied a threshold $\Delta K$ for fatigue crack growth. Later work [11] verified that the threshold had indeed been observed. This discovery came as a result of using already developed methods of observing fatigue crack growth with assistance in attaining higher speed testing by H.R. Hartmann and R. Churchill of MTS Systems Corporation. Yet, later testing by R. Schmidt [12] with electrodynamic equipment first showed the distinct $\Delta K$-dominated and $K_{\text{max}}$-dominated ranges of thresholds with variation in load ratio, $R$. Although observing thresholds adds justification to truncation of the lower load end of many service load spectra, it does not otherwise contribute to development of an accumulation of damage analysis or rule.

3. The next big discovery

Early in 1968 Paris had the pleasure of being taken for dinner by a group of advanced students in Germany. One of the students, W. Elber was eager to discuss some of his observations indicating that crack closure due to interference of opposing surfaces occurs during tensile portions of load cycles. The immediate response was that this was an extremely important discovery and the reason explaining various hitherto anomalous behaviors.

However, he was not happy with the indication that his experiments did not dismiss other possible explanations and that he needed to perform further critically definitive tests. He did so prior to initial publication of his landmark work [12,13], which was done without ambiguity. It was also not readily accepted since it was a shock to many ‘established’ concepts which needed revision in its light. Since then, a large body of supporting evidence has been compiled, leaving little doubt.

On the other hand, with respect to formulating an accumulation of damage model or method, it has created as many problems as it has solved. In particular it requires measurement or prediction of closure levels in order to be incorporated in crack growth calculations. Prediction is the subject of much effort with little agreement on its physical behavior. Further measurements of closure remain subject to wide differences caused by instrumentation and definition of the closure point or load even under cyclic loading conditions, let alone under variable amplitude loading. Nevertheless, it is an important feature of cracking behavior which must be addressed or cleverly avoided by methods as yet undiscovered. From the beginning it was shown by Elber and others that the crack tip ‘action’ would be the load range between the closure load and maximum load in a cycle. This was shown to be reasonably true for higher rates of crack growth well away from threshold (where it was easiest to provide data to test the idea). It was some time later that closure levels for variable amplitude loadings were addressed. Schijve [14] explored this area and Paris [15] also looked specifically at the transient closure immediately following an overload. This work at relatively high $K$ levels showed that growth rates were correlated using the transient opening
load to maximum load in computing the relevant range of $K$. This indicated that there was no apparent major material memory for the prior strain history in the crack tip plastic zone. Making use of this effect would require a way to predict the transient closure levels after overloads or within a program of variable amplitude loads. On the other hand, Schijve suggested that it might be possible to predict a steady state average closure level to deal with most variable amplitude loads.

4. Variable amplitude prediction methods

In 1970 the loss of an F-111 resulted in the requirement for a predictive model for fatigue crack growth for the load spectrum of that aircraft. Using no credit for the delaying effect of high loads in the spectrum would have resulted in life predictions which would have been too conservative for practical use. As a result of the need, Wheeler [16] developed a method based on reduced rates of crack growth for cracks crossing a plastic zone left by an overload (high load in the spectrum). It contained an empirical exponent which was determined by comparison of predictions with spectrum load tests. It was simplistic in its representation of physical reality but served its purpose in making calculations of minimum life with potential flaws present for safety purposes. The Wilenbourg model [17] developed later was a more sophisticated but still empirical method not much better than Wheeler’s. Neither of these methods acknowledged crack closure effects and were thus lacking in representing the physical realities of fatigue crack growth phenomena.

A much more realistic finite element model of the circumstances has more recently been developed by Newman [18]. It employs a strip yield type plastic zone for leaving residually stretched material in the wake of the crack, causing interference between crack surfaces or so-called plasticity-induced closure. Some researchers have questioned the assumption of plasticity-induced closure, especially for plane strain applications. However, this model is far better than other empirical models for life prediction, although it has some assumptions which appear to be subject to further question. Indeed, no model will ever be a perfect representation of reality, and every model can be superseded by a new improved one. Currently, Newman’s finite element model is the best we have and worthy of future improvements.

5. Near threshold observations of closure and growth behavior

In the near-threshold regime of crack growth rates, Donald [19] has recently observed a lack of correlation of growth rates using closure to maximum load based $K$ ranges. (Although at higher rates that method produces adequate correlation, as noted earlier.) He shows that crack tip action affecting growth rates must be taking place below opening load! This observation is directly shown by the test data, which has been obtained with the best of precision and taken for a wide variety of load ratios with aluminum alloys. These effects are most significant only near threshold, since such effects are minimized at higher growth rates where much lower closure loads relative to maximum load occur. These observations provide a vastly improved view of the relationship of closure loads to crack growth rates. Since crack growth life accumulates most of its cycles at or near the lowest growth rates involved, this is a very important discovery for improving life prediction! It is the opinion of his co-authors that Donald’s discovery is as important as the discovery of closure itself.

