NBS SPECIAL PUBLICATION 423

U.S. DEPARTMENT OF COMMERCE/National Bureau of Standards

Mechanical Failure: Definition of the Problem

MFPG
20th Meeting

MFPG
Mechanical Failure: Definition of the Problem

Proceedings of the 20th Meeting of the
Mechanical Failures Prevention Group,
held at the
National Bureau of Standards,
Washington, D.C. 20234
May 8-10, 1974,

Edited by

T. R. Shives and W. A. Willard

Metallurgy Division
Institute for Materials Research
National Bureau of Standards
Washington, D.C. 20234

Jointly Sponsored by

NBS Institute for Materials Research, Washington, D.C. 20234
Office of Naval Research, Arlington, VA 22217
Consumer Product Safety Commission, Washington, D.C. 20207
Frankford Arsenal, U.S. Army, Philadelphia, PA 19137
Naval Air Development Center, Washington, D.C. 20362
American Society for Mechanical Engineers, New York, NY 10017
Federal Aviation Administration, Washington, D.C. 20591
NASA, Goddard Space Flight Center, Greenbelt, MD 20771

U.S. DEPARTMENT OF COMMERCE, Elliot L. Richardson, *Secretary Designate*
James A. Baker, III, *Under Secretary*
Dr. Betsy Ancker-Johnson, *Assistant Secretary for Science and Technology*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Acting Director*

Issued April 1976

Library of Congress Catalog Card Number: 76-600003

National Bureau of Standards Special Publication 423

Nat. Bur. Stand. (U.S.), Spec. Publ. 423, 228 pages (Apr. 1976)

CODEN: XNBSAV

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1976

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (Order by SD Catalog No. C13.10:423) Stock Number 003-003-01451-6 Price $6.10 cents (Add 25 percent additional for other than U.S. mailing).

Abstract

These proceedings consist of 20 submitted papers and discussions from the 20th Meeting of the Mechanical Failures Prevention Group held at the National Bureau of Standards, May 8–10, 1974. The central theme of the proceedings is the definition of the problem of mechanical failure, with emphasis on modes of failure, consequences of mechanical failure, and implications of mechanical failure.

Key words: Brittle fracture; economics of failure prevention; fatigue failure; mechanical reliability; stress corrosion cracking.

Units and Symbols

Customary United States units and symbols appear in many of the papers in these Proceedings. The participants in the 20th Meeting of the Mechanical Failures Prevention Group have used the established units and symbols commonly employed in their professional fields. However, as an aid to the reader in increasing familiarity with and usage of the metric system of units (SI), the following table of conversion factors and references are given:

Length
- 1 inch = 2.54 centimeters (cm)
- 1 foot = 0.3048 meters (m)
- 1 mile = 1.609×10^3 meters (m)
  1.609 kilometers (km)

Mass
- 1 pound (lbm) = 0.4536 kilogram (kg)
- 1 ton (short, 2000 lbm) = 907.2 kilograms (kg)

Stress
- 1 psi = 6.895×10^3 pascals (Pa)
- 1 ksi = 6.895×10^6 pascals (Pa)

Velocity
- 1 mile per hour (mph) = 0.4470 meter per second (m/s)
  1.609 kilometers per hour (km/h)

Volume
- 1 gallon = 3.785×10^{-3} cubic meters (m^3)

Volume/Time
- 1 cubic foot per minute (cfm) = 4.719×10^{-4} cubic meters per second (m^3/s)

References

NBS Special Publication, SP330, 1974 Edition, "The International System of Units."

ISO International Standard 1000 (1973 Edition), "SI Units and Recommendations for Use of Their Multiples."

E380–72 ASTM Metric Practice Guide (American National Standard Z210.1).

Disclaimer:

Certain trade names and company products are identified in order to adequately specify the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the products are necessarily the best available for the purpose.

Contents

*Indicates speaker when a paper had more than one author.

The following registrants gave presentations at the symposium, but did not submit the manuscript for publication:

R. W. Staehle, "Mechanical Failure: An Overview," Session I

H. Corten, "Failure by Brittle Fracture," Session 1

Introduction

Elio Passaglia
Executive Secretary, MFPG

Institute for Materials Research
National Bureau of Standards
Washington, D.C. 20234

Mechanical failures are a pervasive fact of life in our society. Ranging from the failure of small items that all of us have experienced and that many of us take for granted, to the failure of a large complex structure that often becomes front page news, they have undesirable consequences for our society. The large ones many times cause loss of life or cause serious injury to many people. The minor ones sometimes also cause loss of life or injury, and they always cause frustration and anger on the part of the one to whom they occur. Always they cause loss of valuable material, and have undesirable social and economic consequences.

Thus it is not surprising that a large segment of our technological community devotes its efforts to the prevention of mechanical failures, and that numerous agencies of the Government are charged with the responsibility of preventing mechanical failure. These efforts to prevent mechanical failures are wide ranging indeed. They go from the material scientist studying the fundamentals of fracture, to the metallurgist assuring that a steel forging meets specification, to the design engineer designing a new bridge, to the Government administrator considering new specifications to reduce the incidence of mechanical failures. Yet despite these efforts mechanical failures continue to occur, and continue to present a problem whose solution, if possible, would have highly desirable consequences for our society.

The Symposium, the proceedings of which are presented in this volume, was designed to explore with a broad brush the various aspects of mechanical failures. We say "with a broad brush" because no symposium of reasonable length could conceivably grope with all the various aspects of the problem. All that can be done in a reasonable time is to define the problem, not discuss in depth its many detailed aspects and ramifications. As the organizers of the Symposium saw it, the definition of the problem of mechanical failures involved three aspects:

(1) The modes of failure
(2) The consequences of failure
(3) The implications (for action) of the occurrence of failures.

Thus the Symposium was divided into sections discussing each of these parts.

The section on the modes of failure was constructed on the basis that all failures of machine and structural components are eventually traced to material failure. Following this basis, the investigation of large numbers of mechanical failures shows that the primary modes of failure are fatigue, stress corrosion, and brittle fracture. Thus these topics are treated in individual papers in the Symposium, since they are of central importance to the problem of mechanical failure prevention. The reason this is so, and that these modes of material failure are the most common, is that while adequate design procedures are available for designing against simple overload, procedures to design against fatigue, stress corrosion, or brittle fracture, are either not available or are only now beginning to be used. Thus these modes of failure continue to be prominent because design methods are not yet available to deal adequately with them. These topics are dealt with in papers by McEvily, Kruger and Corten.[1]

When components are assembled into machines, particularly where moving parts are involved, the situation becomes far more complicated. Here we have the situation where one part interacts with another across their surfaces, and this leads to the whole problem of wear in its various forms, and its prevention. Various techniques are used to prevent mechanical failures in these circumstances. These include the use of fluid, solid or semi-solid films to separate the parts that are in relative motion, and the "components" aspects of mechanical failures thus refers to the design, kinds of damage asso-

[1] Corten's manuscript was not submitted for publication.

ciated with various types of components, and inspection. These topics are discussed in papers by Widner and Littmann, and by Wilde, and one of the underlying theories of the lubrication process, elastohydrodynamics, is described in some detail by Cheng.

Not all aspects of the modes of failure could be discussed at the Symposium. The papers given were all by renowned world authorities, and the topics they treated were obviously treated with great depth, but many topics had perforce to be left out. Among these might be mentioned crevice corrosion, fretting, erosion, corrosion fatigue, etc. In particular, one of the most important means of failure prevention, namely non-destructive evalution, was given only passing recognition in the paper by Erthal, et al., where the topic of laser holographics, among others, is mentioned.

The consequences of mechanical failures are many and great, and this symposium attempted to discuss only a few of them: the public safety consequences, the economic consequences, and what might be called the "societal" or emotional consequences.

Public safety is the most important reason that the Federal Government is concerned with the prevention of mechanical failures. Indeed, many agencies of the Government owe their existence to the desire of the Government to protect the safety of the general public by the prevention of mechanical failures. Only two were represented at this conference, the National Transportation Safety Board (Henry Wakeland) and the Atomic Energy Commission[2] (Peter Morris). In his paper Mr. Wakeland makes the point that some failures, because of their dramatic nature, assume an importance that transcends their effect on public safety, so that public emotional considerations are important in determining public policy with respect to failure prevention. In the atomic energy field, similar considerations are important, and protection against malfunction or mechanical failure of reactor systems and components in nuclear power plants is achieved by a very sophisticated combination of design and operational procedures. To assure that the probability of serious accidents is acceptably low requires conservative design, construction and operation, using redundant and diverse systems and components and a rigorous program for quality assurance.

The concern with public safety is, of course, not limited to the Federal Government. It is of concern also to the whole private sector, but only one aspect is treated here. This is in the construction field (Feld). Most of the accidents and partial collapse in structures come during the construction period, often with serious effects on the life and well being of people involved in the work as well as the innocent public that is by chance in the area involved. However, accidents do occur even after completion. Mr. Feld discusses the historical aspects of their prevention, the design and legal considerations in the construction industry arising from failures, and discusses the sources of failures and the responsibilities for their prevention.

Another of the important consequences of mechanical failures are economic. No aggregate statistical data exist to measure the losses due to mechanical failures, or even the cost of mechanical failure prevention. Indeed, the measurement of these costs would be a formidable task. However, to the extent that many accidental deaths and injuries are due to mechanical failures, the economic losses undoubtedly represent a substantial sum. However, the measurement of these losses poses a serious problem (Morgan). Economic analysis for quality control, which is one of the most important means of mechanical failure prevention, relies on the techniques of marginal analysis and life-cycle costing. Typically there is an optimal level of quality that is greatest relative to its cost. It is usually not economically feasible to obtain zero defects, which implies that failures at some level of occurrence will always be with us.

Although no aggregate statistical data exist as mentioned before, there exist, however, comprehensive data on a facet of failure in England. In a study underwritten by the Department of Education and Science, the economic and technological significance of tribology was shown (Yost). Tribology is the science and technology of interacting surfaces in relative motion and of the practices related thereto. The study detailed the annual saving which could be achieved by industry in the civilian sector in England through rigorous practice of tribology. This figure came to some $500 000 000. When translated to the United States through the GNP, the relevant figure would indeed be staggering.

---

[2] The Atomic Energy Commission has since been reorganized and its functions divided into two new groups: the U.S. Energy Research and Development Administration and the U.S. Nuclear Regulatory Commission. Peter Morris is with the U.S. Nuclear Regulatory Commission.

Finally, we come to the emotional consequences of mechanical failures. These range from the serious emotional consequences experienced by a large segment of the society when a dramatic failure in a large structure occurs, to the annoyances we all feel when an item of equipment we own fails. A particular case of wide importance of the latter type are automobile failures. Each of us has experienced the aggravation caused by such failures, and the anger with which we take our troubles to the dealer. We have also experienced the anger and frustration caused when the repairs did not fix the failure. These emotions are very real consequences of mechanical failures and are of very real concern to automotive manufacturers. How one of them dealt with the problem is described by Uhlig.

The occurrence of mechanical failures has implications for efforts of all concerned in their prevention. Those concerned range from basic scientists to Government policy makers, and the implications for action on the part of the various groups involved are discussed in papers in the last section of the Symposium.

In the area of basic science, the main concern, as detailed by Hirth, is in the area of fracture, and the understanding of the detailed mechanisms of fracture that is a prelude to its eventual control. In materials that are ideally elastic, the macroscopic theory of fracture is well worked out, but even in this relatively simple case the atomistic details of the fracture process still require considerable attention before real understanding is attained. In engineering materials, where plasticity at the crack tip is important in determining the fracture process, the situation is far more complicated. Dislocation motion at the crack tip must be taken into account in any atomistic treatment of the problem, and mechanisms of void initiation, growth, and coalescence become important in the fracture process. While some continuum treatments of the plasticity problem, such as the J integral, are important steps leading to a formalism that can be used by designers in fracture control design, the identification of the important material properties that control fracture mode await an atomistic treatment. In considering the viewpoint of the design engineer in fracture control, Paxton makes the important point that there is a great deal of impetus to reduce the number of engineering alloys since many specifications are redundant and obtaining data for fracture control design on all possible alloys poses an intolerable burden on the whole engineering community.

The existence of mechanical failures holds implications for action for industry as well. These are detailed by Compton and Ryan, representing two very large industries. Both customer safety and customer satisfaction considerations determine large programs aimed at failure prevention, and necessitate the development of highly detailed engineering data. In the case of fatigue, which is one of the major causes of mechanical failures, very sophisticated testing indeed is done on prototypes. Unfortunately, it was impossible in the time available to have discussions on what is done in failure prevention in small industries that do not have the resources available to the two industries represented here.

Consumers are also deeply concerned with the problem of failure prevention. This group has until recently not had a technical spokesman, but with the formation of the Consumer Product Safety Commission, such a spokesman has emerged, and the point of view of that organization was presented by Kushner, one of the Commissioners. He described how it was the will of the Congress to protect the consumer from hazard, and pointed out that mechanical failure prevention was a deep and abiding concern of the Commission. Of the various technical aspects of failure prevention, Kushner singled out non-destructive testing as having the most promise for application to the prevention of failures in consumer items.

Finally, mechanical failures have had important implications for public policy, a topic discussed by Roberts. The Federal Government has been a leader in mechanical failure prevention because of military activities, which are the sole responsibility of the Federal Government, and because of the space program. In both these areas technology is pushed to the limit. In the military case this is done because military ability must be premier, and in the space area because the accomplishment of the mission required this. In addition to these two areas, the protection of the safety of the public required the Government to be concerned with failure prevention. Roberts highlighted the emerging area of materials conservation as one in which the Federal Government could very possibly pass new laws, and in which failure prevention would play an important part. Of the various technical issues involved, Roberts called for more attention to fracture control design and the greater and more intensive use of non-destructive evaluation.

Clearly, this Symposium does not treat the whole problem of mechanical failures, even with the broad brush used. Each of the topics discussed could be the topic of one or several symposia. Indeed, symposia could be built around the topics of many of the individual papers presented. Nevertheless, as a definition of the problem, or the class of problems posed by mechanical failures, this symposium should serve the purpose.

Intent of the Symposium

D. C. Drucker

University of Illinois
Urbana, Illinois 61801

The intent of this Symposium is to sharpen our understanding of the problem of mechanical failure, its consequences, and its implications for research workers, practitioners, consumers, and the policy makers in government and industry. We hope that clear problem definitions will serve as a guide to engineers and scientists in industrial applications as well as in fundamental studies.

Complete prevention of mechanical failure is, of course, an uneconomic, unattainable ideal. Working toward it, however, will be essential as we move through the energy crisis into the materials crisis which lies ahead. The more effective our anti-failure engineering can become, the more viable will be our economy, the better our environment, and the safer and happier will be our people.

As time goes on, the public (quite understandably) expects us to do more and more, better and more economically. We must respond. Therefore, the arrangers of the program, Fred Ling, Elio Passaglia, and Robb Thomson, have provided superb speakers who will challenge our comfortable notions and prod us to action.

All of this is evident in the program before you. Although now I should speak only for myself, perhaps there are others here in the audience, who like me, have fallen somewhat behind current knowledge in some aspect of mechanical failure prevention. All of us in this uncomfortable but common situation express our gratitude to all those who organized this Symposium which will bring us right up-to-date in so crucial an area in an efficient and incisive manner.

MODES OF FAILURE

Session I

Chairman: Elio Passaglia
National Bureau of Standards

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

What We Can Learn from the Examination of Service Failures

John A. Bennett

Institute for Materials Research
National Bureau of Standards
Washington, D.C. 20234

The study of metal components that have fractured in service provides information that can be of major importance in preventing future failures. The metallurgical and fractographic techniques used in such studies are described so as to permit an evaluation of the reliability of the results. Examples are given of some case histories in which the fracture studies were highly effective, and of others in which the information was ignored and further failures occurred.

Key words: Crack propagation; failure analysis; failure prevention; fracture; fractography; mechanical failure.

1. Introduction

Mechanical failures come in a wide variety of shapes and sizes, and their importance varies from a minor nuisance to a major disaster. When a failure causes injury or serious property damage, there is usually an investigation. However, these are often conducted with no clear understanding of what information can be obtained from the investigation or how to use the information that is obtained. I would like to consider these questions in regard to one class of failures—namely fractures in metal components—and try to show where failure analysis fits into an overall failure avoidance program.

2. Techniques

First, what techniques are available for studying service fractures? Obviously, one needs to know how the information was obtained before he can evaluate its validity. The most important requirement in such a study is to be sure that the fracture being analyzed is the initial one—the one that really started the trouble. This may be difficult to do if, for example, the fracture caused an aircraft to crash, but it can be done in a surprisingly large number of cases.

The characteristic of the initial fracture which is most useful in identifying it is usually the lack of ductility. The cause of failure is apt to be a crack propagation mechanism, especially stress corrosion cracking or fatigue, which has occurred at a stress level well below the yield strength of the metal, so there will be little or no plastic deformation at the origin, in contrast to fractures that have occurred due to overload. This characteristic is demonstrated by the two specimens of identical material in figure 1. The upper specimen was broken in a laboratory fatigue machine by the application of many thousands of cycles of bending load, and the fracture appears completely brittle. The lower specimen was subjected to a larger bending load which caused extensive plastic deformation, showing the inherent ductility of the material.

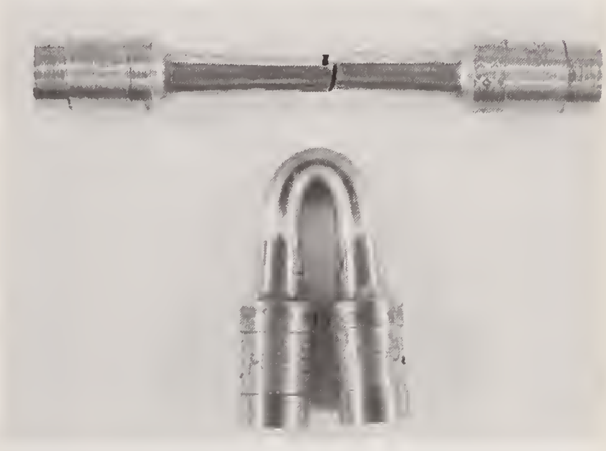


FIGURE 1. *Identical steel specimens loaded in bending.*
Top—broken by many thousands of stress cycles in a fatigue testing machine. Bottom—bent by a single application of load to show the ductility of the material.

In some components a failure due to stress corrosion cracking or fatigue will occur at a different place than one due to a static overload. This is because the cracks in the former cases will be initiated at the point where the local stress is a maximum, while the overload fracture will occur at the location where gross plastic deformation first takes place. A good demonstration of this was given by the failure of some cylinder hold-down bolts from an aircraft engine. The two bolts on the left in figure 2 were broken by fatigue cracks which started in the threads adjacent to the reduced section in the shank. As each bolt failed the loads on the remaining ones increased, until the stress became greater than the yield strength in the shank, and overload fracture occurred there, as in the bolt on the right.

Another good example was provided by the collapse of the Point Pleasant bridge over the Ohio River, which was due to stress corrosion cracking adjacent to the pin hole of an eyebar [1].[1] Eyebars are always de-

[1] Figures in brackets indicate literature references at the end of this presentation.

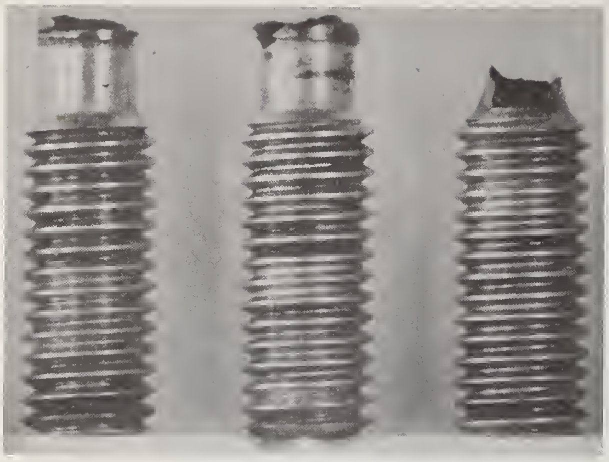


Figure 2. *Cylinder hold down bolts (7/16 in diameter) from an aircraft engine.*
Two bolts at left were broken by fatigue cracking, one at the right by overload.

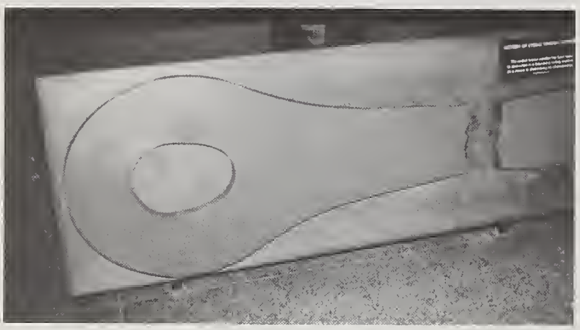


Figure 3. *Part of a tested eyebar showing normal overload fracture in the shank.*
(Smithsonian Museum of History and Technology.)

signed so that the nominal stress in the shank will be greater than that in the head, to make up for stress concentration around the pin hole. This means that when an eyebar is overloaded in tension, fracture always occurs in the shank, as shown in figure 3. Consequently, when the eyebar shown in figure 4 was found in the wreckage from the bridge, it was easy to guess that it had been the initial one to break; this was verified by a very careful investigation of the collapse.

If you are concerned with a large component, such as a welded ship which has broken in two, there is no difficulty in identifying the first fracture, because the component is all one piece. But you would like to know where the fracture started; this is usually important even on smaller components. Here the characteristics of the fracture surface often give the answer. When a tearing type fracture occurs in a reasonably ductile metal, there are irregular ridges formed on the fracture surface which are approximately perpendicular to the crack front. The ends of the propagating crack are held back at the surfaces (fig. 5), so the ridges tend to form a pattern which points back toward the origin. These patterns were first analyzed carefully in plate materials, so they are often referred to as "chevron" patterns. In components of a more complicated shape the patterns may not look at all like chevrons, but they can often be used to trace the development of the fracture back to the origin, if it is kept in mind that the ridges are approximately perpendicular to the crack front. Figure 6 shows the area near the origin of the fracture in a B–52 bulkhead which caused the loss of the plane and six crew members. Even though the wreckage was spread over quite a bit of landscape, there was little difficulty in tracing the fracture back to this area, and it is fairly obvious that the fracture markings radiate from the dark area in the upper right. This failure was the subject of an extensive investigation and a multimillion dollar lawsuit, but no one ever questioned the statement that the fracture started near the point shown by the arrow.

Another type of marking is generally found on fatigue fracture surfaces, particularly in wrought metals. These are the "clam shell" marks which represent the location of the crack front at various stages of its progress. Figure 7 shows the fracture of

Figure 4. *Portion of an eyebar from the Point Pleasant, W. Va. bridge.*
This fracture caused the collapse of the bridge.