It remains to experimentally explore these effects with variable amplitude loading in the near-threshold regime. Furthermore, it remains to develop physical models and explanations of this effect discovered by Donald and therefore a new model with initially promising results for cyclic loading will be presented herein.

6. A new partial crack closure model exhibiting ‘Donald’s effect’

The effect of crack tip action below crack closure load levels motivates one to reconsider historical discrepancies in the physical model of closure. We must admit that these discrepancies were only recalled after the development of the new model presented here. They are:

1. Bowles [20] in his doctoral dissertation observed that fatigue crack surfaces interfere, but not at the tip, even if compressive loads are applied. After this new model was explained to him, Bowles [21] reacted in a letter “the crack closes behind the tip, not at the tip”! He made these observations from plastic replicas of the crack opening in a plane strain region at various load levels within a cyclic loading.

2. More recently, Vasudevan [22] and others have strongly objected to the concept of plasticity-induced closure for plane strain conditions. They argue that with no volume change in plane strain plasticity there should be no net residual material sticking off of the crack surface. However, it is also noted here that they do not object to crack surface interference caused by mismatched roughness. Schijve, in Ref. [14] and earlier papers, adopted similar interference interpretations. (We note that although this is sometimes regarded as a limitation of Newman’s finite element model, assuming plastically stretched material can indeed model the effects of purely roughness-induced interference, since both the roughness and stretch of
elements will be proportional to the so-called ‘residual crack opening stretch’.

3. The residual crack opening stretch left by cyclic load, as described by Paris [23, 15], does not go to zero upon unloading or even reversed loading if crack surface interference occurs.

In summary, all three of these observations are consistent with assuming a closure or crack surface interference model which does not close at the tip of the crack. In addition to a physically opened crack tip, at minimum load crack tip plasticity creates an effective crack length equivalent to further opening beyond the tip.

With the above ‘afterthoughts’ in mind it is easier to accept the physical picture of the partial crack closure model as depicted in Fig. 1a. The crack surface roughness generated by residual crack opening stretch by a ‘sliding-off’ mechanism might mismatch behind the tip, but not immediately at the tip due to various effects, for example, uneven residual stresses in the plastic wake of a propagating crack. Fig. 1b shows the corresponding mathematical model for equivalent elastic computational purposes. It depicts an added layer, 2h, of material inserted into a smooth crack, but not to the tip, to model the interference caused by roughness. The end of this layer is regarded as being at a distance, d, from the effective crack tip. The problem is then to determine the value of the crack tip stress field intensity parameter, $K_{\text{eff}}$, at minimum load in the cycle. Referring to the Tada [24] Stress Analysis of Cracks Handbook, pages 3.11 and 3.12, it is noted that:

$$K_{\text{min-eff}} = \frac{E' h}{2 \pi d} + \sigma_{\text{nom}} \sqrt{\frac{\pi d}{2}}$$  \hspace{1cm} (1)$$

where $\sigma_{\text{nom}}$ is the nominal uniform stress that would be present at minimum load if the crack were absent, as noted on Fig. 2a. On the other hand, for the opening of the crack, refer to Fig. 2b. Presuming d is small enough that the final contact point is within the crack tip stress field, the opening displacements in that region would be parabolic according to Tada [24] (p. 1.3) and thus the opening $K$ is:

$$K_{\text{open}} = \frac{E' h}{2 \sqrt{\frac{\pi}{2d}}}$$  \hspace{1cm} (2)$$

Combining Eqs. (1) and (2) leads to:

$$K_{\text{min-eff}} = \frac{2}{\pi} K_{\text{open}} + \sigma_{\text{nom}} \sqrt{\frac{\pi d}{2}}$$  \hspace{1cm} (3)$$

Notice that the minimum effective $K$ for the load cycle is approximately reduced by the factor $2/\pi$ from the opening $K$, since it is presumed that $d$ is very small. It is of special interest that this reduction factor is independent of both $h$ and $d$!

It is therefore suggested that perhaps the effective range of $K$ between its real minimum and maximum during a fatigue load cycle is:

---

**Fig. 1.** (a) A physical representation of conditions. (b) The computational elastic model.
\[ \Delta K_{\text{eff}} = K_{\text{max}} - \frac{2}{\pi} K_{\text{open}} - \sigma_{\text{nom}} \sqrt{\frac{\pi d}{2}} \]  

(4)

Since \( d \) is small but unknown an obvious further suggestion is to simply drop the final term from Eq. (4) to get the approximate result:

\[ \Delta K_{\text{eff}} \approx \Delta K_{2/\pi} = K_{\text{max}} - \frac{2}{\pi} K_{\text{open}} \]  

(5)