Figure 5. *Sketch showing the development of fracture markings (solid lines) perpendicular to the moving crack front (dotted lines).*

an aircraft axle with a pronounced set of these markings; they result from variations in the magnitude of the alternating stress during crack propagation, so may not be present in some fractures.

It is not necessarily sufficient to find a fracture caused by stress corrosion cracking or fatigue to be sure that it is the initial cause of trouble. Failure in another component may induce an alternating load where there was none in normal service, for example; so the conclusions about the initial failure must always be checked to make sure that they are resonable in view of the service conditions.

Once the initial fracture has been identified with certainty, there are a number of techniques available to the investigator which may give information regarding the factors contributing to the failure. Many of these involve the study of the fracture itself. Something as simple as the location of the origin may be significant. The propeller fracture, figure 8, was found to have started on the inside, figure 9. As one would expect the bending load on the blade to produce a higher stress amplitude on the outside, the location of the origin showed that some extraneous factor had reduced the fatigue strength on the inside. The culprit was corrosion, apparently due to contamination on the lead wool used for balancing, which was packed in the hole in the blade.

The orientation of the fracture surface may be informative. The next two figures show failures of components that were stressed in torsion. Under this loading fatigue cracks may propagate on planes either

Figure 6. *Area near the origin of the fracture in a B–52 bulkhead. X4.*

Figure 8. *Fracture of a propeller caused by fatigue.*

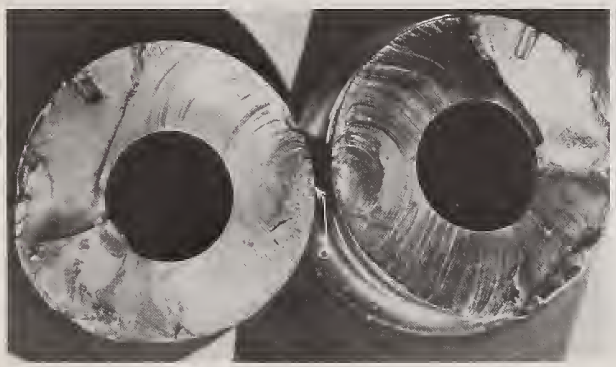


Figure 7. *Fatigue fracture of a 2 in diameter aircraft axle showing pronounced "clam shell" markings.*

Figure 9. *Surface of the fracture of figure 8 showing that the fatigue cracks started at the surface of the central hole.*

parallel to the length of the rod or at 45° to it. Figure 10 is a torsion bar spring from a highway bus; the initial crack was longitudinal and extended from A to B. A 45° fatigue crack extended from B to C. and the final overload fracture ran around from C to D. The fact that the initial crack was longitudinal indicated a weakness in this direction; metallographic examination showed that there were unusually large inclusions which had been elongated by the forging operations and which undoubtedly contributed to the failure. The helical spring in figure 11, on the other hand, cracked on a 45° plane. For a given discontinuity, the stress concentration is higher for the direct stress on the 45° plane than for the shear stress on the longitudinal planes, so this type of cracking suggests that a stress raiser had been a factor in initiating the crack. As a matter of fact. there was a large pit at the origin. probably the result of rolled-in scale.

Finer details of the fracture surface can also provide information regarding the stresses at the origin. In the absence of a stress gradient a fatigue or stress corrosion crack will grow fairly uniformly in all directions. so the crack front will be a circular arc. Figure 12 from reference [2] shows the fracture of a piston rod from a forging hammer where the fatigue crack started at an internal defect and formed a nearly perfect circle. Under a bending load, portions of the crack will be growing into a decreasing stress region, so the front will be elliptical, with the minor axis perpendicular to the neutral axis of the component. Figure 13 shows this effect on a thin metal component from an automobile suspension. The progressive positions

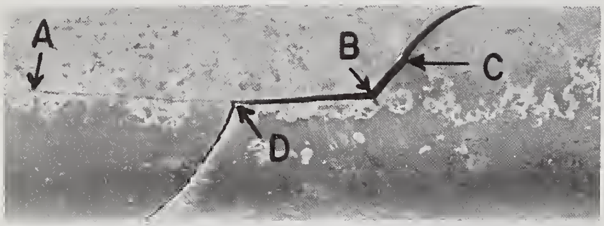


Figure 10. *Torsion spring (1⅛ in diameter) from a bus in which the initial fatigue crack developed on a longitudinal plane.*

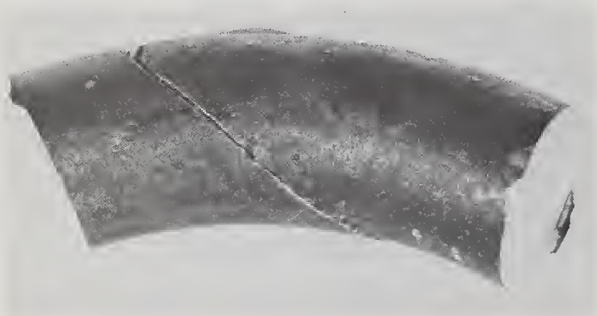


Figure 11. *Portion of a helical spring (½ in diameter rod) in which the fatigue crack had initiated on a 45° plane at a a surface pit.*

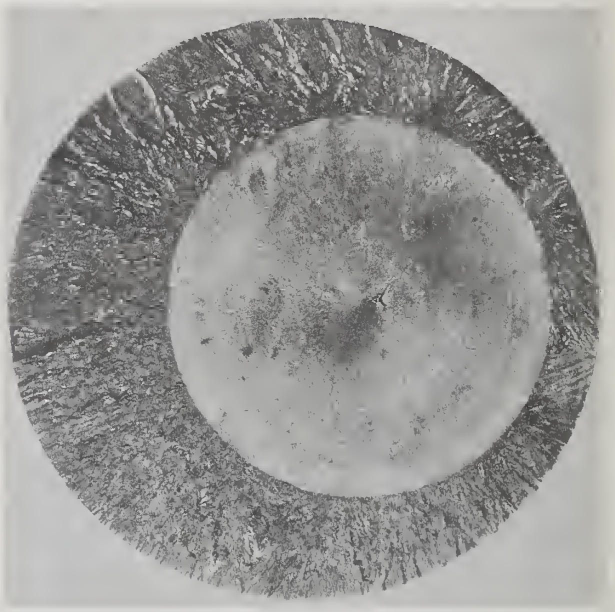


Figure 12. *Fracture surface of a piston rod from a forging hammer.*
The fatigue crack started at an internal flaw in the rod. (From reference [2].)

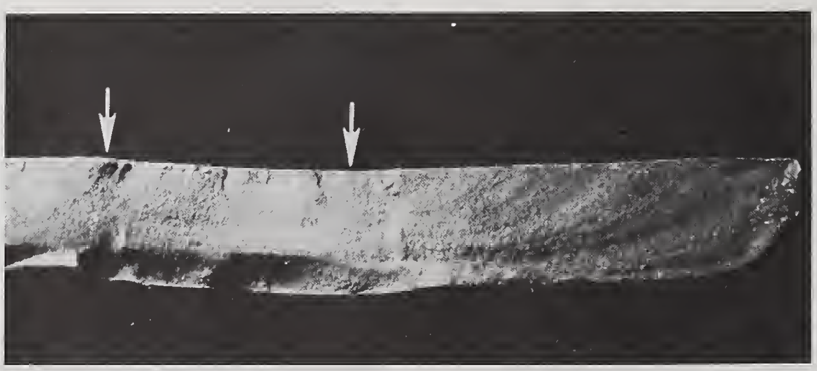


Figure 13. *Fatigue fracture from a ⅛ in thick steel component of an automobile suspension.*
Cracks started at several origins between the arrows.

of the crack front are often well defined by the clam shell markings on fatigue fractures; they are much less apt to be discernible on stress corrosion fractures.

The light microscope is of little use for examination of fracture surfaces above about 30 magnification because of its short working distance and limited depth of focus. For a closer look at the fracture surface, the investigator will go to the electron microscope. If the part cannot be cut up, or if the maximum detail is needed, the transmission microscope will be used by making replicas of the surface. This is a time-consuming process, but the fine detail that can be obtained is shown by figure 14 which is the surface of a laboratory fatigue fracture in an aluminum alloy. The striations are 0.001 mm apart and represent the progress of the fatigue crack during each cycle of stress. Such detail is seldom seen on service fractures, in part because it is destroyed very quickly by even mild corrosion.

FIGURE 14. *Surface of a fatigue fracture in an aluminum alloy laboratory specimen.*
The photograph is a transmission electron micrograph of a replica of the surface.

If the fracture surface can be inserted into the vacuum chamber of a scanning electron microscope the examination can be made much more quickly than replicas can be prepared and the detail will usually be sufficient to show the nature of the fracture. Figure 15 is a scanning electron micrograph of a fatigue fracture in an aircraft flap hinge. The striations can be seen clearly even though the resolution is not nearly as good as the transmission micrograph. The SEM is also very useful in giving an idea of the fracture toughness of a material by showing the amount of cleavage or intercrystalline fracture in the overload part of the surface. In interpreting the results of electron micrography of fracture surfaces, it must be remembered that while the presence of striations is a definite indication of a fatigue fracture, their absence does not mean that it was not fatigue; they may have been obliterated by corrosion before the part reached the laboratory.

After examining the fracture surface, the investigator next wants to learn all he can about the material; for this he will use metallography, mechanical testing, and chemical analysis. Of the three, metallography will probably provide the most extensive information in a majority of cases. The investigator is fortunate if there are other cracks in the component which appear to be similar to the one which was involved in the fracture, because metallographic examination of these will provide information on the nature of the cracks much more readily than examination close to the fracture, where the material is often battered or otherwise damaged. This is particularly true of stress corrosion fractures where the surface of the fracture is often

FIGURE 15. *Scanning electron micrograph of a fatigue fracture in an aircraft component.*
The width of the picture represents 75 μm on the surface.

FIGURE 16. *Fracture of the collar on a* 3/16 *in diameter aircraft fastener.*

uninformative, but metallographic examination of material containing cracks will show the wandering and branching that is typical of this kind of cracking. Figure 16 shows the fracture of a swaged-on aluminum alloy collar on a 3/16 in diameter aircraft fastener. In addition to the fracture, there were numerous other cracks in the collar, and figure 17 is a metallographic section through some of these. The nature of the cracks is typical of stress corrosion, caused in this case apparently by the residual stress resulting from the swaging operation. Similarly, figure 18 shows the cracks in a section of the fractured eyebar which I showed earlier; these are also believed due to stress corrosion cracking.

Metallographic examination is the best way to discover many material defects that may have influenced a service fracture. Figure 19 shows the two parts of a failed aircraft wing spar fitted back together. The spar was assembled by welding steel tubing, and the fatigue crack had started at the toe of the weld at the left. It is obvious that the weld caused a stress concentration, but in addition metallographic examination showed another significant factor—decarburization (fig. 20). The origin of the fatigue crack was in the low carbon content material at the surface of the tube where the fatigue strength was estimated to be only about half as great as it would have been in sound material.

FIGURE 17. *Section of the collar of figure 16 showing cracks remote from the fracture. X 75.*

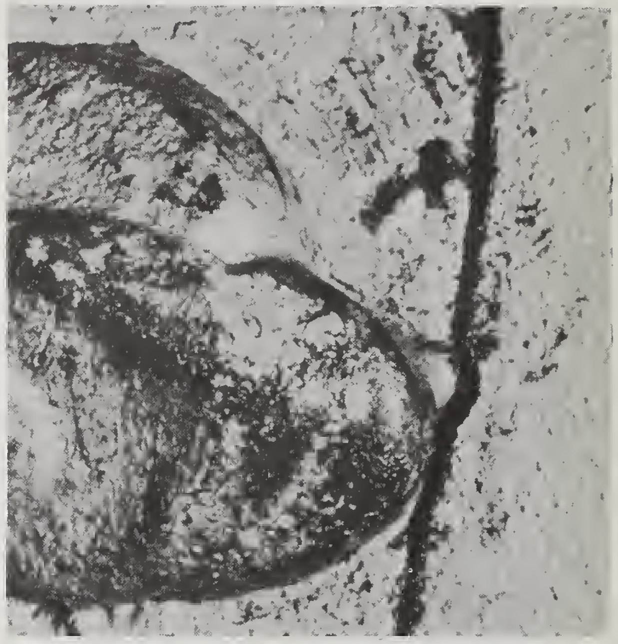


FIGURE 19. *Fracture in a light aircraft wing spar.*
The fatigue crack started at the toe of the weld. × 5.

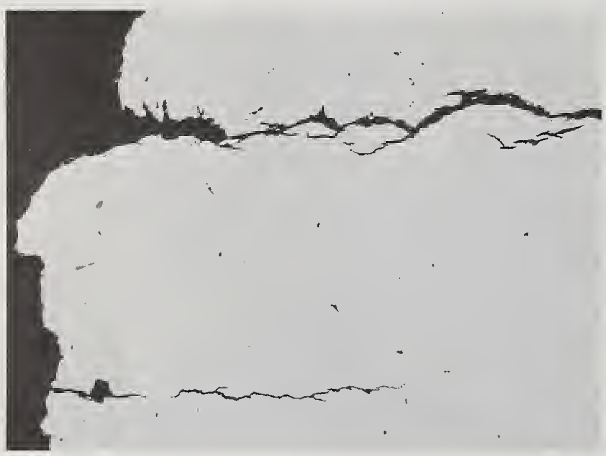


FIGURE 18. *Section of the eyebar of figure 4 showing cracks a short distance from the fracture. X 60.*

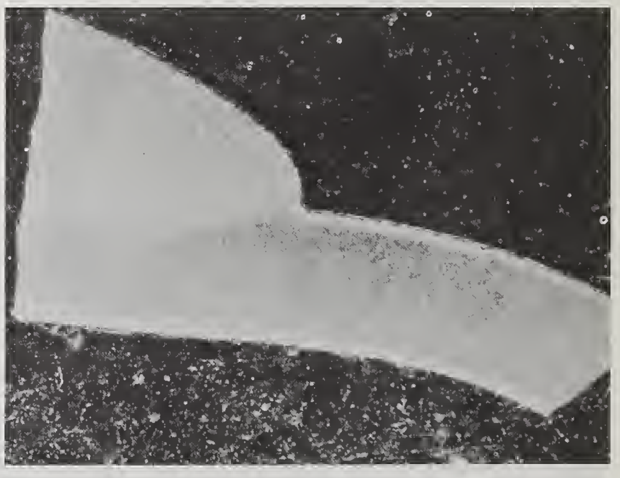


FIGURE 20. *Cross section of the tubing and weld of figure 19 showing decarburization (light-etching material near the surface).*
The wall thickness of the tubing is 2.5 mm.

Some of the other defects which have been found by metallography are: excessively large inclusions, laps and seams due to poor forging techniques, and improper heat treatment.

Probably the most common technique used in trying to learn about the material which has fractured is the hardness test. It has the advantages of low cost, rapidity, and very little damage to the failed part, yet it can sometimes give enough information regarding the mechanical properties of the metal so that no further testing is required. Figure 21 shows a cross section of a stainless steel reaction plate from a dam gate; these plates were found to be cracking after they were fastened to the structure but before they had been loaded. Near the surface where the cracking started the hardness was 45 Rockwell C, an unusually high value for the 410 type steel from which the plates were made. The deformation of the plate after cracking showed that there was a very high residual stress present, and apparently this combination was sufficient to initiate cracking under atmospheric conditions. The problem was thought to be due to lack of tempering or improper tempering; these preliminary conclusions based on only the visual examination and hardness tests were verified by more extensive tests.

There are, of course, many more sophisticated techniques for determining the mechanical properties of metals, but these are of relatively little use in the investigation of fractures; they are however of great importance in evaluating materials in order to prevent future failures, and you will be hearing more about these techniques from other speakers.

It is almost routine to ask for chemical analysis of the failed component, but I cannot think of a single instance in which the results were of real significance in the investigation of a service fracture. This does not mean that the analysis should be skipped, because the component just might have been made of the wrong alloy, but it doesn't happen often. The chemist may be also called upon for the analysis of corrosion products in an effort to determine the corrodent responsible for stress corrosion cracking or corrosion fatigue. These attempts have not been successful in many cases because of the small quantities of material involved and the complexity of corrosion processes. These techniques will become more and more useful as our knowledge of corrosion is increased.

FIGURE 21. *Cracks in a 2.5 cm thick reaction plate which occurred before the plates were put in service.*

3. Use of Information

I hope that this discussion has given an idea of the means which are available for investigating fractures, and has shown that the techniques are capable of revealing a lot of important information concerning the conditions leading to the fracture. The question now is, how can this information be used to best advantage in avoiding further fractures?

If there are other components, similar to the fractured one, which are still in service, I believe that it is important to verify the results of the investigation by duplicating the failure in the laboratory. If the fracture was in a large assembly, such as an airplane wing, the duplication will have to be done by stressing test samples which represent only a small part of the total assembly. This introduces some uncertainty, but there is no other way to get a base value from which to measure the effect of corrective measures. And unless the failure can be produced in the laboratory under conditions that might reasonably be expected in service, there is no assurance that the conclusions of the investigation are correct. The duplication need not use a full-sized component; for example figure 22 shows the fatigue fracture of a truck axle in which the fatigue crack propagated more than half way across before final fracture occurred. This indicated that the operating stress on the axle was quite small, so the material must have had a very low resistance to fatique crack initiation. This was attributed to severe decarburization combined with the grooves made by shearing off the forging flash. These factors could have been evaluated on laboratory specimens much smaller than the actual axle, and the observed improvement due to eliminating them could be applied to the full-sized parts with good confidence.

FIGURE 22. *Fatigue fracture in a truck axle.*
The vertical dimension of the axle is 7.5 cm.

Service failures due to stress corrosion cracking are more difficult to reproduce in the laboratory than those due to fatigue, but it is just as important to make the effort. Because of the long exposure times required, samples incorporating the possible corrective measures should be tested at the same time as those representing the component that failed. For example, if the corrective measure is an improved coating, samples with the new coating should be tested simultaneously with those being tested to duplicate the service fracture. Also the testing time can be shortened by examining specimens metallographically after various exposure times to see if cracks have developed, rather than waiting for complete fracture.

A wide variety of corrective measures have been used to reduce the likelihood of repetition of fractures, and generally these will be indicated by the results of the service failure investigation. For the examples that I have shown the indicated measures included, proper torquing of bolts, elimination of corrosive contaminants, stricter surface quality requirements, cleaner steel, grinding off of weld reinforcements, and eliminating decarburization. Unfortunately, in our work on service failures at NBS we were seldom informed as to what corrective action was taken; we could only assume that if we did not hear about further failures, some successful modification was made. Figure 23 shows a case that we know a little more about; this is a blower shaft which was broken by fatigue cracks initiating in the severely fretted material under a collar. The collar was held on by two set screws which permitted considerable movement of the collar. When the replacement shaft was installed, the method of attaching the collar was changed to eliminate the possibility of movement relative to the shaft, and no further trouble has been reported.

FIGURE 23. *Severe fretting on a 10 cm diameter fan drive shaft where a collar contacted the shaft.*
The fatigue cracks started in the fretted area and resulted in fracture of the shaft.

In closing, I would like to say that despite a large amount of failure analysis work done in the United States, failures continue to occur. This is illustrated by an excerpt from the Washington Post of June 18, 1973.

> . . . . . A letter to the [House Government Operations] committee from Rep. Edward Mezvinsky (d.-Iowa) referred to the crash of a Beech Model 18 last April 19. The crash was the result of a structural failure of a wing spar, according to the Safety Board preliminary investigation.
>
> "Such a structural defect was first identified by a National Bureau of Standards metallurgical examination during the investigation of a 1947 fatal crash. Since that time, seven additional fatal crashes have been attributed to wing failure," Mezvinsky wrote. . . . .

More encouraging is the experience subsequent to the failure of the Point Pleasant bridge. Another bridge, identical in practically every detail, existed at St. Mary's, West Virginia. After the investigation of the collapse of the Point Pleasant bridge (see figs. 4 and 18), it became clear that there was no way to make sure that a dangerous crack did not exist in the St. Mary's bridge. Accordingly the authorities, in the face of violent local opposition, removed the bridge. This was a courageous decision which may have saved many lives.

A great deal can be learned about failure prevention from the analysis of service failures, and from the related research in fatigue, fracture, stress-corrosion, wear, and non-destructive evaluation. One of the most important tasks we face in the future is the more effective utilization of information obtained from these activities.

4. Reference

[1] Bennett, J. A. and Mindlin, Harold, Metallurgical aspects of the failure of the Point Pleasant Bridge, Journal of Testing and Evaluation, JTEVA, Vol. **1**, (1973), pp. 152–161.

[2] Bending and Tensile Failures in Shafts, Chapter 7 of How Components Fail, by Donald J. Wulpi, American Society for Metals, Metals Park, Ohio, 1966.

5. Discussion

W. E. Littmann, The Timken Company: You spoke of metallographic examination often revealing excessively large inclusions as a cause of failure. How do you tell the difference between a normal inclusion and one that is excessively large and a cause of failure?

J. A. Bennett: I don't know any simple answer to that, but comparison may be the best way. In the spring from the highway bus, the inclusions were terrible. An inclusion at 100 magnifications couldn't be covered on a 4 x 5 print. I do have one theory, and that is, in fatigue at least, the thing that is important is the size of the largest inclusion in the highly stressed region. Most inclusion rating systems result in some sort of an average inclusion size. I would like to see a little more emphasis on the extreme values theory in order to try to determine what is apt to be the largest inclusion in the highly stressed region.

J. Kruger, National Bureau of Standards: You said you had never seen a case where chemical analysis helped you with a failure analysis. Surely this is not true. Analysis of corrosion products sometimes identifies a damaging species.

J. A. Bennett: You misquoted me a little. I said it hadn't been of major importance, and I was thinking primarily of the analysis of the composition of the failed component. I think the analysis of corrosion products is becoming of increasing importance. The difficulties lie in the complexity of the corrosion process and the small amount of material that is available. But I do think that it is definitely one of the things that is coming up in importance as techniques for small sample analysis get better and as our knowledge of all the complicated reactions that go on in corrosion improves.

A. J. Babecki, NASA, Goddard Space Flight Center: One aspect you didn't mention which we find in the space business to be rather significant is the fabrication of the part from the bulk material without due consideration of the anisotropy. Mechanical properties do vary with orientation. You didn't mention in any of your analyses that this was the major problem.

J. A. Bennett: That certainly is an important consideration, although not in any of the examples that I've shown. Years ago, Logan showed that in extruded hinge fittings, the susceptibility to stress corrosion cracking is much greater in the short transverse direction.