This result gives a slight overestimate of the \( K \) range. In order to bound our results the \( 2/\pi \) factor could be applied to the range of load from minimum load to opening load to also find an underestimate of the effective \( K \) range, which gives:

\[ \Delta K_{\text{eff}} \approx \Delta K_{2/\pi} = K_{\text{max}} - \frac{2}{\pi} K_{\text{open}} - \left(1 - \frac{2}{\pi}\right) K_{\text{min}} \]  

(6)

Both of these approximate forms for the range of \( K \) separately have interesting features and both can be determined by measuring opening loads. Together they tend to bound the effective range of \( K \), if this model is relevant. In order to contrast them with the apparently true range of \( K \) they are shown diagrammatically on Fig. 3. Of course it should also be noted that for high load ratios, \( R \), where closure does not occur, the range of \( K \) should simply be computed from the applied maximum and minimum loads.

A limitation of this model is that their bounding of
effective $K$ ranges requires measurements or predictions of opening loads, which are avoided by some of Donald’s other methods [19]. On the other hand, having a simple and possibly realistic physical model associated with the proposed ‘partial closure model’ is a distinct advantage. But before going further, it seems appropriate to test the model against the available cyclic loading data from Donald [19,25].

7. Crack growth rate correlation using the new partial closure model

A significant amount of data on five aluminum alloys with load ratios, $R$, from −1.0 to 0.7 has been produced in the Fracture Technology Associates Laboratory [19,25] including analysis of load displacement records to determine closure loads as well as growth rates. For each of these materials the data have been plotted using four methods for evaluating the range of the effective applied $K$. They are:

(a) the ‘appl’ method using the actual applied minimum and maximum loads for the range, i.e. $\Delta K$ in Fig. 3;
(b) the ‘op’ method using the measured opening load to maximum for the range, i.e. $\Delta K_{\text{open}}$ in Fig. 3, etc.;
(c) the ‘$2/\pi$’ method using the minimum load, the opening load, and the maximum load to evaluate the $K$ range according to Eq. (6);
(d) the ‘$2/\pi 0$’ method according to Eq. (5).

Fig. 4a–d (as defined above) show data on 6013 alloy for load ratios, $R=−1.0, 0.10, 0.30, 0.50$, and 0.70. For at least the highest load ratio, 0.70, no closure occurs. It can be noted that strikingly better correlation of data occurs using the partial closure model, i.e. Fig. 4c and d. Figs. 5 and 6 show similar data on 2324 and 7055 alloys, respectively, which also demonstrate better correlation using the partial closure method, again plots (c) and (d) for each alloy. Indeed, it appears that the most simplistic $2/\pi 0$ plots, i.e. (d), show slightly better correlations for each. However, the reader is warned that...
Fig. 5. (a) 2324 with $\Delta K_{\text{appl}}$. (b) 2324 with $\Delta K_{\text{open}}$. (c) 2324 with $\Delta K_{2/\pi}$. (d) 2324 with $\Delta K_{2/\pi^0}$.

Further adjustments for maximum $K$ sensitivity might favor the other partial closure method. Without bias for either partial closure model, dramatic improvements are noted in correlation compared with the (a) and (b) plots for all of these alloys. A further two alloys, 6061 and 2024, are plotted on Figs. 7 and 8, respectively, but each with only two load ratios, $R=0.10$ and 0.70. Since at the higher load ratio the data are closure free the 'appl' method was used on all plots (a) through (d). For the lower load ratio closure is always present and is thus compared by the various methods for more detailed correlation results. The data are self-explanatory.

Donald [19] has developed other stress intensity range correlation parameters which give equally good correlations compared with the partial closure model. However, at this time simple physical models have not been discovered as their bases. We shall not rush to suggest one or the other as 'best'. Further critical experimentation and evidence would be prudent before a final choice is made. Meanwhile, the 'partial closure model' remains a leading candidate for further exploration.

Finally, perhaps that judgment should be made based on which model best assists correlation and understanding of variable amplitude behavior. For example, can the simplest ‘$2/\pi^0$’ model be used with transient closure levels following overload or for spectrum loading by just adjusting closure loads by the $2/\pi$ factor for computing the effective load range? If so, might it be possible to incorporate such a result into finite element schemes such as Newman’s to make better predictions of crack growth? ‘Stay tuned for the next exciting episode!’

In any eventuality, the subject of development of an accumulation of damage model for a wide variety of service load conditions remains for improvement. Some further suggestions here are that based on prior history, continuum analytical models have been very productive in understanding fatigue crack growth behavior; more than was or could have been expected prior to 1960. The success of using elastic stress intensity factors over the full range of growth rates compels further simple modeling attempts. After over 40 years of trying to develop the damage rule for fatigue crack growth for variable
amplitude situations it is still appealing to search for a simple solution. Such an example has come to light recently in the work of Hertzberg [26] which can be compared with the partial crack opening model results here.