O. Jahari, IIT Research Institute: In regard to the effect of inclusions on fracture, we at the Institute have found an approach that I think has worked extremely well. Instead of looking at the microstructure, we generate fractures and examine them with the scanning electron microscope. We have reported cases where two materials, with identical microstructures have widely different fatigue properties because of the location of the largest inclusion. In fact, many of the property differences which we have been attributing to statistical analysis or statistical variations could be explained by this fact. When there are existing cracks, a good approach is to open the cracks and examine the fracture surfaces. Recently I worked on a copper casting in which there were existing cracks. When the cracks were opened, inclusions were found that contained copper and phosphorus. Combining chemical analysis with scanning electron microscopy offers a very useful approach to failure analysis, particularly where mechanical properties are concerned.

J. A. Bennett: I think that is a very good point. We have had cases where we could see secondary cracks, that is cracks not close to the fracture, which were clearly initiated at inclusions.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Failure by Fatigue

A. J. McEvily

University of Connecticut
Storrs, Connecticut 06105

The lifetime of manufactured items under intended service conditions is often limited by the processes of corrosion, wear and fatigue. Since such processes represent economic loss and perhaps a safety hazard to the consumer, much effort has gone into the understanding of these phenomena as well as into improved design procedures to guard against their occurrence. In the field of fatigue in particular a considerable advance in recent years has been made in the quantitative treatment of the fatigue process, especially with respect to the matter of fatigue crack growth. Improved understanding of the fatigue crack growth process is timely as in certain circumstances, as for example in the case of welded structures, it is not the initiation of cracks but rather the growth of cracks from preexisting defects which is the critical aspect in determining service lifetime. Other advances have been made in improving the resistance of materials to fatigue either through the control of chemistry or by control of processing variables. Such procedures are generally more important in affecting the crack initiation rather than the crack propagation stages. In this presentation a review of the current status of fatigue will be given from the mechanistic as well as the design viewpoints. Areas in need of further understanding such as corrosion fatigue, creep-fatigue, and fatigue under variable amplitude loading will also be considered.

Key words: Crack initiation; crack propagation; fatigue crack growth; fatigue failure; metal fatigue.

1. Introduction

In this paper we are concerned with the problem of fatigue and the steps being taken to minimize its occurrence. Fatigue failure, of course, is due to the repeated application of stress. This type of failure can occur in crystalline as well as noncrystalline materials, with metals and polymers being of principal concern. Fatigue failure involves both the initiation and propagation of cracks, and irreversible plastic deformation plays a key role in both, and it may be surprising, therefore, that a completely elastic, i.e., brittle material, is not subject to fatigue. However, such materials are not generally useful as engineering materials, a notable exception being a composite material which contains brittle fibers encased in a protective matrix. Therefore, although ductility is a highly desired characteristic of a metal, under cyclic loading it leads to failure. Perhaps Aesop could find a moral here.

In considering the various components subject to fatigue, sometimes a division of these components into two categories is made. The first of these are components which comprise the primary load bearing members of structures such as airplanes, cars, trucks, bridges, pressure vessels, etc. The second category includes the remainder of products subject to fatigue, a category which includes the bulk of mechanical products ranging from can openers to aircraft turbine blades. This later category can also be broadened to include certain types of wear failures which in fact result from contact fatigue. Of these two categories, more attention has been given in general to structural failures due to fatigue. In addition to fatigue, other forms of structural failure are: overload failure of a ductile nature, overload failures of a brittle nature, creep, stress corrosion and instability. Of these failure types fatigue is the most common form of failure simply because few structures are subjected to the static loading assumed in design [1].[1] In fact Freudenthal [2] states that 95 percent of all structural failures are fatigue failures. Structural fatigue failures often attract considerable attention because of their sometimes catastrophic nature, which result in loss of life as well as economic losses, as in the case of the Comet failures. Nonstructural failures are usually less dramatic, but nonetheless the integrated economic loss is undoubtedly significant as is the drain on natural resources as a result of such failures, although statistics on the frequency of occurrence as well as the economic factors associated with such losses are not generally available. In any event, such figures must be carefully interpreted, for a low frequency of occurrence may not necessarily indicate an absence of a fatigure problem but may instead reflect a competent handling of the problem. Such is the situation in the transport aviation field in recent years, where improvements in fatigue analysis, fatigue testing, and in the capabilities of nondestructive inspection have served to control the fatigue problem. In other areas of the consumer product field, design against fatigue may not be as carefully carried out and fatigue can be a more common occurrence. Some 700 000 product liability law suits are currently pending against manufacturers [3], an increase by an order of magnitude in a span of a few

[1] Figures in brackets indicate literature references at the end of this presentation.

years [4]. Such an increase reflects a growing expectation for product reliability and safety on the part of the consumer. Undoubtedly many of these failures, rightly or wrongly, are attributed to fatigue resulting from poor design.

Fatigue has now been the subject of research investigations for over one hundred years, and despite the progress made, failures continue to occur. This situation is due in part to the complex nature of the fatigue process and the stress-material-environmental interactions involved therein. However, much of the fatigue problem is simply due to poor communications. Many designers do not know how to deal properly with fatigue and do not treat it as a matter of concern until product failures start to occur. This situation can lead to a rapid but expensive development of an appreciation of the complex nature of fatigue and of the procedures available to aid in avoiding such failures. This is not to say that we can now design Holmes' one-horse shay, or match the capabilities of Neville Shute's hero who accurately predicted that the plane on which he was a passenger was about to experience fatigue failure, but we are getting closer. Today there are two approaches to design against fatigue [5]. One of these is to design for a safe life in terms of the S–N properties of the material. The other approach is to design damage tolerant (fail-safe) structures so that in the event that fatigue cracks do develop their presence will not be catastrophic. Nondestructive inspection is an important aspect of this latter approach. The purpose of the remainder of this paper will be to review the basic aspects of these two design approaches, with emphasis placed upon the unit processes of crack initiation and growth in simple specimens. Knowledge of these processes is important in the design stage, with simulated service testing being relied upon to provide a final check of the fatigue analysis of many complex structures.

2. Lifetime Determination

A representative S–N curve is shown in figure 1 [6]. To emphasize that scatter of test results is an important characteristic of fatigue, survival probability estimates are also included. The safe-life design approach uses such curves, coupled with component testing, to establish fatigue reliability, particularly in the high cycle range.

Fatigue is generally a surface sensitive phenomenon with cracks in polished specimens initiating at sites such as slip bands in or at inclusions, figure 2, depending upon the nature of the metal. In total life determination of simple specimens, the distinction between crack initiation and propagation in the failure process is not made, at least at the design level. Cracks of the order of a millimeter in size are usually visible only in the last stages of life, although they may be present but of smaller size for a much greater portion of the total lifetime, especially for lifetimes less than 1000 cycles. This total lifetime will be influenced by the environment. For example, at elevated temperatures oxidation can exert a deleterious effect, as shown in figure 3 [7]. Note that in this case the fatigue lifetime is plotted as a function of plastic strainrange, rather than stress amplitude. At elevated temperatures the time dependent nature of deformation involving creep or stress relaxation can also become important. Figure 4 shows a variety of hysteresis loop shapes as well as a thermal fatigue cycle [8]. Analysis of cyclic loading under such conditions is obviously complex, but new methods of analysis to deal with such loading histories are being developed, as for example, the strainrange partitioning methods due to Manson and co-workers [9]. The importance of surface finish on fatigue is indicated in figure 5 [10]. Allowance for scatter, surface finish, corrosion, and in-service surface damage can be combined in certain instances to reduce the

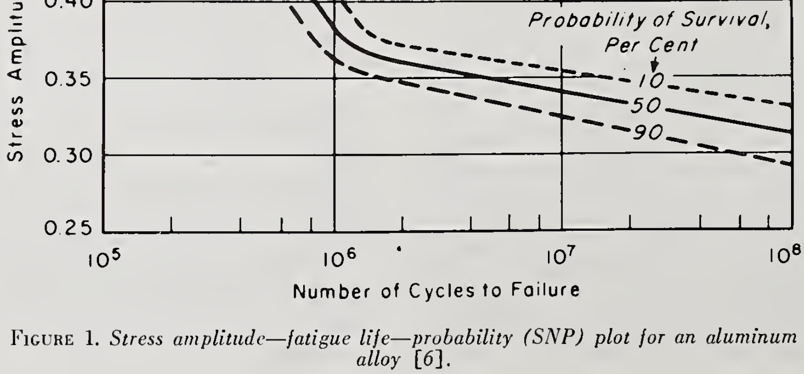


Figure 1. *Stress amplitude—fatigue life—probability (SNP) plot for an aluminum alloy [6].*

allowable design stress by as much as a factor of ten below the average smooth specimen stress for a given lifetime.

In the low cycle range where inelastic deformation is dominant it is more usual to relate the lifetime to a strain range rather than a stress range because relatively simple relationships, as indicated in figure 3, can exist between plastic strainrange and lifetime stress amplitude for fully reversed strain. These relationships are as follows, using the notation of Morrow [11]:

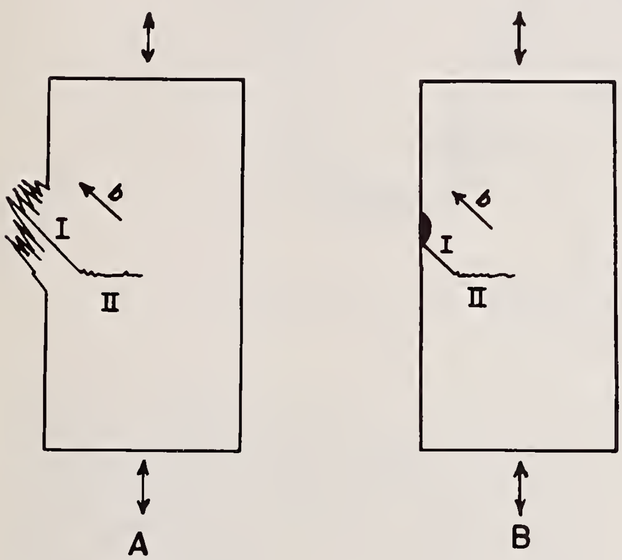


FIGURE 2. *Fatigue crack initiation sites in (A) a slip band, and (B) at an inclusion.*
The STAGE I of crack growth is along primary slip planes; the STAGE II of crack growth is perpendicular to the principal tensile stress axis.

$$\frac{\Delta\varepsilon_p}{2} = \varepsilon'_f\ (2N_f)^c \qquad (1)$$

$$\frac{\Delta\varepsilon_e}{2} = \frac{\sigma'_f}{E}\ (2N_f)^b \qquad (2)$$

$$\sigma_a = K'\left(\frac{\Delta\varepsilon p}{2}\right)^{n'} \qquad (3)$$

where $\Delta\varepsilon_p$ is the plastic strainrange
ε'_f is the fatigue ductility coefficient
$2N_f$ is the number of reversals to failure
c is the fatigue ductility exponent
$\Delta\varepsilon_e$ is the elastic strainrange
σ'_f is the fatigue strength coefficient
b is the fatigue strength exponent
σ_a is the true stress amplitude
K' is the cyclic strength coefficient
n' is the cyclic strain hardening exponent
E is the elastic modulus

Equations (1) and (2) can be combined to yield

$$\frac{\Delta\varepsilon_t}{2} = \frac{\Delta\varepsilon_e}{2} + \frac{\Delta\varepsilon_p}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f\ (2N_f)^c \qquad (4)$$

Maraging steel is an example of an alloy which follows these average lifetime relationships as shown in figure 6 [12]. High strength aluminum alloys and titanium alloys generally do not exhibit such good agreement. In such cases a more complete testing program must be employed to obtain reliable fatigue life relationships.

Design situations often involve a mean stress whereas much fatigue test data, particularly those obtained in the low-cycle range, are obtained under fully reversed cycling. A Goodman diagram, figure 7 [13] for example, can be used to determine the allowable stress amplitude for a given lifetime, σ_a, at any mean stress, σ_m. If it is assumed that the stress amplitude for

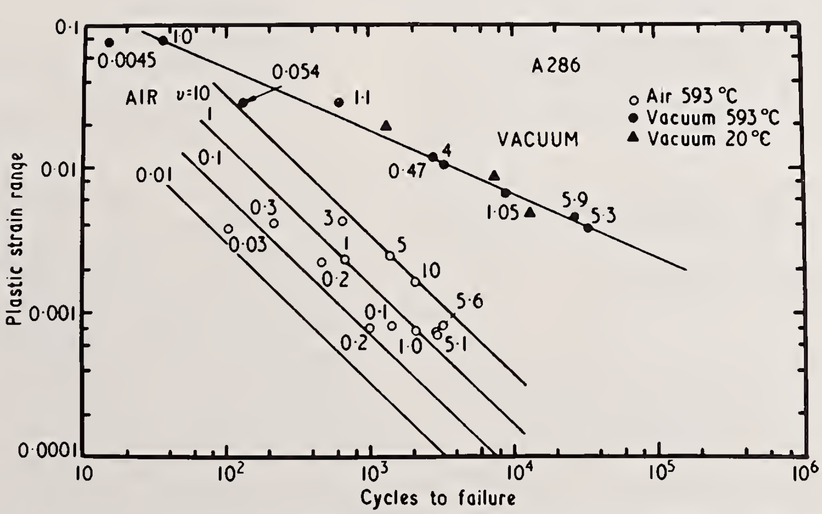


FIGURE 3. *Plastic strain range versus fatigue life for A286 in air and vacuum at 593 °C.*
Numbers adjacent to test points indicate frequency in cycle/min. After Coffin [7].

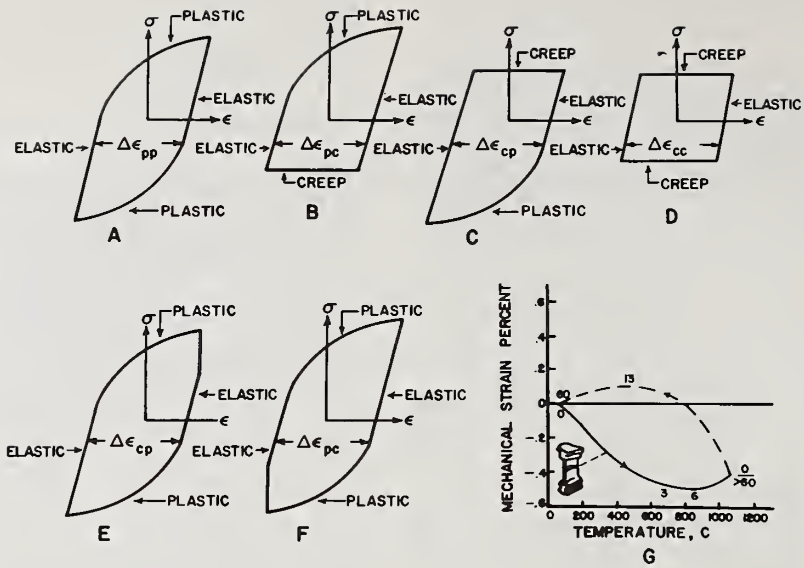


FIGURE 4. *A–F indicate a variety of hysteresis loop shapes which can be developed at elevated temperatures as a result of plastic deformation, creep and stress relaxation.*
G indicates a thermal fatigue cycle in which both temperature and strain vary.

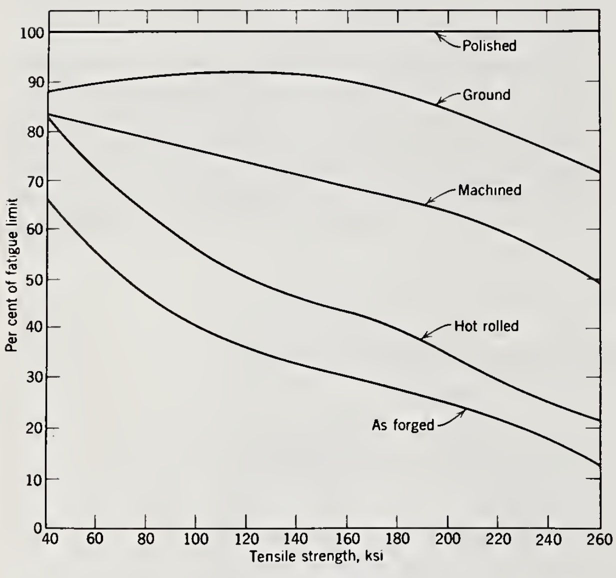


FIGURE 5. *Effect of surface on fatigue as a function of tensile strength of steel [10].*

a given lifetime varies linearly with mean stress then we can write

$$\sigma_a = \sigma_{\bar{a}} \left(1 - \frac{\sigma_m}{\sigma_u}\right) \tag{5}$$

where $\sigma_{\bar{a}}$ is the stress corresponding to a given lifetime for fully reversed loading, and σ_u is the ultimate tensile strength. For some materials it may be preferable to use the true stress at failure, σ_f, in place of σ_u.

Another design complication is that the in-service loading is rarely of the constant amplitude variety. Random loading has to be considered, and a first problem is the selection of a statistical counting method assuming that adequate load data is available. Current statistical counting methods are modifications of the three basic types indicated in figure 8 [14]. These three types are characterized by:

a. The variable stress reaches a maximum or a minimum.
b. The variable stress changes from a minimum to a maximum, i.e., it describes a positive or negative range.
c. The variable stress crosses a given level in a positive or negative direction.

Recent methods such as the rain-flow and range-pair cycle counting methods are variations of type *b* in which a second counting condition, the instantaneous mean value, is introduced. Figure 9 is a strain time history in which a pairing of ranges to form a cycle is made. For example, in figure 9, a range is counted between peak 1 and peak 8, and each range that is counted is paired with the next straining of equal magnitude in the opposite direction to make up a complete cycle. In figure 9 part of the range between peaks 8 and 9 is paired with the range counted between peaks 1 and 8. In figure 9 the counted ranges are marked with solid lines and the paired ranges with dashed lines. Each peak is taken in order as the initial peak of a range, except that a peak is skipped if the part of

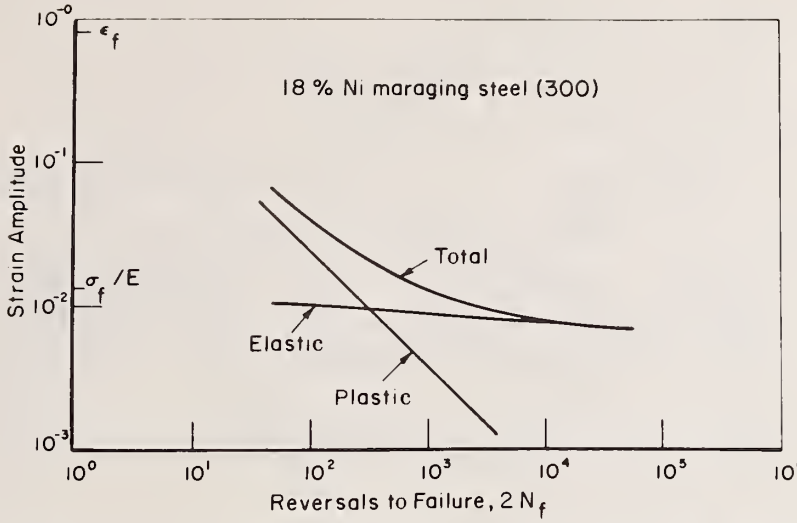


FIGURE 6. *Low cycle fatigue of a maraging steel as a function of elastic, plastic, and total strain amplitudes* [*12*].

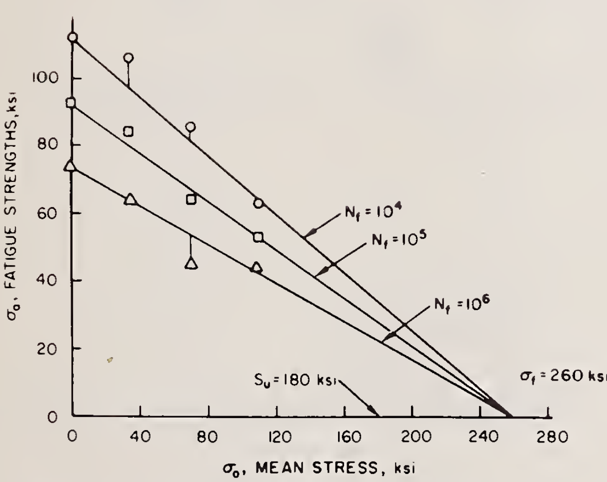


FIGURE 7. *A Goodman diagram for a steel of 180 ksi tensile strength* [*13*].

the history immediately following it has already been paired with a previously counted range [15].

Each strainrange can be converted to an equivalent strainrange for fully reversed loading by using eq (5). The fraction of life expended, $\frac{1}{N_i}$, for each cycle can then be computed. Miner's law, $\sum_{i=1}^{i=i} \frac{n_i}{N_i} = 1$, is then used to predict the total number of cycles which can be applied before failure. Dowling [15] used this approach to predict the fatigue life of 2024—T4 aluminum alloy under a complicated stress-strain history and found predicted lives to be within a factor of three of the actual lives in 83 tests.

Thus far, we have considered lifetime prediction for simple specimens and modifications due to effects of surface finish, mean stress and random loading. Another important consideration is the effect of a notch on fatigue properties. In cases where eqs (1), (2), and (3) hold they can be used in conjunction with Neuber's rule [16] (to be defined) to predict the fatigue lifetime as a function of the applied R ratio (R being the ratio of minimum to maximum stress in a simple loading cycle). In considering the plastic behavior at the root of a notch it can be shown that the local R ratio differs from the macroscopic value of R [17, 18]. For example, under R = 0 loading, the value of R at the notch root may be equal to −1. In order to estimate the local R value use is made of the Neuber rule: [16]

$$K^2_T = K_\sigma K_\varepsilon \qquad (6)$$

which leads to

$$K^2_T \frac{S^2}{E} = \sigma_l \varepsilon_l \qquad (7)$$

where K_T is the theoretical stress concentration factor
K_σ = the local stress concentration factor
K_ε = the local strain concentration factor
S = the applied elastic stress
σ_l = the local peak stress at the notch
ε_l = the local peak strain at the notch.