8. An ideally simple law predicting cyclic load crack growth rates

Paris’s first graduate student, R.W. Hertzberg, arrived with a Master’s degree in Metallurgy from M.I.T. to study fatigue crack growth at Lehigh University. Telling him that Anderson’s observation of $\frac{D K}{E}$ correlating data from various metals showed the lack of importance of metallurgical structure, was done to bother him. Some 30 years later [26], he continued to ponder that fact and put together the following results. He was amused by it having been called the ‘Hertzberg (Paris/McClintock) Law’ of fatigue crack growth. McClintock [27] independently developed related results.

In addition to knowing Anderson’s observation for many years, Hertzberg also knew that leveling to a threshold occurs when the crack growth rate is equal to a Burger’s vector, $b$, per cycle. Combining this with knowing that the effective $K$ range is the important parameter, he first predicts the threshold corner at:

$$\frac{da}{dn} = b \text{ and } \frac{\Delta K_{\text{eff}}}{E b} = 1$$

Next he noted that further dimensional analysis and observing growth rates approximately proportional to $\Delta K_{\text{eff}}^3$ leads to:

$$\frac{da}{dn} = b \left( \frac{\Delta K_{\text{eff}}}{E b} \right) \frac{1}{\frac{\Delta K_{\text{eff}}}{E}}^n \left( \frac{\Delta K_{\text{eff}}}{E} \right)^3$$

The first opportunity arises here to test this ‘Hertzberg’s

Law’ against data, where the effective $K$ range is determined by the partial closure model.

Figs. 9 and 10 show the data for 6061 and 2024 alloys, respectively, for a load ratio, $R$, of 0.10 for which significant closure effects are observed indicating that the method of analysis sizably affects the position of the data. For each of these figures both partial closure methods are shown as:

1. the $2/\pi$ method;
2. the $2/\pi_0$ method.

On each of Figs. 9 and 10 the Hertzberg Law line is plotted starting with the lower star at the predicted threshold corner point and extended for three log cycles of rates with the appropriate slope of 3 and arbitrarily ended at the upper star just to demonstrate that slope. Needless to say, the agreement between the data and the prediction line is quite satisfactory! When it is recalled that the prediction line location depends only on the elastic modulus, $E$, and the Burger’s vector, $b$, of the material, it can only be viewed as an amazingly good predictor. Therefore the ‘Hertzberg (Paris/McClintock) Law’ stands as an example of the power of simple continuum models to describe fatigue crack growth behavior. Hopefully, the also simple partial closure model will be equally successful with the test of time.

9. Prospects for improvements in service load fatigue damage methods

Returning to the major objective of this discussion, it remains to consider the prospects and future approaches to accumulation of damage methods for variable amplitude loading. The preceding discussions of the partial closure model and the Hertzberg Law serve this objective by indicating the possible types of approaches which may be fruitful. It seems relevant, for example, to try something as simplistic as applying the $2/\pi$ factor on
closure loads to get effective $K$ ranges in current closure-based damage methods, such as Newman’s finite element method or others. Further, it may be possible to develop new analytically based methods, such as the residual crack opening stretch method initiated 20 years ago [15,23] but never fully exploited. That method and others might also benefit from the incorporation of partial closure model results or other methods developed by Donald [19,25]. Analysis improvements have developed only slowly in this area.

Further, the current superior experimental capabilities for measurements of closure effects and automated data analysis make it seem timely to perform further tests to assist in developing the analysis models. Especially tailored programs of critical experiments to explore the results of various modeling possibilities seem in order including single and multiple overload tests as well as service load spectrum tests with measurements of transient closure levels and transient crack growth rates. Although many of such tests have been done in the past, they have not been designed to specifically address questions for the development of analytical models.

Above all, an objective here has been to show that fruitful modeling can be kept simple, and necessarily so if it is to be useful in promoting better understanding and universally productive usage.

In summary, the suggestions here can be condensed into the following rules and starting points.

1. First rule: KEEP IT SIMPLE!
2. Try continuum analytical models.
3. Explore whether transient effective stress intensity ranges really control transient growth rates.
4. Attempt two-dimensional plane strain modeling first at near-threshold levels, since this is the most relevant applications area for service use.
5. Continue to review ultimate simplifications such as Schijve’s suggestion of assuming and determining steady state mean closure levels.
6. After 40 years of trying and mainly addressing details,
let’s stick to solving the major problem of a reasonable yet simple accumulation of damage model.

Acknowledgements

The authors wish to thankfully acknowledge the invitation of Dr. A. K. Vasudevan of the Office of Naval Research to prepare and present this discussion at Second Engineering Foundation Conference on Fatigue Damage of Structural Materials (1998). The joint sponsorship of the test program for the data shown in this discussion by the ONR and Fracture Technology Associates is also gratefully acknowledged.

References