To take into account notch sensitivity and notch size effect (an effect related to the volume of material sub-

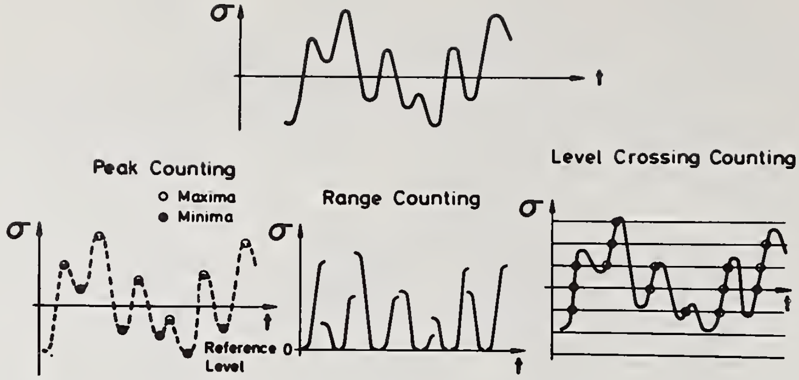


FIGURE 8. *Example of three basic types of counting methods in the analysis of variable amplitude loading [14].*

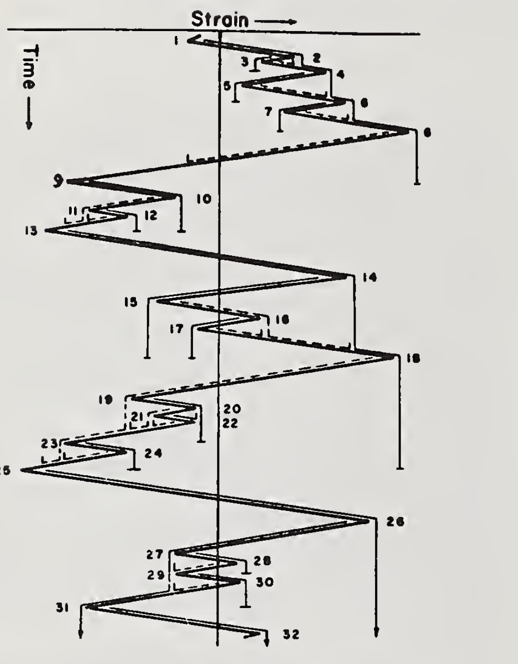


FIGURE 9. *Example of range pair counting method [15].*
Solid and dashed ranges are paired to form cycle of corresponding mean strain.

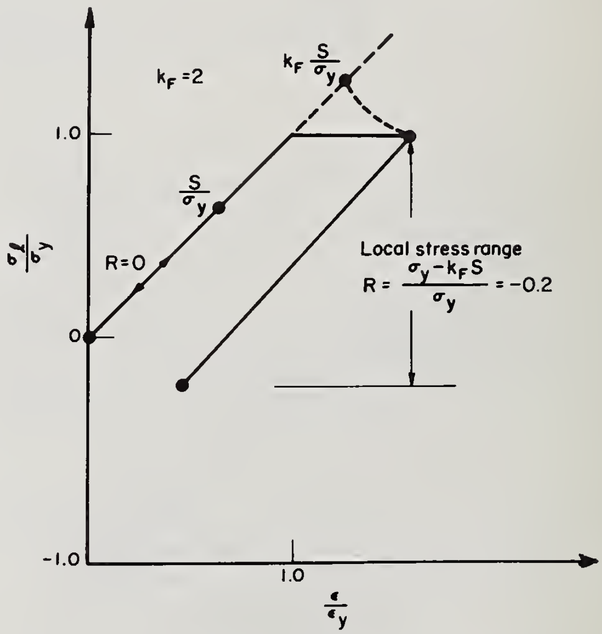


FIGURE 10. *Stress range at root of notch as influenced by localized yielding.*
Stresses and strains normalized with respect to yield stresses and strains, respectively. Applied R ratio equal to zero; fatigue notch factor, $K_F=2$.

ject to the peak stress at the tip of a notch), eq (7) can be written as

$$K^2_F \frac{S^2}{E} = \sigma_l \varepsilon_l \qquad (8)$$

where K_F is the fatigue stress concentration factor determined at lifetimes such as 10^7 and defined as the ratio of the unnotched to notched fatigue strengths for fully reversed loading.

Figure 10 [19] is a simplified version of the stress-strain behavior at the root of a notch when yielding occurs on the tensile part of a cycle and is based upon the experimental work of Wetzel [17] and Crews and Hardrath [18].

Again, eq (5) can be used to convert a stress amplitude at a given stress to an equivalent stress or strain amplitude fully reversed loading or vice versa. To obtain the $R = -1$ baseline information, two approaches can be used. One is to use eqs (1), (2), and

(3) together with the Neuber rule, eq (6), to obtain

$$S_{\bar{a}} = \frac{\sqrt{K'E}}{K_F}\left\{ [\varepsilon'_f \ (2N_f)^c]^{n'+1} + \frac{\sigma'_f}{E} \ (\varepsilon'_f)^{n'} \ (2N_f)^{2b} \right\}^{1/2} \qquad (9)$$

where $S_{\bar{a}}$ is the applied stress amplitude for fully reversed loading.

Alternatively, one can use the fatigue data for unnotched specimens tested under fully reversed loading in the nominally elastic range. To obtain a value of $S_{\bar{a}}$ for any lifetime, simply divide the unnotched stress value by K_N. Then to obtain the value of S_a for that lifetime corresponding to the local R value, reduce $S_{\bar{a}}$

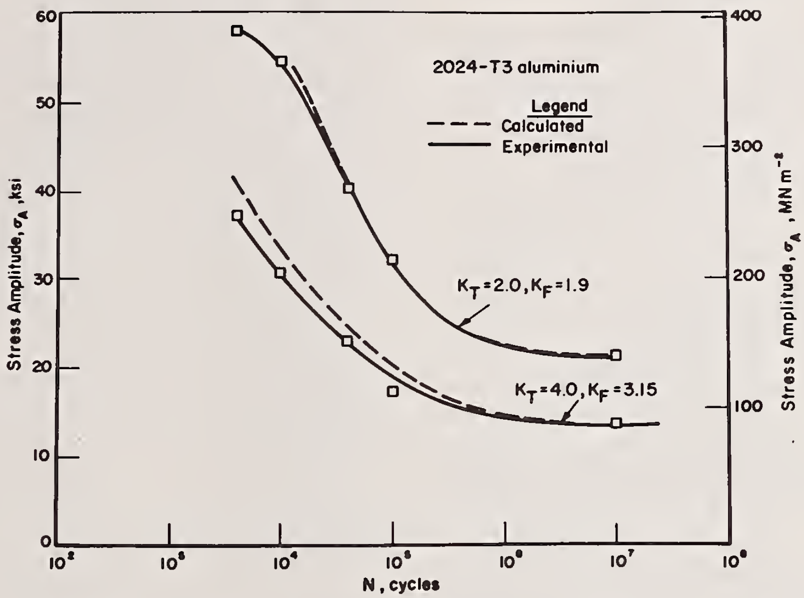


FIGURE 11. *Comparison between experimental and predicted fatigue lives for two stress raisers in an aluminum alloy.*

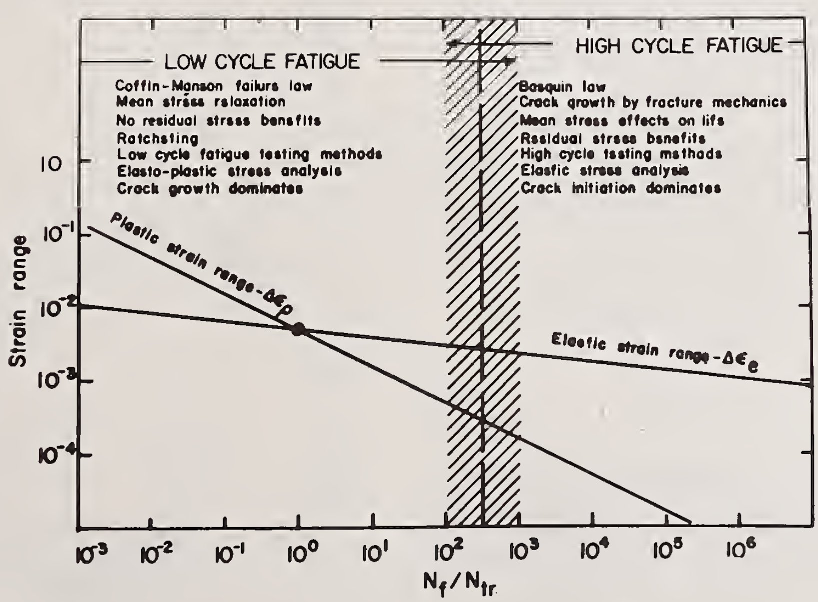


FIGURE 12. *Characteristics of low and high cycle fatigue ranges as a function of the transition life.*
After Coffin [20].

by the ratio of $\sigma_a/\sigma_{\bar{a}}$ given in eq (5). This latter procedure was followed to prepare the curves shown in figure 11. For the low stress range the local yield stress is not exceeded and the following relation is used:

$$K_{F_0} = K_F \left(\frac{\sigma_u + \dfrac{\sigma_{\bar{a}u}}{K_F}}{\sigma_u + \sigma_{\bar{a}u}} \right) \qquad (10)$$

where K_{F_0} is the fatigue notch factor for $R = 0$ loading, and $\sigma_{\bar{a}u}$ is the fatigue strength of an unnotched specimen at a specified number of cycles for $R = -1$ loading. It is noted that as the applied stress increases, the local R value changes rapidly and takes on a value of -1.

These are the basic ingredients of the safe-life design approach. In certain cases other factors may also have to be considered. These may include fretting fatigue, contact fatigue, corrosion fatigue, and cyclic stability, for example. The principal subdivision of the safe-life approach is based upon consideration of whether the fatigue process is of a low cycle or high cycle nature. A distinction between these two can be made by comparing the elastic and plastic strain amplitudes. The low cycle fatigue region is that region in which the plastic strain amplitude is larger than the elastic, and the high cycle fatigue region exists where the elastic strain amplitude is larger than the plastic. At some strain the elastic and plastic strain amplitudes are equal as seen in figure 6. Figure 12 [20] summarizes the important characteristics associated with low and high cycle fatigue as related to the transition lifetime.

3. Fatigue Crack Propagation

Consideration of fatigue crack propagation is important in the damage-tolerant approach to fatigue design. An analytical approach to crack propagation can be obtained by the application of fracture mechanics concepts to this subject. The most important aspect of the use of fracture mechanics is the single-valued correlation in the linear elastic range between the stress intensity factor and the rate of fatigue crack growth, $\Delta a/\Delta N$, where a is the crack length and N is the number of cycles. The stress intensity factor, K, is related to the stress concentration factor, K_σ, through the following definition of the stress intensity factor.

$$K = \lim_{\rho > 0} K_\sigma \sigma \frac{\sqrt{\pi \rho}}{2} \qquad (11)$$

with K_σ and σ based upon the gross cross-sectional area, and ρ is the tip radius. For the case of a central slit in a sheet specimen subjected to tensile loading at right angles to the slit the expression for K (with the crack length much smaller than the specimen width) becomes

$$K = \sigma\sqrt{\pi a} \qquad (12)$$

In general the stress intensity will be of this form, but modified to account for particular geometry, in which case we can write

$$K = \sigma\sqrt{\alpha \pi a} \qquad (13)$$

where α is a factor which takes into account the particular geometry. Consideration is now being given to the effect of structural modifications such as stringers on the rate of crack growth; however, the bulk of research in the past has involved specimens of constant thickness in which material response rather than structural response can be established.

The characteristic dependence of the rate of fatigue crack growth on the stress intensity factor is indicated in figure 13. There are two asymptotic limits to the curve. The upper limit is set by the fracture toughness of the material, K_c. The lower limit is referred to as the threshold for crack growth, K_{TH}. This latter quantity is currently of considerable interest, for a new design philosophy based upon the quantity is emerging. Many parts before going into service already contain crack-like defects as in the case of welded joints. Some of these parts may have to be designed for infinite safe-life. In the absence of defects this would entail designing at stresses based on the 10^8 life of the S–N curve, but in the presence of defects the new approach is to insure that the stress intensity associated with defects is kept below the threshold level for crack growth. Wells [21] has considered the implications of this approach for both internal and surface flaws. The present capability for the detection of internal flaws varies from

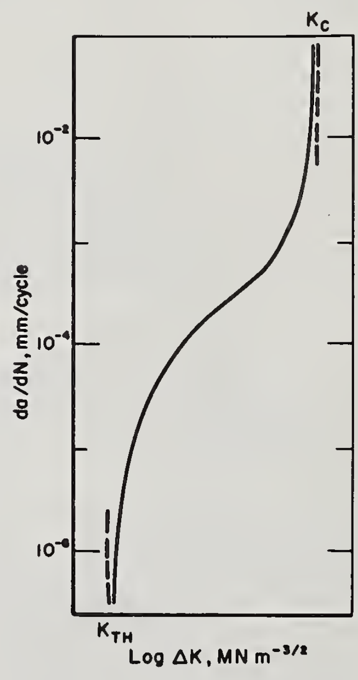


FIGURE 13. *Schematic dependence of the rate of crack growth, $\frac{da}{dN}$, on the stress intensity factor K.*

0.5 mm up to 5 mm depending upon particular conditions. If design were based upon the maximum size flaw in the most critical position, the allowable stresses would be unacceptably low. Rather than use such low stresses, material processing is controlled so that the probability of such damaging defects is extremely low. The capability for detection of surface cracks is much greater than for internal cracks, with surface crack lengths of 0.05 mm being within the range of current capabilities. For these smaller flaws the fracture mechanics approach could be used in the control of crack growth even above the threshold level. However, for these small surface cracks there may still be difficulties with the use of the fracture mechanics approach as indicated in figure 14 [22] where it is seen that surface flaws grow more rapidly than do through-cracks at low stress-intensity values. Such a lack of correlation with the stress intensity factor suggests that the continuum approach may break down for crack sizes comparable to metallurgical features such as grain size, for example. Application of the fracture mechanics approach above the threshold level will be therefore more straightforward in dealing with structures which contain cracks of more significant size. For example, the large scale of an aircraft wing structure favors the approach, where the small scale of an aircraft turbine blade may restrict the approach.

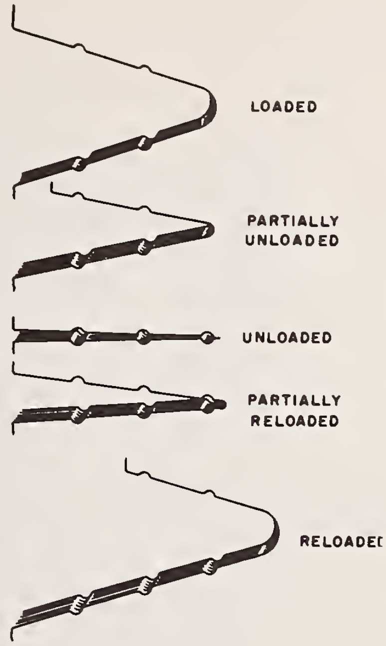


FIGURE 15. *Schematic of the fatigue crack growth process during a loading cycle in the absence of static modes of separation* [23].

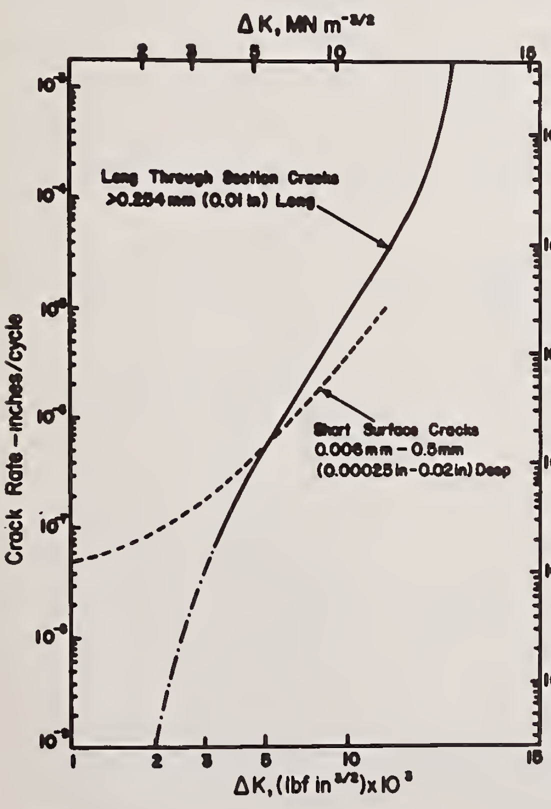


FIGURE 14. *Comparison of crack growth rate as a function of K for short surface cracks and through cracks.*
After Pearson [22].

The rate of fatigue crack propagation can be expressed as a function of the stress intensity factor, but to put such an expression on a rational basis, consideration should be given to the modes of separation involved. At low crack growth rates, i.e., below 10^{-4} in at $R = 0$, a ductile mode of crack advance associated with plastic blunting on loading, and tip sharpening on unloading as illustrated in figure 15, is thought to be dominant [23]. At higher rates of crack growth where the peak stress intensity approaches the fracture toughness, static modes of separation such as ductile rupture become operative and accelerate the rate of crack growth. The following expression based on crack-opening displacement considerations has been developed [19] to account for the contribution of each of these modes to the amount of crack growth per cycle:

$$\frac{\Delta a}{\Delta N} = \frac{A}{\sigma_y E}(\Delta K^2 - \Delta K^2_{TH})\left(1 + \frac{\Delta K}{K_c - K_{max}}\right) \quad (14)$$

Since $K_{max} = \dfrac{\Delta K}{1 - R}$, the effects of mean stress are incorporated in the equation. Figures 16 and 17 show a comparison with predicted and experimental results for two aluminum alloys of different toughness levels. Note that mean stress effects are much more pronounced in the case of the low toughness alloy, and they are virtually absent in the high toughness alloy. The threshold level in both alloys is sensitive to mean stress, and current research is aimed at understanding this dependency. Crack growth rates near the threshold are

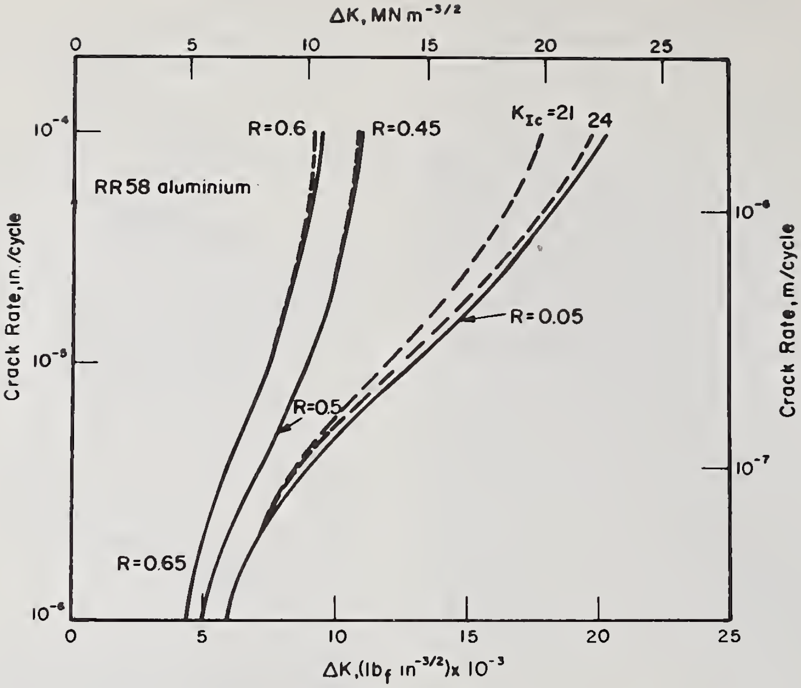


FIGURE 16. *Dependence of the rate of crack growth on mean stress and the range of the stress intensity factor for a low toughness aluminum alloy.*
Experimental results after Pearson [25] shown in solid lines; equation (14), dashed lines.

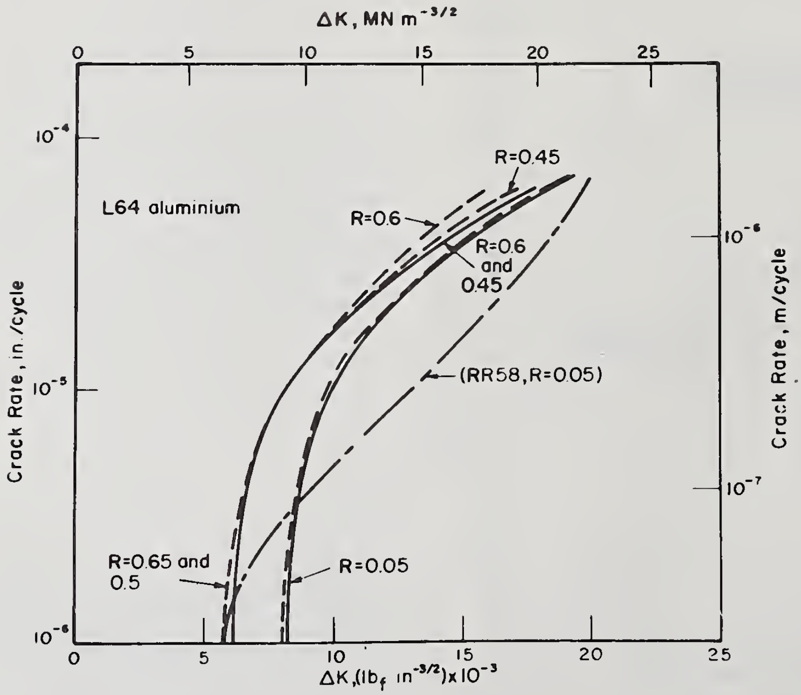


FIGURE 17. *Dependence of the rate of crack growth on mean stress and the range of the stress intensity factor for a high toughness aluminum alloy.*
Experimental results after Pearson [25] shown in solid lines; equation (14), dashed lines.

of the order of 10^{-8} in per cycle or less, representing increments of growth on the order of the interatomic spacing. Further, since oxides can form very rapidly in metals tested in air, the growth must be occurring primarily in the surface oxide at these low rates rather than in the metal itself. The appearance of the fracture surface should, therefore, be influenced since oxide rupture is a brittle process compared to the base metal blunting process. The characteristic fatigue striations are not observed at such low growth rates since they are resolvable only at crack growth rates above 10^{-6} in per cycle, but even in this range their ease of detection will be material dependent.

The effects of variable amplitude loading on the rate of fatigue crack growth are of interest. Barsom [24] has recently reported on crack growth tests of an A514–B steel (yield strength 129 ksi) subjected to random loading. He found that the rate of crack growth could be expressed as

$$\frac{\Delta a}{\Delta N} = A' \left(\Delta K_{rms}\right)^n \tag{15}$$

where A' and n are material constants, and ΔK_{rms} is the root-mean-square stress-intensity-factor fluctuation. It is of interest to compare predictions based upon eq (14) with this result. Since Barsom observed that crack growth in this steel was independent of mean stress for crack growth less than 10^{-4} in per cycle, the contribution due to static modes in eq (14), i.e., the $\frac{\Delta K}{K_c - K_{max}}$ term, can be neglected. An average rate of crack growth under variable amplitude loading, $\left(\frac{\Delta a}{\Delta N}\right)_{AVG}$, can be taken to be equal to the average contribution from each of the cycles over the increment of crack growth considered, that is

$$\left(\frac{\Delta a}{\Delta N}\right)_{AVG} = \frac{A}{\sigma_y E}\sum_{i=N}^{i=N+\Delta N} \frac{\Delta K^2_i}{\Delta N} - \Delta K^2_{TH} \tag{16}$$

Note that the quantity $\frac{\Sigma \Delta K^2}{\Delta N}$ is the square of ΔK_{rms}, so that eq (16) can be written

$$\left(\frac{\Delta a}{\Delta N}\right)_{AVG} = \frac{A}{\sigma_y E}\left(\Delta K^2_{rms} - \Delta K^2_{TH}\right) \tag{17}$$

Figure 18 shows a comparison between eq (15) and eq. (17) and the data obtained by Barsom. (In evaluating eq (17), A was taken to be 0.023 and ΔK_{TH} to be 8 ksi$\sqrt{\text{in.}}$) Agreement of both equations is quite good; however, for other more extreme types of loading the agreement may not be as good. For example, a single overload can greatly retard the rate of crack growth rate at a lower amplitude, and the above equations would over-estimate the rate of crack growth. Nevertheless, for a large number of random cycles, the amplitudes of which do not fluctuate too widely, the agreement shown above indicates that reasonable predictions can be made.

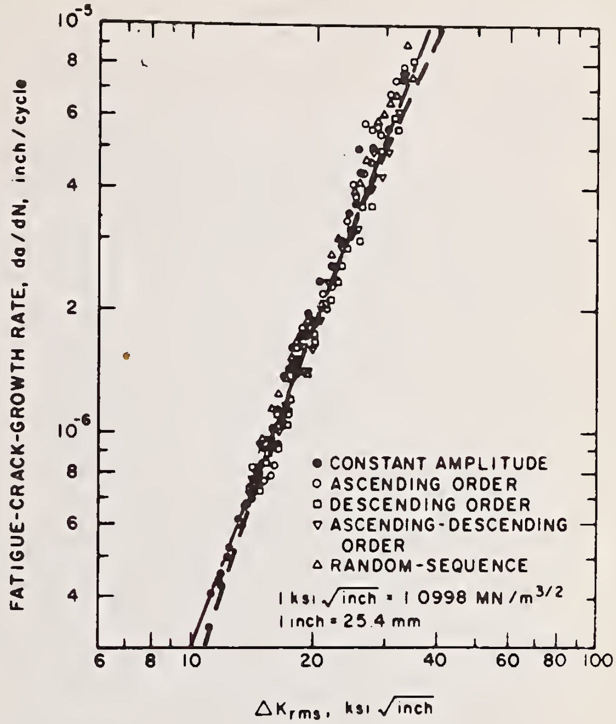


Figure 18. *Dependence of the rate of crack growth for variable amplitude loading on* ΔK_{rms}.
Equation 15 represented by the solid line; equation (17) represented by the dashed line. Data after Barsom [24].

4. Concluding Remarks

This brief review has primarily surveyed analytical approaches to fatigue. To recognize that this review is not in depth, one has only to note that over 125 papers on various aspects of the subject were presented at a recent conference. However, much of the nature of the current approach, if not all of the specific detail, has been indicated. In recent years, improvements in analysis rather than improvements in materials resistance to fatigue have been obtained. While it is to be hoped that material improvements will be forthcoming, thus far, this goal has been a difficult one to attain. Perhaps more modest goals such as reduction in scatter, in notch sensitivity, and in environmental sensitivity may be more realizable and useful than improvements in average properties. At any rate, a wider dissemination of information already available about fatigue design would constitute a positive step toward the elimination of fatigue failures.

The support of this study under Air Force Grant No. AFOSR–74–2703 is gratefully acknowledged.

5. References

[1] Weck, R., Failure of steel structures: causes and remedies, Proc. Roy. A **285,** No. 1400, (6 April 1965). p. 3.

[2] Freudenthal, A. M., Fatigue mechanisms, fatigue performance and structural integrity, Proc. of the Air Force Conference on Fatigue and Fracture of Aircraft Structures and Materials, Ed. by Wood et al, AFFDL TR 70–144, (1970), p. 9.

[3] Dolan, T. J., So, you are going to testify as an expert, ASTM Standardization News **1,** (1973), p. 30.

[4] Speigelman, A., An insurance approach to product liability, ibid, p. 26.

[5] Hardrath, H. F., A unified technology plan for fatigue and fracture design, presented at the Seventh ICAF Symposium, London (July 1973).

[6] Hardrath, H. F., A guide for fatigue testing and statistical analysis of fatigue data, ASTM STP 91–A (1963).

[7] Coffin, L. F., The effect of vacuum on the high-temperature, low cycle fatigue behavior of structural metals corrosion fatigue, NACE–2 Nat. Assoc. of Corrosion Engineers, Houston, 1972, p. 590.

[8] McEvily, A. J. and Crosby, S. R., Lifetime prediction methods for elevated temperature fatigue, presented at the AGARD Specialists Meeting on Low-Cycle High Temperature Fatigue, Washington, D.C., April 1974, to be published.

[9] Manson, S. S., The challenge to unify treatment of high temperature fatigue—A partisan proposal based upon strainrange partitioning, in Fatigue at Elevated Temperatures, ASTM STP **520**, (1973), p. 744.

[10] Lipson, C. and Juvinall, R. C., Handbook of Stress and Strength (Macmillan Co., New York, N.Y.).

[11] Morrow, J., Cyclic plastic strain energy and the fatigue of metals, ASTM STP **378** (1965), p. 45.

[12] Landgraf, R., The resistance of metals to cyclic deformation, ASTM STP **467** (1970), p. 3.

[13] Landgraf, R., Fatigue Design Handbook, Society of Automotive Engineers, 1968.

[14] Buxbaum, O., Methods of stress-measurement analysis for fatigue life evaluation, in Fatigue Life Prediction for Aircraft Structures and Materials, AGARD–LS–62, 1973.

[15] Dowling, N. E., Fatigue failure predictions for complicated stress-strain histories, J. Mater. ASTM **7**, (1972), p. 71.

[16] Neuber, H., Theory of stress concentration for shear strained prismatical bodies with arbitrary nonlinear stress strain law, J. Appl. Mechanics, ASME **28**, (1961), p. 544.

[17] Wetzel, R. M., Smooth specimen simulation of fatigue behavior of notches, J. Mater. **3**, (1968), p. 646.

[18] Crews, J. H., and Hardrath, H. F., A Study of cyclic plastic stresses at a notch root, Exp. Mech. **6**, (1966), p. 313.

[19] McEvily, A. J., Phenomenological and microstructural aspects of fatigue, presented at the Third International Conference on the Strength of Metals and Alloys, Cambridge, 1973, to be published.

[20] Coffin, C. H., Fatigue at high temperature—prediction and interpretation, Proc. Inst. of Mech. Eng. **188**, (1974), p. 109.

[21] Wells, C. H., Design procedures for elevated temperature low-cycle fatigue, Presented at the AGARD Specialists Meeting on Low-Cycle High Temperature Fatigue, Washington, D.C., April 1974, to be published.

[22] Pearson, S., Initiation of fatigue cracks in commercial aluminum alloys and the subsequent propagation of very short cracks, R.A.E. Tech. Report 72236, Sept. 1973.

[23] McEvily, A. J., Boettner, R. C., and Johnston, R. C., On the formation and growth of fatigue cracks in polymers "in" fatigue—an interdisciplinary approach, Ed., Burke et al. (Syracuse University Press, 1964), p. 95.

[24] Barsom, J. M., Fatigue crack growth under variable amplitude loading in ASTM A514–B Steel, ASTM STP **536**, (1973), 147.

[25] Pearson, S., The effect of mean stress on fatigue crack propagation in half-inch thick specimens of aluminum alloys of high and low fracture toughness, Eng. Fract. Mech. 4, (1972), p. 9.

6. Discussion

J. M. Karhnak, Jr., U.S. Army, Ft. Belvoir: You showed several different curves that looked like hysteresis-type curves in which you said that different time durations led to different reactions, etc. Who is doing further work in this area?

A. J. McEvily: The group that worked with Manson at NASA, Lewis Research Laboratory. A rather comprehensive treatment of the subject is available in ASTM STP–520, *Fatigue at Elevated Temperatures.*

A. Beerbower, Esso Research & Engineering Company: We've been running into something we call sliding fatigue; fairly high stresses approaching a plastic region, very environmentally responsive, and tending to produce spallation. We have run into a lot of criticism for calling this fatigue because it only takes between 2 and 30,000 cycles to generate complete surface roughness, sometimes called scuffing. Do you think we could be seeing what you are calling low cycle fatigue?

A. J. McEvily: Yes, I think so. There is an area called fretting fatigue which seems to be close to what you are describing. I also think that a large number of what we would call wear failures, like in gear teeth, are actually contact fatigue failures. Each time the contact is made, a Hertzian type of stress system is developed which leads to cracking on a micro-level. If you want to broaden what the category of fatigue might include, I would think what you described as sliding or anything that is repetitive of a contact nature could have an ingredient of fatigue associated with it. In some recent scanning electron microscopy studies, I was struck by the fact that a wear process looked very much like a contact fatigue process involving subsurface nucleation of a crack and a sheet of material being removed from the surface.

MODES OF FAILURE

Session II

Chairman: Frederick F. Ling
Rensselaer Polytechnic Institute

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Failure by Stress Corrosion Cracking—Current Approaches Toward Failure Prediction*

Jerome Kruger

Institute for Materials Research
National Bureau of Standards
Washington, D.C. 20234

Stress corrosion cracking (SCC) produces failures in a material when it is subjected to the combined effects of mechanical stress and reaction with an environment. All proposed mechanisms of SCC seek to explain how the combination of a given level of stress, a particular material, and a given environment can lead to the initiation and propagation of cracks. The three major categories of mechanisms that are generally proposed are (1) active path dissolution, (2) stress-sorption, and (3) embrittlement.

In recent years attempts to determine the mechanisms have been enhanced by new concepts and techniques that flow from the three major scientific disciplines that underlie the complex problem of SCC—chemistry, metallurgy, and mechanics. From chemistry have come two main currents: (a) increased awareness of the importance of characterizing the altered environment inside a growing crack, and (b) a recognition of the importance of the regrowth rate of a protective film on a bare surface exposed when that film is broken by stress. The main emphasis in the metallurgy of SCC has been on the crucial role played by structure in general, and particularly at the tip of a crack. Aiding this objective has been the use of the high voltage electron microscope which can directly look at the interaction of the environment with structural defects, e.g., dislocations. Finally, the major thrust in the mechanics of SCC has been the application of the concepts and techniques of fracture mechanics.

The impact of these new concepts and measurement techniques on predicting and preventing SCC failure will be discussed.

Key words: Corrosion; failure prediction; film rupture-metal dissolution; hydrogen embrittlement; stress corrosion cracking; stress-sorption

1. Introduction

Far too often the failure of a metal part involves an alloy that was chosen because of its good corrosion resistance to the environment in which it is designed to operate. Moreover, it failed at stress levels well below its normal fracture stress. More than likely, this and similar such unexpected failures were due to stress corrosion cracking (SCC). The insidious nature of this mode of failure results from the situation just described where, in the absence of the possibility for SCC, one would expect both good fracture and corrosion resistance. Figure 1 from a paper by Staehle [1] [1] illustrates how SCC markedly reduces the load carrying cross section of a part over that to be expected from environmental attack (corrosion) alone or from purely mechanical failure. Thus, it is from this synergistic interaction of corrosion and stress that SCC arises, and the definition of the phenomenon reflects this fact. For example, the ASTM Committee on the Corrosion of Metals defines it as "a cracking process requiring the simultaneous action of a corrodent and sustained tensile stress. This excludes corrosion-reduced sections which fail by fast fracture. It also excludes intercrystalline or transcrystalline corrosion which can disintegrate an alloy without either applied or residual stress." This definition embraces many of the other terms that focus on particular mechanisms such as "hydrogen embrittlement," "caustic embrittlement," "season cracking," etc.[2] This diversity of terms has tended to produce confusion. Many workers, for example Staehle [1] and Brown [2], have proposed that the term "stress corrosion cracking" be used as a generic term to avoid entangling and confusing the macro phenomenon with any special mechanism that may govern a particular situation. This proposal is being increasingly adopted in the modern literature.

Because of the diversity of mechanisms affecting SCC, and because it takes place in what are thought to be innocuous environments, this mode of failure was not generally considered too important as late as 1940. In a discussion accompanying a 1940 paper by Hodge and Miller [3] which clearly highlighted the problems of SCC, Brooks pointed out that ". . . the industrial implications of this phenomenon do not loom large. Failures are rare and will become more so as understanding spreads. . . ." Not only were his pre-

*This presentation was previously published as a part of NBSIR 74–583.
[1] Figures in brackets indicate the literature references at the end of this presentation.

[2] It does not include "corrosion fatigue" which involves a cyclic stress rather than a sustained static stress.

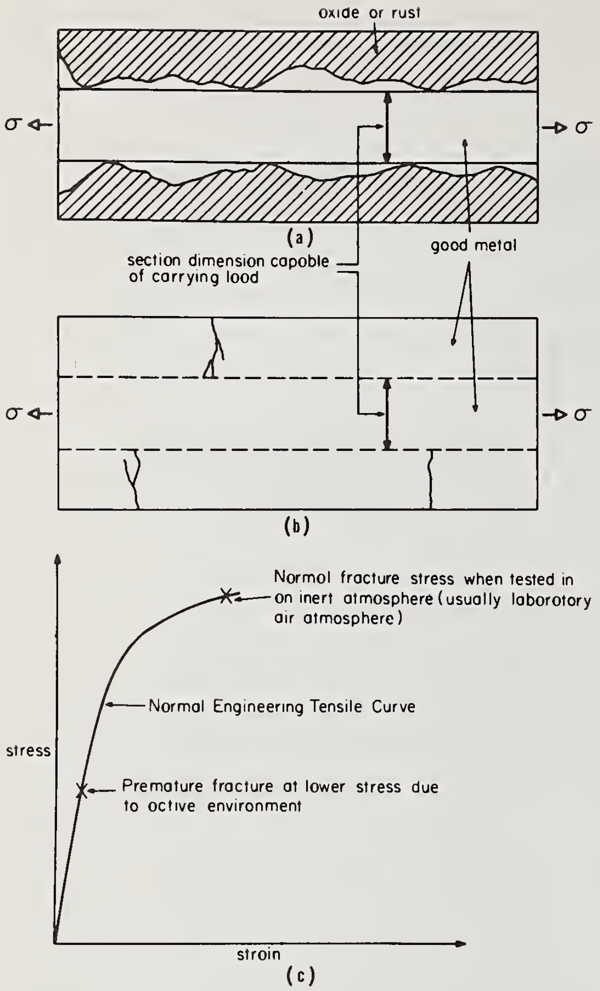


FIGURE 1. *A comparison in the reductions in load-carrying cross section as a result of (a) general corrosion and (b) stress corrosion.*
The effect of stress corrosion on the tensile curve is shown in (c). From Staehle [1].

dictions not realized, but the systems where SCC is recognized as the cause of a failure have increased markedly throughout this century. It was first recognized at the end of the 19th century when brass cartridge cases (containing residual stress) failed because of exposure to the ammonia present in stables. Table 1 documents the marked increase in the recognition of SCC as a source of failure, going from the days of transportation by horse-drawn vehicles to the space vehicles of today. The number of systems where previously failures were not expected continues to grow, especially as greater demands are made on materials.

Because of the increased identification of SCC as a major cause of failures, there has resulted a significant increase in research directed toward understanding the phenomenon and developing mechanisms to explain it. Two recent proceedings of major meetings [5,6] and a forthcoming one [7] give excellent accounts of

TABLE 1. *History of Stress Corrosion Problems*
(Adapted from Brown[4]*)*

When	What
Late 19th Century	Brass in ammonia
End 19th Century	Boiler steels in caustics
Early 20th Century	Mild steels in nitrates
Early 20th Century	Aluminum alloys in moist atmospheres
1930's	Stainless steels in chlorides
1930's	Magnesium alloys in moist atmospheres
1950's	Martensitic steels in aerospace environments
1950–60's	Titanium—Hot salts, N_2O_4, methanol

the current state of thinking on the origins of SCC. This paper will briefly outline the major mechanisms that are currently believed to explain SCC and then show how these mechanisms have pointed the way to the development of new experimental approaches which, besides enhancing our understanding of SCC, have, more importantly, provided us with opportunities for the development of new capacities for failure prediction and prevention.

2. Mechanisms

In order to consider the important mechanisms that have been proposed to explain the origins of SCC, we must first list the characteristics, most of which must be present to produce this mode of failure. This is necessary because a valid mechanism must explain the necessity for these characteristics to be present if SCC is to occur. Brown [4] has listed the following features generally common to SCC:

(1) The existence of a tensile stress either applied or present as a residual stress.

(2) The alloy usually shows good resistance to corrosive attack by the environment causing cracking. This is why alloys chosen because of their corrosion resistance to an environment may be subject to SCC.

(3) SCC for a given alloy usually occurs more readily in a rather specific environment. Moreover, the key damaging species in that environment need not be present in large concentrations. Recent studies [9,10] have indicated that the dependence of SCC on environmental specificity is not as pronounced as originally thought in early SCC studies.

(4) The microscopic appearance of stress corrosion cracks is that of a brittle fracture in spite of the fact that the metal itself may still be quite ductile in a mechanical test. The cracks frequently exhibit multiple branching, but this does not always occur.

(5) In order for SCC to occur a threshold stress or stress intensity must be exceeded.

To Brown's list must be added another characteristic becoming increasingly recognized [9,11]. There exists a critical potential. Any metal in an environment where its potential lies below this critical potential is thought by some to be immune from SCC. This is a more recently proposed characteristic that has been strongly advocated by Uhlig [11].

At one time it was believed that SCC could not occur for pure metals, but some recent work [13,14] has found SCC for high purity copper. As with all the other characteristics, one can usually find exceptions.

Using these characteristics of SCC as a basis, a number of mechanisms have been developed. At one time some believed that there existed a universal mechanism that applied to all systems where SCC was found. It is now widely held that SCC can be due to a number of different mechanisms, and that there are alloy-environment systems where one or more of the mechanisms outlined here may apply. Of the many proposed, three major mechanisms are currently the ones most actively used. Moreover, the others proposed can be considered as variations or sub-sets of the three major ones. All of the mechanisms involve the complex interactions of the three disciplines that control SCC: chemistry (electrochemistry), metallurgy, and mechanics. None of the mechanisms will be discussed in great detail, and a critical evaluation of their pros and cons is outside the scope of this paper. Instead these mechanisms will be briefly described, their strong and weak points detailed, their failure prevention implications indicated, and, in the next section of this paper, the techniques that they point to as possible tools for failure prediction will be discussed.

Stress-Sorption Mechanism—This mechanism, most actively advocated by Uhlig [14,15], proposes that a specific species adsorbs and interacts with strained bonds at the crack tip causing a reduction in bond strength, or, thermodynamically stated, a reduction in the surface energy, γ, which leads to a lowering in the stress required to produce brittle fracture. The relationship between the theoretical stress to produce brittle fracture, σ_f, and γ is given by the well-known expression

$$\sigma_f = \left(\frac{E\gamma}{d}\right)^{1/2}, \tag{1}$$

where E is the elastic modulus and d is the atomic spacing. The adsorption of a damaging species at a crack tip lowers σ_f and therefore the threshold stress for SCC is the minimum value to which γ is lowered by adsorption. Figure 2 outlines the mechanism schematically. Uhlig proposes that an important factor arguing for this mechanism is the existence of a critical potential below which SCC does not occur. He explains that the effect of potential can be attributed to the strong dependence of adsorption on potential. He bases this on many studies by electrochemists who have found that the adsorption of an anion (usually

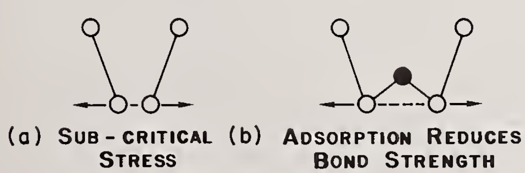


Figure 2. *Stress-sorption mechanism for SCC.*
(a) Chemisorption of ion at tip of crack.
(b) Weakening of interatomic bond at the tip of the crack so that applied tensile stress leads to the breaking of this bond.

the damaging species for SCC) takes place at potentials above the point of zero charge for a given surface in a given environment. (This strong dependence of failure on potential explains why SCC can be prevented by cathodic protection in some systems.) Environment plays a role in the stress-sorption mechanism because of the specificity of adsorption of a particular species on a surface, and because environment brings the potential to a value where adsorption can or cannot take place. Moreover, if one can introduce into the environment an extraneous species that displaces the damaging bond-weakening adsorbed species, cracking becomes more difficult. Uhlig [14] gives examples where such extraneous species inhibit SCC. The stress-sorption mechanism is more widely accepted in the case of liquid metal embrittlement or stress cracking of plastics. In the former case, for example, Preece [15] points out that the theory is successful in explaining the observations of a brittle to ductile transition with increase in temperature (desorption), an "instantaneous" failure mode, the increase in severity of embrittlement with increase in strain rate and flow strength, and, possibly, specificity.

For SCC in aqueous environments, however, a number of problems with the stress-sorption mechanism have been pointed out [9,10]. Among them are the following:

(1) In the cracking of ductile alloys, the ductility present is sufficient to produce extensive plastic deformation at a crack tip. The blunting that this produces eliminates the infinite sharpness required for this mechanism to work.

(2) The large amount of metal dissolution observed in many straining electrode experiments (to be discussed later) poses the question of what happens when the metal dissolves out from under the adsorbed film.

(3) Crack growth rates are considerably slower than one would expect from this model.

(4) Studies [17] have shown that thin protective films exist on metals in SCC environments, and that these films reform rapidly, on the order of milliseconds in some cases. The weakening of metal bonds requires that rapid direct access to the base metal surface is necessary, but the existence of these films would prevent or strongly influence this adsorption process.

The measures suggested by the stress-sorption mechanism to promote SCC failure avoidance are as follows:

(1) Lower the potential of a metal to be protected below the critical potential for SCC either by coupling another more active metal to it or by cathodic protection using an applied current.

(2) Introduce inhibiting anions in the environment.

To properly apply these measures requires techniques that enable both the determination of the critical potential for SCC and the potential of the metal in an environment. The latter requirement is not as easily met since the environment that counts is that environment existing in flaws or cracks. In addition to a realistic characterization of the environment, there exists an equal need for realistic characterization of the stress system acting on a crack tip, since the model requires adsorption at strained bonds or dislocations. Thus, techniques that yield both the geometry of stress application and expected dislocation configuration would be valuable for failure prediction using the guidelines provided by this mechanism. Finally, ways to measure the rate of adsorption of the relevant species as a function of stress geometry and potential would be invaluable if this mechanism has merit.

Film Rupture-Metal Dissolution Mechanism—This mechanism was originally suggested by Logan [18] and Champion [19]. It has also been refined and extended by a number of workers [10,17,20–24]. The mechanism proposes that the protective film normally present on a metal surface (if it were not present the metal would fail by corrosion alone) is ruptured by continued plastic deformation at a crack tip where the exposed metal becomes a very small and restricted anodic region where metal dissolution takes place. The rest of the metal surface, especially the walls of the crack, acts as a cathode. Susceptibility depends on the rate at which the metal exposed by film rupture is repassivated vis à vis the rate of metal dissolution. Figure 3 gives a schematic representation of this mechanism. A clear picture of this process has been provided by Vermilyea [24] in a recent paper where he lists the following five steps in the process that leads to crack extension:

(1) The film is ruptured at time zero.

(2) Crack growth by dissolution occurs for a distance of L during the time of repassivation t_L.

(3) The dissolution process removes the material at the old crack tip and thereby creates a strain transient just ahead of the newly created crack tip as shown in figure 4.

(4) At a time t_L a critical strain builds up to a value sufficient to cause rupture of the newly reformed film at the base of the new crack.

(5) The process repeats itself, the crack propagation rate, $\frac{dL}{dt}$, given by Vermilyea as

$$\frac{dL}{dt} = \frac{L}{t_c}. \qquad (2)$$

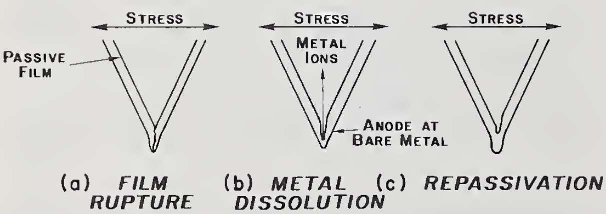


FIGURE 3. *Film rupture-metal dissolution mechanism for SCC.*
(a) Rupture of the protective film at tip of a crack by the application of tensile stress.
(b) Dissolution of the metal exposed when film is ruptured. Tip of crack is an anode and film covered walls of the crack are a cathode.
(c) Repassivation of the exposed metal at the tip of the crack.

A sub-set or variation of the active path dissolution mechanism just described is the brittle-film mechanism [25,26]. It differs from the dissolution mechanism in that crack advance is mainly by the production of a penetrating brittle film which undergoes fracture when tensile stress is applied. The crack produced enters the ductile metal below this brittle film and a new layer is formed on the bare metal so produced. The process continues by a repetition of the film fracture-reformation cycle.

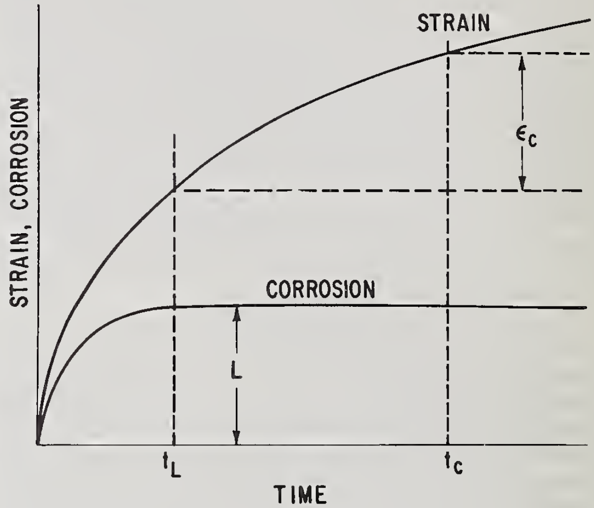


FIGURE 4. *Time dependence of events occurring after film rupture. Repassivation time is* t_L.
The time to produce the critical strain, ϵ_c, necessary to cause film rupture is t_c. From Vermilyea [24].

One can conceive of at least three factors that affect the general mechanism.

(1) The rate of production of bare sites, which is dependent on the magnitude of the stresses applied, the metallurgy of the material, and the geometry of the stressed system (the value of t_c is important here).

(2) The rate at which the bare sites produced by the applied stress are repassivated (t_L is important here).

(3) The resistance of the protective film (its ductility) to rupture, see [24] (again t_c is important).

The nature of the species in the environment has an effect on t_p, or probably more importantly, the rate of repassivation [17]. This is so because small impurities can affect the composition of the newly reforming film and rates of growth are sensitive to the presence of impurities [27]. Also, environment can affect t_c through its influence on film ductility [28]. Potential affects both film growth kinetics and metal dissolution rate so that the balance of these two competing factors is important, the latter affecting L, the distance of crack advance. The concept of a critical potential for SCC is as valid for the film rupture-metal dissolution model as it is for the stress-sorption mechanism.

There are a number of facts that create problems in accepting the film rupture-metal dissolution model for all cases of SCC. Among the most important are the following:

(1) The quasi-cleavage character of some SCC transgranular fracture surfaces indicates brittle failure rather than dissolution.
(2) Crack propagation occurs at a surface which is itself rapidly dissolving at a high rate [29].
(3) There is a possible discrepancy between crack growth rate and the currents observed in straining electrode experiments which purport to measure the current involved in dissolution at crack tips [22].

The measures suggested by the active path dissolution model to promote SCC failure avoidance will include one suggested for the stress-sorption model, i.e., control of potential to a value below the critical potential for SCC. Another measure is the control of repassivation kinetics by environment or alloy modification. In order to apply the latter measure for SCC avoidance, techniques for measuring repassivation kinetics as a function of potential and environment are required. However, in order to make such measurements one must, as suggested earlier for stress-sorption, be able to characterize the environment and potential existing in cracks. Finally, and most difficult, one must have a technique for determining t_c, the time required to build up a rupturing stress in a film as a function of the metallurgy of the alloy whose failure we wish to prevent. This requires the measurement of the ductility of the protective film on an alloy and the geometry of the stress system acting on a crack.

Hydrogen-Embrittlement Mechanism—This is considered by some to be a different phenomenon from SCC, but since cathodic reactions (which produce hydrogen under the proper conditions) are as much a part of corrosion as anodic ones (producing films and/or metal dissolution), it is difficult to exclude it in a discussion of SCC. In this mechanism fracture results from the production of a brittle region at the crack tip because of the introduction of hydrogen into the alloy via cathodic reactions (see fig. 5). The production of this brittle region has been attributed to a number of causes among which decohesion [30], new phase formation [31–33], and pinning of dislocations [34] can be mentioned. The environment can affect the process by influencing the rate of production of hydrogen ions and by enhancing or inhibiting the entry of hydrogen into the metal by the interposition of adsorbed or reaction layer films. In aqueous systems the potential at the crack tip determines whether the production of hydrogen at the surface is thermodynamically feasible. The composition of the environment influences what potential is necessary to produce hydrogen, its rate of production, and by means of the presence of inhibiting or enhancing species, whether it can readily enter the metal lattice. In gaseous systems, hydrogen pressure plays the some role as potential in aqueous environments. Inhibiting species, e.g., oxygen, can also be added to gases [35].

It has been suggested by some workers, for example, Troiano and co-workers [36], that almost all aqueous SCC is hydrogen embrittlement. They found that austenitic stainless steels, which have been assumed to be non-susceptible to hdrogen embrittlement, could be made less ductile by hydrogen charging, and that one could observe the permeation of hydrogen at positive currents (where one should only expect metal dissolution). Other workers [31–33,37–39] have found the formation of a phase or phases capable of undergoing hydrogen embrittlement.

All of these findings do point to greater universality of application for the hydrogen embrittlement mechanism. However, as always, there are counter arguments that rule against such universality. Recent work by Wilde and Kim [40] showed that for an austenitic stainless steel while rate of hydrogen permeation was higher at negative potentials, time to failure was also higher at these hydrogen producing potentials. This finding also agrees with results that show that cathodic protection (which should promote the production of hydrogen) does retard SCC for a number of alloys [16]. Indeed, to confuse the picture further, it has been found by Bernstein and Pickering [41] that anodic dissolution can occur in the cracks of cathodically polarized ferrous alloys.

This difficulty in determining whether hydrogen embrittlement is an operative mechanism is most important in developing measures for failure prevention. For example, unless one knows the operative mechanism, a change of the potential to what is believed to be a protective value may actually accelerate failure. Obviously techniques that determine what are the real potential and environment at the tip of the crack are crucial. Likewise, electron micrography of fracture surfaces which help in determining the mode of failure and can thereby suggest the proper prevention measures are also of utmost importance.

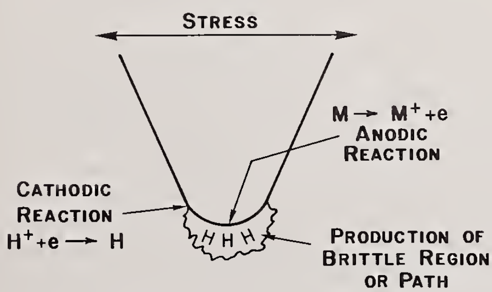


FIGURE 5. *Hydrogen embrittlement mechanism for SCC.*
Cathodic reduction of H^+ at the tip of a crack produces hydrogen which enters the vicinity of the tip and embrittles the metal immediately ahead of an advancing crack.

3. Techniques Suggested by Mechanisms

In the preceding section the discussion of the major mechanisms of SCC pointed to the types of information that are needed in order to develop our ability to predict SCC susceptibility and to institute suitable measures for prevention. In this section the techniques developed in recent years, in many instances under the influence and stimulation of the concepts of these mechanisms, will be discussed. The techniques brought into being in this manner have their roots in the three broad disciplines that control the phenomenon of SCC: chemistry, metallurgy, and mechanics.

Techniques Dealing with the Chemistry of SCC—Two areas have been attacked with increasing interest in recent years. As pointed out in the previous section they are both important in all three major mechanisms although not always well recognized. They are the techniques directed towards determining the environment in cracks and the kinetics of repassivation.

(a) The Environment in Cracks—While it has been recognized for a number of years that the environment trapped in the interior of restricted regions on a metal (cracks, pits, and crevices) has a different composition than the bulk environment surrounding the metal, it has been only in recent years that such knowledge was considered important, and serious attempts made to measure the composition of the occluded environment in a crack. Brown and co-workers [42–44], using the insights provided by the large body of work of Pourbaix (see the review in [45]) that systematized the role of pH and potential in all corrosion processes, set about to measure these parameters in SCC cracks. Their original approach was to immobilize the corrodent within SCC cracks by freezing it in liquid nitrogen, breaking open the crack mechanically, and then analyzing the thawed out solution removed from the crack. A later development is illustrated in figure 6 where microelectrodes determine the pH and potential in the crack's environment. Other developments in this fast moving field have been the use of a capillary to sample the crack solution [46] and the combination of a capillary technique with thin layer chromatography [47].

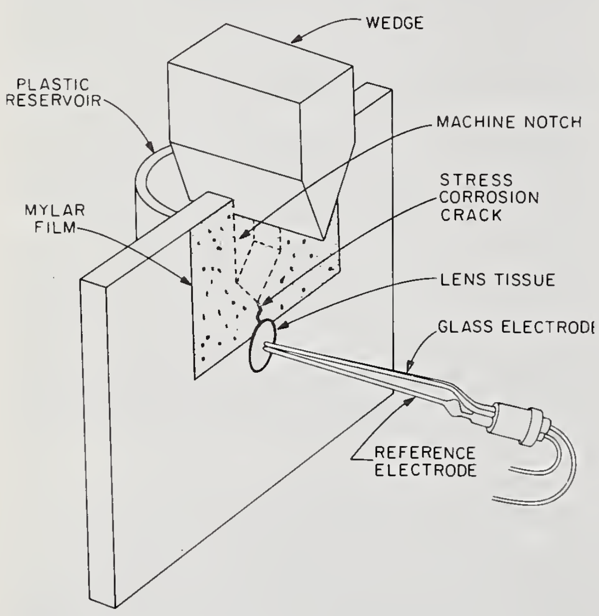


Figure 6. *Experimental arrangement for determining the acidity (pH) and potential at a crack tip.*
The glass and reference electrodes measure the pH electrochemically by sampling the solution that wets the lens tissue after the solution travels the length of the crack. From Smith, Peterson, and Brown [44].

The value of these crack chemistry techniques becomes evident when figure 7 is inspected. It shows for a large number of steels of different compositions in environments of either high or low pH that the potential and pH inside the crack falls in a region of the pH-potential diagram (Pourbaix diagram) where one expects both corrosion and, because the points on figure 7 are below the line marked H, the presence of hydrogen. Thus, any failure prevention measures using bulk composition or potentials could be based on non-realistic conditions and lead to an improper choice. A good discussion of the electrochemical ramifications of knowing the proper environment in a crack is given by Pourbaix [45.] In addition to knowing the nature of the solution in a crack and the potential at the crack tip, it is of great importance to know another aspect of the events occurring at the crack tip, e.g., how much hydrogen is entering the metal lattice? As the discussion on the hydrogen embrittlement mechanism stresses, this knowledge is of crucial importance. A tool for obtaining this information has been developed based on the work of Devanathan and Stachurski [48]. By using a thin membrane of the metal studied it is possible to measure the amount of hydrogen that is evolved on one side of the membrane when it passes through the metal by oxidizing it electrochemically in a chamber on the other side of the membrane. The anodic current density at steady state involved in the oxidation of the permeated hydrogen is a measure of the amount of hydrogen entering the lattice on the hydrogen evolution side of the membrane. A discussion of the technique, its difficulties, and validity has been given recently by Bockris [49]. Figure 8 from a paper by Kim and Wilde [40] illustrates the importance of this technique. It shows that the rate of hydrogen penetration cannot always be related to SCC susceptibility.

(b) Repassivation Kinetics—In all the mechanisms proposed for SCC, access of the environment must be provided to the metal surface. This surface, if it is suitable for withstanding ordinary corrosion, usually has a protective film. Therefore, as discussed earlier, the breaking of this film and the rate at which it regrows (repassivation kinetics) are of great importance in determining the time during which the environment can interact with the bare metal surface in any one of the different ways the various mechanisms propose.

Scully [20, 50] was one of the earliest workers to emphasize the importance of repassivation kinetics, and it has become an important element especially in the film rupture metal-dissolution mechanism because coupled with the concept of slip step dissolution it explains why restricted lateral dissolution occurs; the slower the repassivation, the more of the metal slip step that has emerged and ruptured the film can be dissolved before new film forms. If too little dissolves, no SCC occurs; if too much dissolves, the dissolved areas are not restricted enough, and the crack is too blunt to propagate. Only when the environment and alloy provide a system where the repassivation kinetics produce a highly restricted region for dissolution is SCC found. In addition to Scully, Staehle [10] has given a detailed discussion of the significance of repassivation kinetics.

These considerations have resulted in a number of techniques aimed at determining the repassivation ki-

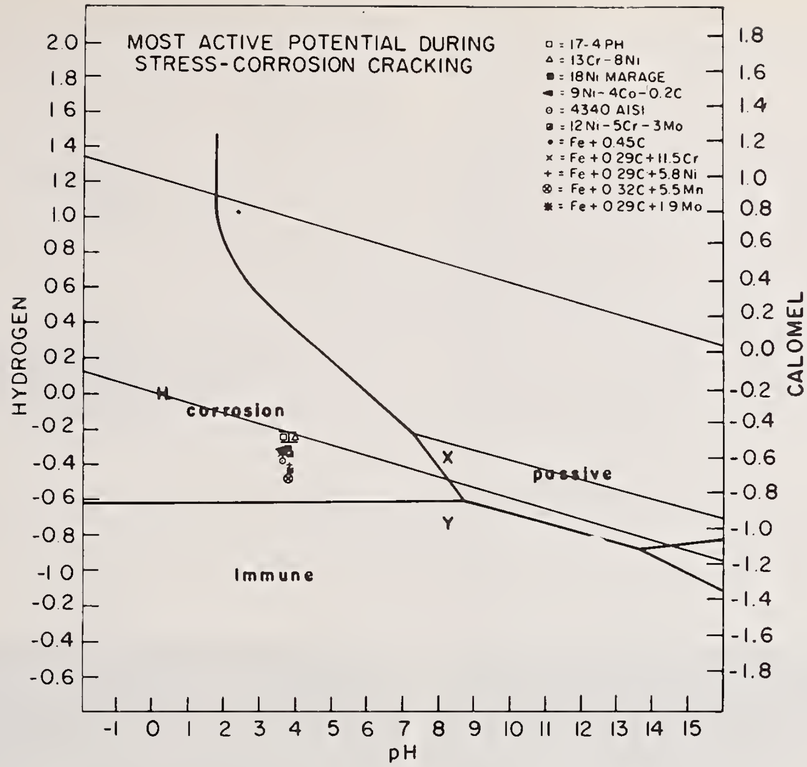


FIGURE 7. *Potential-pH diagram showing the potential-pH values in the crack tip environments for a number of high strength steels.*
From Brown in reference [6], page 197.

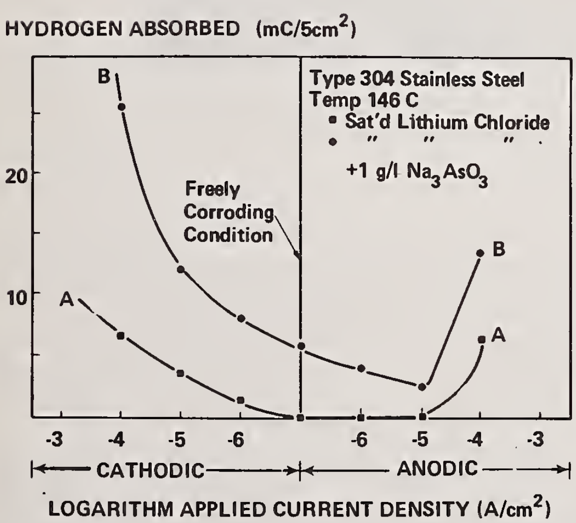


FIGURE 8. *The amount of hydrogen absorbed in 304 stainless steel as a function of applied current.*
Na_3AsO_3 (curve B) enables greater amounts of hydrogen to enter the metal. In spite of the fact that more hydrogen enters the metal in the cathodic region no SCC failures were observed for currents greater than a cathodic current of around 10^{-7} A/cm^2. From Wilde and Kim [40].

netics [51–56]. These techniques produce a bare surface either by straining an alloy wire or by scratching or abrading a surface. In most of these techniques, the specimens are either strained or scratched while maintaining their potentials at a fixed value and measuring the current transient that takes place during the dissolution-repassivation event that occurs subsequent to exposure of bare metal. The rate of current decay is related to the rate of repassivation (figure 9). In these mainly electrochemical techniques, one cannot separate out the current involved in dissolution from that involved in passive film formation. A technique recently developed by Ambrose and Kruger [56], however, does attempt to measure separately the film regrowth transient by optical means (ellipsometry) and to compare it to the overall dissolution-repassivation current. With this technique, called tribo-ellipsometry one removes the protective film on a metal in an environment by means of a polishing wheel and then records both the film growth kinetics and the current as shown in figure 10. In figure 10 we can see that the rate of film growth on a titanium alloy is considerably slower in the susceptible chloride environment than it is in the non-susceptible nitrate solution. With tribo-ellipsometry one can determine the ratio of the total current to that involved in

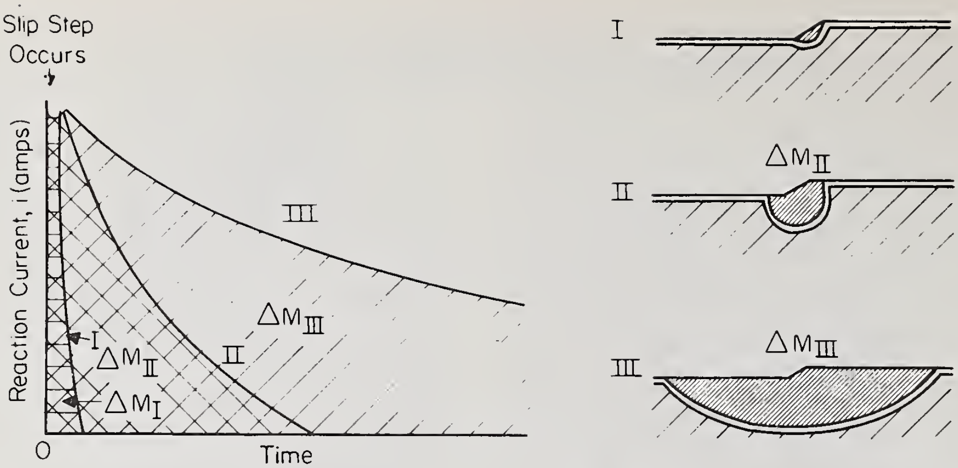


FIGURE 9. *A schematic representation showing the different current transients expected from straining electrode experiments and relating these transients to the amount of material dissolved after the rupture of a protective film by an emergent slip step.*

I, rapid repassivation with little metal dissolution. II, intermediate repassivation rate producing restricted dissolution and favoring SCC. III, slow repassivation producing extensive metal dissolution. From Staehle [10].

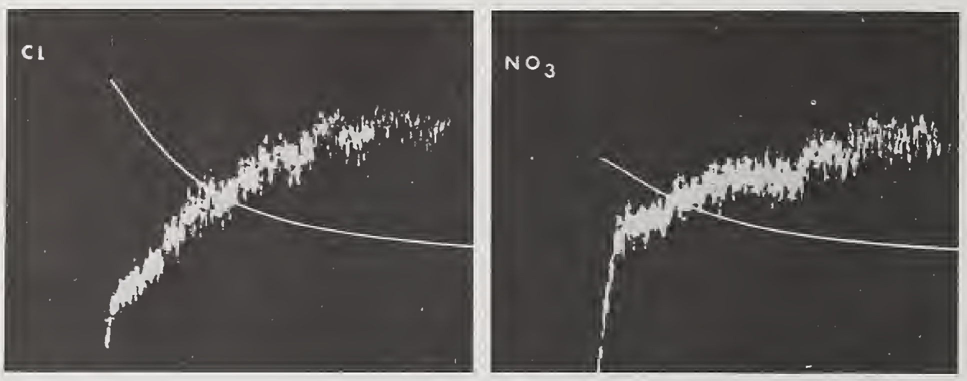


FIGURE 10. *A comparison of oscilloscope repassivation transients for* Ti–8Al–1Mo–1V *alloy in* 1N Nacl *and* $NaNO_3$ *solutions at +706 mV (SHE) (ellipsometric film growth—jagged trace; current—solid trace).*

From Ambrose and Kruger [57].

repassivation only. This ratio R_p^* when the exposed surface area is know is given by the expression

$$R_p^* = 1 + \frac{Q_d}{Q_r} \quad (3)$$

where Q_d = charge involved in dissolution and Q_r = charge involved in repassivation.

R_p^* is a measure of the effectiveness of the repassivation process. Work on the cracking of low carbon steel [17] has shown that SCC does not occur when the effectiveness of repassivation is neither too high nor too low. Thus, for low carbon steel at the time required to form a complete film R_p^* is 2.8 in 1 N $NaNO_2$ (25 °C) where it is nonsusceptible; R_p^* is 26 in 4 N $NaNO_3$ (90 °C) where it does undergo SCC, and R_p^* is >75 in 1 N NaCl (25 °C) where it undergoes widespread pitting attack but not SCC. Thus, a delicate balance between repassivation and dissolution must be maintained to lead to SCC. The determination of this balance is dependent on the environment, the alloy, and finally the metallurgy and mechanics of the system that control the rate at which bare metal sites (generally slip steps) are produced as well as their density and size.

This section, devoted to chemistry, has described techniques for evaluating the role of the environmental side of the equation; the next two sections are concerned with the other side of the equation.

Techniques Dealing with the Metallurgy of SCC—In recent years two experimental aproaches have become increasingly important in providing insights into the role of metallurgy in SCC. They are constant strain rate tests and a variety of applications of electron microscopy.

(*a*) Constant Strain Rate Tests—This test, which produces experimental data as depicted in figure 11, can be applied both to mechanistic studies as well as used for a routine testing. The technique has been applied to studies of mild steels [59], Ti alloys [60] and Mg alloys [23]. The mechanistic importance of the technique is that it ties the metallurgically controlled rate of bare metal exposure, resulting from film rupture by slip step emergence, to the chemical process of repassivation of the exposed metal. This can be seen by examining the significance of the three regions shown in figure 11.

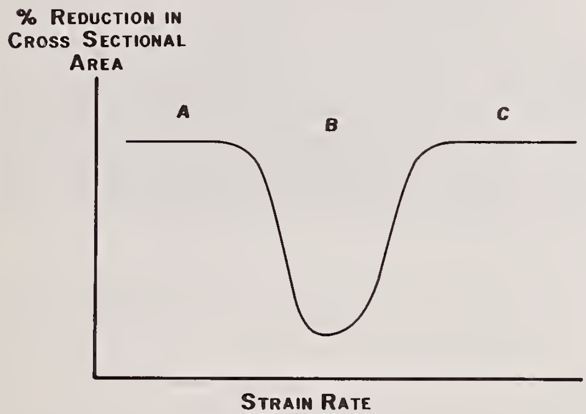


FIGURE 11. *A schematic representation of the results obtained in constant strain-rate tests.*
Region A—repassivation is rapid enough and strain rate slow enough to prevent damaging interaction with the environment. Region B—strain rate high enough so that repassivation does not prevent environmental access. Region C—strain rate is so rapid that ductible pulling apart takes place.

For Region A where strain rates are too low to lead to SCC, the rate of repassivation is high enough to prevent any significant and damaging interaction between the environment and the bare metal exposed during straining. At the opposite extreme, Region C, the strain rate is so rapid that the interaction of the exposed metal with the environment is of no consequence because the rate that the metal fails by ductile fracture exceeds the rate by which the environment can affect fracture via any of the possible SCC mechanisms (dissolution, hydrogen embrittlement, or adsorption of damaging species). It is only in the intermediate range of strain rates, Region B, where the rate of production of bare metal sites has a value sufficiently high so that the rate of repassivation does not prevent environmental aecess. However, the strain rate is sufficiently low so that environmental interaction and not ductile pulling apart produces fracture. Under these conditions SCC takes place. If SCC is not possible for the system studied, the dip shown in Region B will not occur.

Does this test, whose development and increasing use was stimulated for mechanistic reasons and was mainly guided by the implications of film rupture and repassivation, provide a valid routine material-evironmental evaluation test? Parkins [58] has pointed out that there are certain features of the constant strain rate technique that make it a useful routine laboratory test. These features are as follows:

(1) It is a relatively severe test so that it will promote laboratory SCC failures where other tests on smooth specimens will not unless inordinately long testing times are used;

(2) It always produces fracture by SCC or some other mechanism and therefore is a positive test;

(3) The time of testing is relatively short.

These three somewhat interrelated attributes, severity, positiveness, and rapidity, are especially valuable in comparative testing where one wishes to intercompare materials and environments.

The promise that the constant strain rate approach offers for laboratory testing is somewhat tempered by the lack of good ways to quantify the results of the tests. For example, how does one relate the strain rate in Region B where SCC susceptibility is the greatest to crack growth rate? Moreover, it is not known what relation, if any, this strain rate bears to the strain rate that exists at the tip of a growing crack. Parkins [23] has made some valuable qualitative connections between this value and the threshold stress for SCC or the threshold value of the stress intensity factor (described in the next section on mechanics) but the quantitative relationships require new theoretical and experimental advances. Despite these problems, the constant-strain rate tests have provided evidence of SCC in systems where other tests have been difficult or not possible to apply [61–63].

(*b*) Electron-Microscopy of SCC—With the advent of sueh powerful tools as the scanning electron microscope (SEM), and high voltage electron microscopy, have come new opportunities for the diagnosis of SCC failures as well as for the prediction of the likelihood of failure. The predictive ability provided by these tools is based more on the mechanistic insights they provide (by allowing the examination of the crack tip itself at nearly atomic resolution) than on any testing procedure employing them. For example, using mechanisms that rely on repassivation as a step in the SCC process, Scully [50] has proposed that the width of the slip steps produced when a metal is stressed plays a role in determining SCC susceptibility. He suggests that the wider the step, the more time is allowed for the migration (under a potential gradient) of a damaging ion (usually chloride) to the point on the metal surface where the slip step is emerging. Thus, a greater concentration of this ion will build up at the emergence point and thereby retard repassivation or enhance dissolution at this point. For fine slip steps, the time for the build-up of concentration of damaging ions via migration along the emerging step is less and repassivation is more effective. Therefore, Scully points out that the metallurgical substructure that determines the width of the step becomes crucial. By means of transmission electron microscopy one can determine the metallurgical substructure. For example, it has been found that some alloys that suffer transgranular SCC have lattices of co-planar arrays of dislocations (figure 12a) which produee wide slip steps (e.g., α brasses [64]). On the other hand, alloys where cross slip is easy (figure 12b) are less susceptible to SCC because deformation results from fine slip and the process described by Scully would be less likely to lead to cracking. The width of the slip steps produced is probably not critical for all combinations of environments and alloys and, indeed, has been strongly questioned [65]. Nevertheless, it is mentioned to illustrate

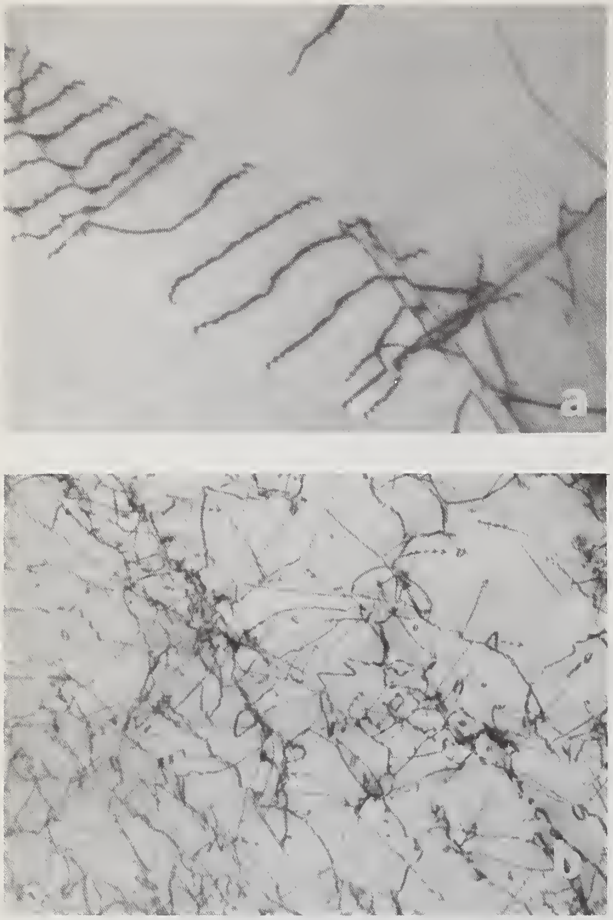


FIGURE 12. *Transmission electron micrographs showing:*
(a) *Coplanar array of dislocations leading to wide slipsteps.*
(b) *Tangled array of dislocations leading to fine slipsteps.*
(Photographs courtesy of L. K. Ives.)

how a transmission electron microscopic study of metallurgical substructure can be used as a failure predictive tool.

Metallurgical substructure plays an important role in the formation of corrosion tunnels which have been found by some workers using electron microscopy to be associated with SCC [66,67]. Scamans and Swann [68] describe a technique whereby they are able to look at U-bend specimens in a high voltage electron microscope and thereby observe the effects of stress and environmental interactions. High voltage transmission electron microscopy (500–1000 kV) allows the use of specimens approximately six times thicker than those used on conventional electron microscopes (100 kV) and thus come closer to studying phenomena in bulk material. With double tilting capabilities they could analyze the three dimensional morphology of the crack tip. By this means they could relate attack morphology of austenitic stainless steels to slip traces and found that more susceptible alloy compositions showed a greater tendency to promote the formation of corrosion tunnels. They found tunnel diameters increased with decreasing susceptibility. For the purposes of this paper, a detailed discussion of the role of attack morphology and susceptibility is not necessary. What is important is to stress that by means of transmission microscopy one can determine morphological or metallurgical features (tunnel formation, dislocation array, etc.) which can serve as potential failure predictive aids when their significance has been more firmly established.

Unlike the just mentioned examples of applications of transmission microscopy, the use of SEM for the study of fracture surfaces (fractography) is more useful for diagnostic, rather than predictive, purposes. As such, it it extensively used for failure analysis which serves as a basis for suggesting measures to prevent future failures. A comprehensive discussion of the use of fractography in the study of SCC is given by Scully [69]. Figure 13 shows a fracture surface that is indicative of a failure that was brought by the hydrogen embrittlement mechanism of SCC. The flat crystallographic damage planes are typical of the brittle mode of fracture that would result from this kind of failure. In contrast, figure 14 shows the dimpled fracture that results from ductile failure that could not be attributed to hydrogen embrittlement. The introduction of the SEM in recent years into many laboratories has given an enormous impetus to fractographic studies. This is so because, as pointed out by Scully [69], the SEM has the following valuable advantages:

(1) The microscope has a wide range of magnifications. This is important in SCC failure analysis because it can quantitatively determine dimple/cleavage or dimple/intergranular ratios and thereby give information about the general pattern of crack propagation. Also, by working at different magnifications it can look at the different varieties of fracture features that can result from different grain orientations.

(2) The microscope has an extremely large depth of field. Because in SCC where corrosion, corrison products, and rough surfaces are common, this feature is of great, actually essential, utility.

(3) No replication is necessary so that one need not be concerned about an alteration of the fracture surface (removal of corrosion products, for example) by the replication process.

(4) It can distinguish between the metallic and nonmetallic materials (corrosion products, inclusions) making up the fracture surface. This can be a source of misinterpretation when replication techniques are used.

Therefore, the advent of SEM fractography has provided a powerful tool for the failure diagnosis which is an essential step in the prevention of SCC failure.

Techniques Dealing with the Mechanics of SCC—Up to this point most of the new techniques described have had their origins in the different mechanisms of SCC and are aimed at providing predictive abilities. However, while some offer hope that such opportunities to predict failure will materialize, the engineer's best present hope lies either in utilizing past experience or in utilizing laboratory tests that adequately model service conditions. Presently, the tests developed using the con-

FIGURE 13. *A scanning electron micrograph of the fracture surface of a high strength steel showing transgranular quasi-cleavage typical of hydrogen embrittlement.*
Arrows show dimpled rupture where some ductility is occurring at a few local sites. (Photograph courtesy of C. G. Interrante and G. E. Hicho.)

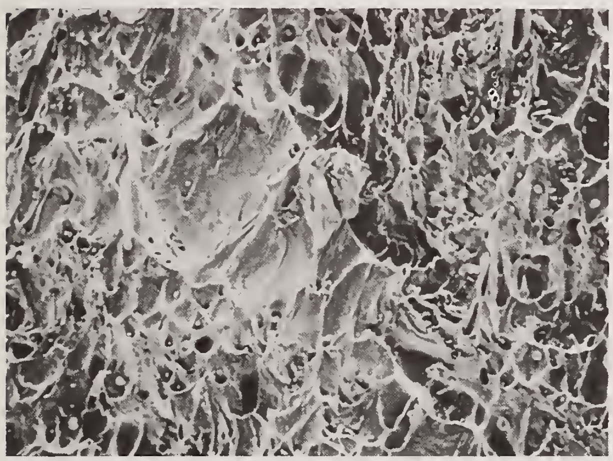


FIGURE 14. *A scanning electron fractograph showing typical ductile fracture of a high strength steel which indicates no hydrogen embrittlement.*
(Photograph courtesy of C. G. Interrante and G. E. Hicho.)

cepts of linear elastic fracture mechanisms come closest to answering this later need. A good review of the theory and application of these techniques to SCC is given in a book edited by Brown [70]. Previously most laboratory tests for evaluating the susceptibility of an alloy to SCC in a given environment used smooth specimens subjected to a stress and measured the time required to observe a failure of the stressed specimen. One description of the kinds of specimens and experimental apparatus used can be found in the book by Logan [71]. The two major criticisms of the test methods that do not use the concepts of linear elastic fracture mechanics are that the geometry of the stress field around a crack is poorly defined or unknown and that time to failure involves the time for crack initiation as well as crack propagation. Brown and Beachem [72] showed, for example, that titanium alloys, which appeared to be completely immune to SCC because they did not pit readily and thereby provide initiation sites, were highly susceptible to SCC once initiation sites were present. Therefore, since initiation may be highly variable depending on whether or not flaws exist in the specimen studied, the time to failure can be an unreliable and hence unreproducible measure of susceptibility.

The testing techniques for characterizing SCC susceptibility by the use of fracture mechanics seek to overcome these shortcomings of time-to-failure techniques by providing a pre-existing flaw or crack and by clearly defining the stress field around a crack through the use of specimens whose geometry allows the characterization of a parameter K [3] the stress intensity factor. This parameter depends on the loading and configuration of the specimen, including crack size, and determines the magnitude of the local stresses. If a certain value of K is found to cause the growth of a stress corrosion crack, any combination of structure and crack geometry which duplicates this level of K will produce the same crack growth rate for the same material and environment. Thus, using K allows one to relate SCC tests carried out for one crack and body geometry to another. Brown and Srawley [73] have made calculations that allow the determination of K for a number of practical test specimens. The pre-existing flaw used to overcome the problem of crack initiation is for many specimens a fatigue crack intentionally introduced for this purpose. The argument for this approach is that most structures have flaws in them so that crack initiation is not of great practical importance.

In SCC tests using specimens designed for the application of fracture mechanics concepts, results similar to those shown schematically in figure 15 will be obtained. In these tests the rate of crack growth is measured as a function of K and, as is pictured, there can be regions (I, II, III) where the functional dependence of crack growth rate on K differs markedly. Of more importance for SCC testing is the existence of a threshold value of K, the K_{ISCC} [4] shown in figure 15, below

FIGURE 15. *Schematic diagram of the data obtained from fracture mechanics techniques applied to stress corrosion cracking.* From Smith and Piper [74].

[3] For a crack of length $2a$ in an infinite plate where the crack is normal to the tensile stress field of value σ, $K = \sigma\sqrt{\pi a}$.

[4] The I in the subscript of K_{ISCC} represents Region I and could be left out because SCC occurs only in Region I [4].

which the crack growth rate is zero or insignificant for practical purposes. Determining K_{ISCC} is an important goal in SCC testing because one can relate it to a critical flaw depth, a_{cr}, and the yield-point stress that may exist in a structure whose SCC behavior is being determined. Brown [73] has shown that

$$a_{cr} = 0.2\left(\frac{K_{ISCC}}{\sigma_Y}\right)^2, \qquad (4)$$

where σ_Y is the yield strength. He proposes that a_{cr} be used as a figure of merit which includes both K_{ISCC} and any yield strength stress whether applied or residual. Figure 16 shows how one can use a_{cr} and σ_Y to evaluate SCC characteristics of materials. Thus, a material whose value of K_{ISCC} in a given environment gives a value of X as shown in figure 16 would suffer SCC (crack propagation) if it had a flaw of 2.5 mm, but not if it had a flaw of 0.25 mm.

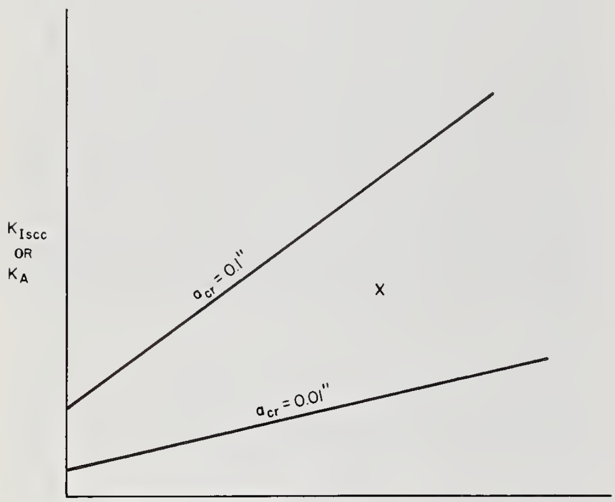


Figure 16. *A plot of Equation [4] using 0.1" and 0. 01" values for* a_{cr} *assuming long surface flaws and yield strength stresses operating.*

A material with a KISCC value of X would not undergo SCC in the environment where measurements were made if surface flaws were no deeper than 0.01". If, however, the flaws were as deep as 0.1", SCC would occur. From Brown [73].

There have been a number of criticisms of the general applicability of the precracked specimens to SCC testing. Parkins [58] has spelled out a number of these shortcomings. Among these are the following:

(1) In order to achieve conditions of plane strain (the conditions upon which the relationship between K and specimen geometries are based) for ductile materials, test specimens of impractical size must be used; [5]

(2) There may exist problems when precracked specimens are usually cracked transgranularly although the SCC process for the alloy is intergranular;

(3) ·The significance of the initiation stage may be to produce a proper environment for cracking rather than just provide a notch to promote crack propagation;

(4) The frequent observation of crack branching may modify the significance of the values of K_{ISCC} determined in fracture mechanics testing.

These shortcomings are not serious for the many cases where they do not apply and where fracture mechanics techniques have been successful, but it is obvious that one cannot rely exclusively on one testing approach. Moreover, a recent paper [23] describes ways to interrelate results obtained from precracked specimens using fracture mechanics concepts and those obtained from smooth specimens used in constant strain rate tests. Thus, a combination of the two approaches may be in the offing.

4. New Opportunities for Failure Prediction and Diagnosis

This brief survey of new approaches towards developing failure predictive tools to cope with SCC have identified new opportunities provided by the three disciplines that govern the phenomenon of SCC. From chemistry comes the recognition of considering the changes in environment that occur in the crack and the important effect of environment on repassivation kinetics. From metallurgy comes the recognition of the importance of metallurgical substructure and the role that the rate of bare metal production via slip step emergence plays. From mechanics comes the recognition of properly characterizing the geometry of the stress field around a crack. When these tools have been refined further, both through experiment and theory, the eventual goal of being able to predict susceptibility to SCC in going from one environment to another (chemistry), from one metallurgical structure to another (metallurgy), and from one geometry to another (mechanics) will be realized. This paper has described some of the beginning steps toward such a goal.

I am most grateful to the Office of Naval Research which supported this work under contract NAONR 18–89 NRO 36–082.

5. References

[1] Staehle, R. W., Proc. Conf. Fundamental Aspects of Stress Corrosion Cracking, R. W. Staehle, A. J. Forty, and D. Van Rooyan, Eds., Nat. Assoc. Corros. Eng., Houston (1969), p. 3.

[2] Brown, B. F., Naval Research Laboratory Report 7130, June 1970.

[3] Hodge, J. C., and Miller, J. L., Trans. A.S.M. **28**, 25 (1940).

[4] Brown, B. F., Stress-Corrosion Cracking in High Strength Steels and in Titanium and Aluminum Alloys, B. F. Brown, Ed., Naval Research Laboratory. (1972), p. 2.

[5] Brown in a private communication has pointed out that because all cracks are growing prior to general yielding, the stress may be characterized by elastic mechanics. Therefore, the specimen size is determined by the ratio of K_{ISCC}/σ_{Y}. Thus, one could have small specimens for relatively soft materials.

[5] Staehle, R. W., Forty, A. J., and Van Rooyan, D., Eds., Proc. Conf. Fundamental Aspects of Stress Corrosion Cracking, Nat. Assoc. Corros. Eng., Houston (1969).
[6] Scully, J. C., Ed., The Theory of Stress Corrosion Cracking in Alloys, N.A.T.O., Brussels (1971).
[7] Proc. Intl. Conf. on Stress Corr. Cracking and Hydrogen Embrittlement of Iron Base Alloys, St. Etienne, 1973, to be published by Nat. Assoc. Corros. Eng., Houston.
[8] Brown, B. F. in Stress Corrosion Cracking of Metals—A State of the Art, H. L. Craig, Jr., Ed., STP518, A.S.T.M., Philadelphia, (1972).
[9] Parkins, R. N., in ref. [6] p. 167.
[10] Staehle, R. W., in ref. [6] p. 223.
[11] Uhlig, H. H., and Cook, Jr., E. J. Electrochem. Soc. **116**, 173 (1969).
[12] Pugh, E. N., Montague, W. G. and Westwood, A. R. C., Corrosion Sci., **6**, 345 (1966).
[13] Escalante, E., and Kruger, J., J. Electrochem. Soc. **118** 1062 (1971).
[14] Uhlig, H. H., in ref. [5], p. 86.
[15] Uhlig, H. H., in ref. [7].
[16] Preece, C. M., in ref. [7].
[17] Ambrose, J. R. and Kruger, J., Proc. 5th Intl. Cong. on Metallic Corrosion, Tokyo, 1972 in press.
[18] Logan, H. L., J. Res. Nat. Bur. Stand. (U.S.), **48**, 99–105 (1952), RP2291.
[19] Champion, F. A., Intl. Symp. on Stress in Metals and Alloys, Inst. of Metals, London, (1948), p. 468.
[20] Scully, J. C., Corrosion Sci., **7**, 197 (1967).
[21] Parkins, R. N., in ref. [7].
[22] Hoar, T. P., and Hines, J. G., J. Iron Steel Inst. **184**, 166 (1956).
[23] Wearmouth, W. R., Dean, G. P., and Parkins, R. N., Corrosion, **29**, 251 (1973).
[24] Vermilyea, D. A., in ref. [7].
[25] McEvily, Jr., A. J., and Bond, A. P., J. Electrochem, Soc. **112**, 638 (1965).
[26] Pugh, E. N., in ref. [6] p. 418.
[27] Lawless, K. R., Energetics in Metallurgical Phenomena, **1**, M. W. Mueller, Ed. (Gordon and Breach, New York, 1965), p. 345.
[28] Kruger, J., Ambrose, J. R., and E. Escalante, Nat. Bur. of Stand. (U.S.), Report 10594 (May 1971).
[29] Nielsen, N. A., Corrosion, **27**, 173 (1971).
[30] Oriani, R. A., Ber Bunsen Gesell. f. Phys. Chemie **76**, 848 (1972).
[31] Troiano, A. R., Trans. A.S.M., **62**, 54 (1960).
[32] Birley, S. S., and Tromans, D., Corrosion, **27**, 63 (1971).
[33] Whiteman, M. B., and Troiano, A. R., Corrosion, **21**, 53 (1965).
[34] Latanision, R. M., Staehle, R. W., Scripta Met. **2**, 667 (1968).
[35] Hancock, G. G., and Johnson, H. H., Trans. Met. Soc., A.I.M.E. **236**, 513 (1966).
[36] Shively, J. H., Hehemann, R. F., and Troiano, A. R., Corrosion **22**, 253 (1966).
[37] Holzworth, M. L., and Louthan, Jr., M. R., Corrosion **24**, 110 (1968).
[38] Vaughn, D. A., Phalen, D. I., Peterson, C. L., and Boyd, W. K., Corrosion **19**, 315 (1963).
[39] Nechay Ye. P., and Popov, K. V., Fiz. Metal Metalloved. **19**, 612 (1965).
[40] Wilde, B. E., and Kim, C. D., Corrosion, **28**, 350 (1972).
[41] Bernstein, M., and Pickering, H. W., Carnegie-Mellon University Report 036–099–2, May 1974.
[42] Brown, B. F., Fujii, C. T., and Dahlberg, E. P., J. Electrochem. Soc. **116**, 218 (1969).
[43] Sandoz, G., Fujii, C. T. and Brown, B. F., Corrosion Sci. **10**, 839 (1970).
[44] Smith, J. A., Peterson, M. H., and Brown, B. F., Corrosion **26**, 539 (1970).
[45] Pourbaix, M., in ref. [6] p. 17.
[46] Leidheiser, Jr., H., and Kissinger, R., Corrosion **28**, 218 (1972).
[47] Bogar, F. D., and Fujii, C. T., Naval Research Laboratory Report 7690, March 1974.
[48] Devanathan, M. A. V., and Stachurski, Z., Proc. Roy. Soc. **270A**, 91 (1962).
[49] Bockris, J. O'M., in ref. [7].
[50] Scully, J. C., Corrosion Sci., **8**, 513 (1968).
[51] Hoar, T. P., and West, J. M., Proc. Roy. Soc. **268A**, 304 (1962).
[52] Leidheiser, Jr. H. and Kellerman, E., Corrosion **26**, 99 (1970).
[53] Murata, T., and Staehle, R. W., Proc. 5th Intl. Cong. on Metallic Corrosion, Tokoyo, 1972, to be published by Natl. Ass. Corr. Eng., Houston.
[54] Tomashov, N. D., and Vershinina, L. P., Electrochimica Acta **15**, 501 (1970).
[55] Pessal, N., and Liu, C., Electrochimica Acta **16**, 1987 (1971).
[56] Ambrose, J. R., and Kruger, J., Corrosion **28**, 30 (1972).
[57] Ambrose, J. R., and Kruger, J., J. Electrochem. Soc. **121**, 599 (1974).
[58] Parkins, R. N., in ref. [6] p. 449.
[59] Humphries, M. J., and Parkins, R. N., Corrosion Sci. **7**, 747 (1967).
[60] Powell, D. T., and Scully, J. C., Corrosion **24**, 151 (1968).
[61] Parkins, R. N., Mazza, F., Royuela, J. J., and Scully, J. C., Brit. Corr. J. **7**, 154 (1972).
[62] Scully, J. C., and Powell, D. T., Corrosion Sci. **10**, 371 (1970).
[63] Deegan, D. C., and Wilde, B. E., Corrosion, **29**, 310 (1973).
[64] Tromans, D., and Nutting, J., Fracture, J. Gilmon, Ed. (Interscience, New York, 1963), p. 637.
[65] Nielsen, N. A. in ref. [7].
[66] Nielsen, N. A., Corrosion, **27**, 5 (1971).
[67] Swann, P. R. in ref. [6] p. 113.
[68] Scamans, G. M., and Swann, P. R. in ref. [7].
[69] Scully, J. C., in ref. (6) p. 127.
[70] Brown, B. F., Ed., Stress-Corrosion Cracking in High Strength Steels and in Titanium and Aluminum Alloys, Naval Research Laboratory. U.S. Government Printing Office, Washington, 1972.
[71] Logan, H. L., The Stress Corrosion of Metals (John Wiley and Sons, New York, 1956), pp. 273–285.
[72] Brown, B. F., and Beachem, C. D., Corrosion Sci. **5**, 745 (1965).
[73] Brown, Jr., W. F., and Srawley, J. E., Plane Strain Crack Toughness Testing of High Strength Metallic Materials, A.S.T.M., STP410 (1966).
[74] Smith, H. R., and Piper, D. E., Stress-Corrosion Cracking in High Strength Steels and in Titanium and Aluminum Alloys. B. F. Brown, Ed., Naval Research Laboratory. U.S. Gov't Printing Office, Washington, 1972, p. 17.

6. Discussion

W. J. Baxter, General Motors Corporation: You indicated a distinction between fine slip and coarse slip, and that one was passivated more quickly than the other. I don't understand the reason for this.

J. Kruger: The rate of passivation is the same for both of them. The problem is that for fine slip, since there is less surface to deal with, the whole slip step can be repassivated without forming a notch at the last portion that's repassivated. For coarse slip, the whole thing might repassivate, and a little notch forms at the end because there is more surface.

W. J. Baxter: I don't understand why repassivation is like a creeping fungus. Why doesn't the whole surface just uniformly repassivate?

J. Kruger: Because the whole surface is not exposed at the same time. The first part that comes out is exposed first.

W. J. Baxter: You're saying that you get a little fine slip and then it's repassivated and then the next one forms?

J. Kruger: That's right, it's a dynamic process. But for coarse slip there is not enough repassivation time to repassivate the whole thing, so there is more corrosion occurring.

D. C. Drucker, University of Illinois: Have you tried any calculations which would combine the mechanics approach with the chemistry approach; that is, to see whether, for example, a particular species sitting at the root of a crack interacts with the stress field in the mechanics sense, not quite the fracture mechanics sense, but the more detailed mechanics sense? Does it make any sense to do that? Can you then show that you should expect bond separation or bond breakage?

J. Kruger: We haven't tried that, but the kind of thing you are talking about is related to the stress-sorption picture. I believe for metals in aqueous environments that the reactions that are taking place, for example the formation of a passive film or the dissolution of metal, are so rapid compared to the rate of adsorption of a species that it is not important. For non-metals, for example glasses, it is indeed important and the kind of calculations you suggest would be worthwhile.

A. J. McEvily, University of Connecticut: You indicated a repassivation trend for the precipitate free zone of the Al–Zn–Mg alloy and for aluminum alloy 7075. To which of the stages of the stress corrosion process does that trend apply?

J. Kruger: You mean the 3 regions in the V–K type curves?

A. J. McEvily: Yes.

J. Kruger: Region 1 is really the only one that most people are concerned with in stress corrosion.

A. J. McEvily: So that's more like a crack initiation stage?

J. Kruger: It certainly is involved in crack initiation, but I think it can also be involved in crack propagation. Whether the crack will continue to grow, or whether metal dissolution or active path cracking is involved depends on the relative rate of repassivation and the relative rate of slip step production versus the rate of corrosion.

A. J. Babecki, NASA, Goddard Space Flight Center: One solution for preventing stress corrosion cracking is to put the surface in compression, usually through shot peening or glass bead peening. Peening of some metals such as magnesium increases the corrosivity of the metal. If the corrosion reaction of the surface is increased by the peening procedure, does that also tend to increase the stress corrosion cracking phenomenon?

J. Kruger: Not at all. Whereas the shot peening may increase the rate of corrosion, it would decrease the susceptibility to stress corrosion.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Elastohydrodynamics in Concentrated Contacts

Herbert S. Cheng

Northwestern University
Evanston, Illinois 60201

Typical geometrical features of elastohydrodynamic (EHD) contacts and major operating, lubricant, and material parameters governing lubrication in these contacts are described. Operating regimes for various machine components, such as gears, rolling element bearings, cams, and compliant bearings, are identified. A state-of-the-art survey is given to the predictability of major EHD characteristics based on existing analytical and experimental EHD research. The significance of film thickness, pressure fluctuations, friction, and temperature in affecting failures of concentrated contacts by fatigue, scuffing, and wear is discussed. Finally, suggestions are made to incorporate the EHD effects in predicting fatigue life, scuffing load, and wear rate.

Key words: Contact lubrication; elastohydrodynamic contacts; fatigue; lubricant film thickness; scuffing; wear.

1. Introduction

In concentrated contacts, load is generally supported over a tiny contacting area known as the Hertzian contacting area, named after Heinrich Hertz who first developed the elasticity theory to determine the size and shape of these contacts. Even under a moderate load, the contacting pressure developed within a Hertzian area can become enormous, and often can exceed the tensile yield strength of the contacting solids. Because of this severe condition, it was believed earlier that a continuous lubricant film cannot exist, and load is strictly supported by metallic contact through the asperities. Lubrication is provided mainly by a thin surface film of one or several molecular layers recognized as boundary lubrication.

The concept that concentrated contacts operate strictly in the regime of boundary lubrication was unchallenged until 1949 when Grubin and Vinogradova [1] [1] theorized that a fluid film of the order of many times the surface roughness can be developed hydrodynamically. This thick film is mainly attributed to a rapid increase of viscosity with pressure under high pressures for mineral oils, and to enlargement of the loading area caused by elastic deformation.

Grubin's work marked the beginning of an important field known as elastohydrodynamic lubrication, devoted to understanding of the lubrication behavior where elastic surface deformation plays a dominant role in governing the process. During the past two decades, rapid progress has been made both in the theoretical and experimental understanding of major variables in EHD lubrication, including detailed film, pressure, temperature, and traction distribution in lubricated Hertzian contacts. Comprehensive reviews have been made by Dowson [2] and more recently by McGrew, et al. [3].

The objectives of this presentation are to provide a general background of EHD lubrication in concentrated contacts such as rolling element bearings, gears and cams, to discuss how various failure processes are influenced by EHD, and to suggest methods to incorporate some significant EHD effects in failure predictions in concentrated contacts.

2. EHD Geometry

Considering a general case of a lubricated Hertzian contact between two bodies as shown in figure 1, the geometry at the point of contact is characterized by the principal radii, R_{x1}, R_{y1} for body 1 and R_{x2}, R_{y2} for body 2. In general, the principal planes containing R_{x1} and R_{x2} may not coincide; however, for most EHD contacts, such as roller or ball bearings and gears, the principal radii R_{x1} and R_{x2} do lie in the same plane. These surfaces can be classified as convex, concave, or saddled depending upon whether R_x and R_y are both positive, both negative, or mixed. Thus, contacts between two convex surfaces such as roller-inner race contacts or gear teeth contacts for external spur gears are counterformal. Contacts between one convex and one concave surface such as roller-outer race contacts or gear teeth contacts for internal spur gears are conformal. For contacts between one convex and one saddled surface such as deep-groove ball-inner race contacts, the result is conformal in one direction and counterformal in the perpendicular direction.

In these heavily loaded contacts, the normal load is distributed over an elliptical area approximately conforming to the Hertzian ellipse of a dry contact as shown

[1] Figures in brackets indicate the literature references at the end of this presentation.

at the top of figure 1. In most cases, the direction of rolling is along the minor axis of the ellipse with surface velocities u_1 and u_2 for bodies 1 and 2, respectively. The degree of sliding is measured by a quantity known as the slide to roll ratio denoted by $S = 2\left(\frac{u_1 - u_2}{u_1 + u_2}\right)$ with $S = 0$ corresponding to pure rolling and $S = 2$ to simple sliding.

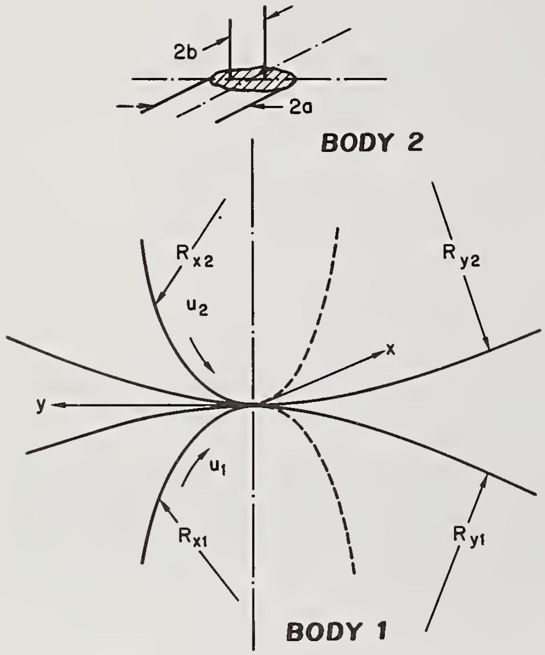


FIGURE 1. *Geometry of an EHD contact.*

3. Governing Parameters

The variables which are well known in fluid film lubrication also play important roles in EHD lubrication. These include speed, load, viscosity, and geometry. In addition to these, the elastic modulus of solids and the pressure-viscosity dependence of lubricants also have a major influence in governing the EHD process. These variables can be grouped together in four major dimensionless parameters listed below:

$$U = \frac{\mu_0 u}{ER} \qquad \text{speed parameter}$$

$$P = p_{HZ}/E \qquad \text{load parameter}$$

$$G = \alpha E \qquad \text{pressure-viscosity parameter}$$

$$S = 2\left(\frac{u_1 - u_2}{u_1 + u_2}\right) \qquad \text{slide to roll ratio.}$$

The above parameters can be readily calculated from properties of the contact material, properties of the lubricant, and operating conditions. Table 1 gives practical ranges of load, rolling speed, sliding speed, lubricant pressure-viscosity coefficient, and elastic modulus of the solid, commonly encountered in EHD contacts such as gears, cams, rolling element bearings, and compliant bearings. In this table, the ultra high-speed applications such as those encountered in aircraft applications have been excluded. Using the limiting values in table 1, the range of EHD load and speed parameters can be readily computed for the above-mentioned EHD contacts. Using load and speed parameters as the coordinates, one can readily show in figure 2 in which regions these contacts operate. It is seen that gears and cams occupy the bottom left corner because of their relatively lower rolling speed and lower contacting stress. Rolling element bearings settle in the middle region, and compliant bearings occupy the upper right corner because of low elastic modulus of solids.

TABLE 1. *Range of operating and material variables in EHD contacts*

	μ	u/R	u_s/u	E	p_{Hz}	α
Gears	10 ↓ 250 c.p.	10 ↓ 250 s^{-1}	0 ↓ 0.35	33×10^6 ps	2.5×10^5 psi	0.0001 ↓ .0002 in^2/lb
Cams	10 ↓ 250 c.p.	10 ↓ 250 s^{-1}	2.0	33×10^6 ps	1.5×10^5 psi	.0001 ↓ .0002 in^2/lb
Rolling element bearings	10 ↓ 100 c.p.	50 ↓ 2000 s^{-1}	0 ↓ 0.02	33×10^6 ps	3.5×10^5 psi	.0001 ↓ .0002 in^2/lb
Compliant bearings	1 ↓ 100 c.p.	10 ↓ 1000 s^{-1}	2.0	300 ↓ 1000 psi	10 ↓ 100 psi	0

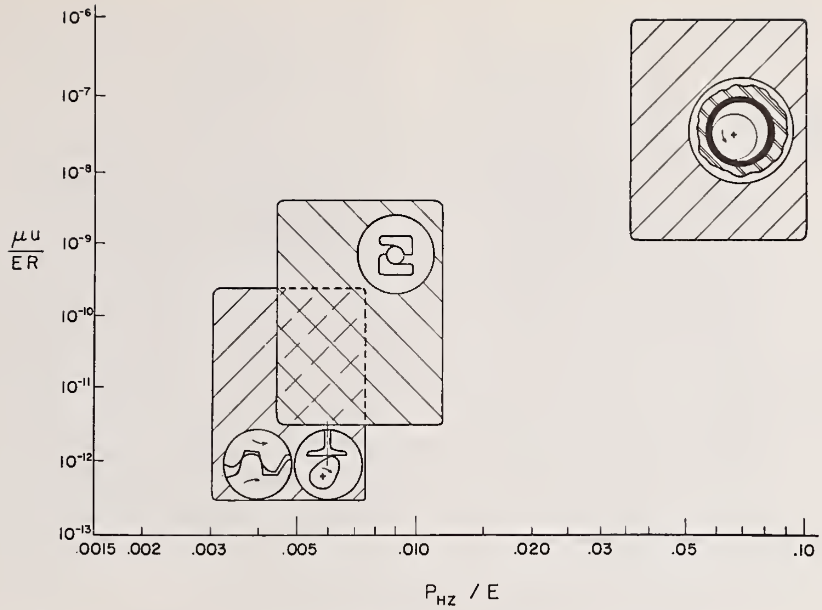


FIGURE 2. *Regions of operation for gears, cams, rolling element bearings, and compliant bearings.*

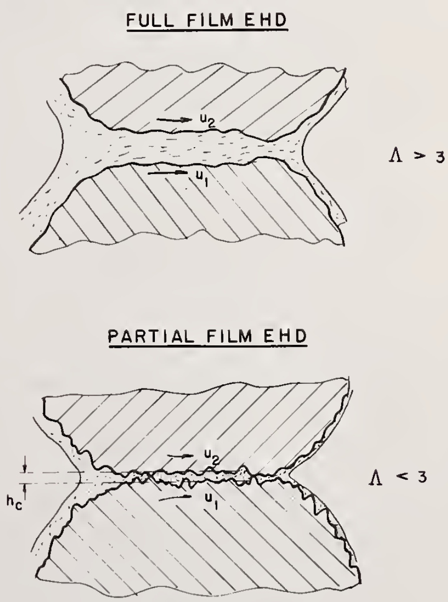
4. Major Characteristics

The behavior of EHD contacts is characterized by the distribution of film thickness, pressure, temperature, and frictional force within the pressurized region. These variables will be discussed separately below.

4.1. Film Thickness

The effectiveness of EHD lubrication is largely measured by how well the asperities are separated by the lubricant film. A fully effective EHD lubrication is generally agreed to be where the center film thickness is at least three times the composite r.m.s. surface roughness of both contacting surfaces. A film thickness inferior to this threshold condition is considered to be a partial EHD lubrication. Figure 3 depicts the essential difference between these two conditions.

Effort in the past has been mostly concentrated on determining the nominal level and detailed distribution of film thickness in line contacts as well as in elliptical contacts. Among the experimental techniques [12, 13, 14] for the study of film thickness, the optical method is most satisfactory in resolving the detailed gap distribution in EHD contacts. Figure 4 shows that the film shape in a heavily loaded elliptical contact consists of a region of approximately uniform film thickness covering the inlet and central parts of the contact. In the exit region, the sudden drop in pressure causes the surface to protrude slightly and produces a constriction around the exit edge of the contact. The film shape is describable by local film thicknesses measured at three strategic locations within the contact: h_c, the center film or the average gap separating the contacting surfaces in the inlet and central regions; h_T, the trailing minimum film thickness along the center strip; and h_s, the side minimum film thickness at the two ends of the elliptical contact.



$$\Lambda = \frac{h_c}{(\sigma_1^2 + \sigma_2^2)^{1/2}} = \frac{h_c}{\sigma}$$

FIGURE 3. *Full film EHD versus partial film EHD.*

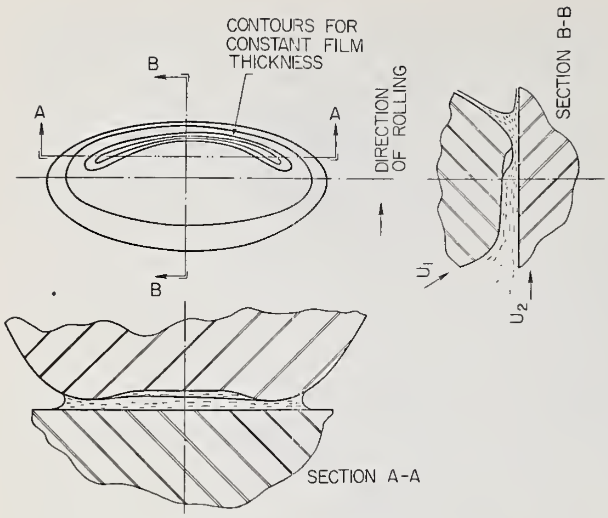


FIGURE 4. *Film shape in elliptical contacts.*

The present ability in predicting the center film thickness h_c in EHD contacts of very small b/a ratios is shown in the left map of figure 5. The operating ranges of load and speed for gears, cams, and rolling element bearings are blocked out as dotted rectangles in this map. The predictability of h_c in each subregion in this map is indicated by a number with 10 for 100 percent accuracy and 0 for a total absence of predictability.

In the partial-film region, the mean central film thickness will be influenced by surface roughness effects, and very little analytical work has been developed so far. Preliminary analysis accomplished by Chow and Cheng [15] based on Christensen's stochastic model [16] for transverse surface roughness showed that the mean value h_c can be altered by surface roughness. Since there are no experimental data available for comparison, no great confidence can be attached to any approximate means of estimating the mean h_c in the partial-film region. Consequently, the entire region is covered with variable degrees of shade indicating marginal to poor predictability.

Figure 6 gives a survey of expected accuracies for the

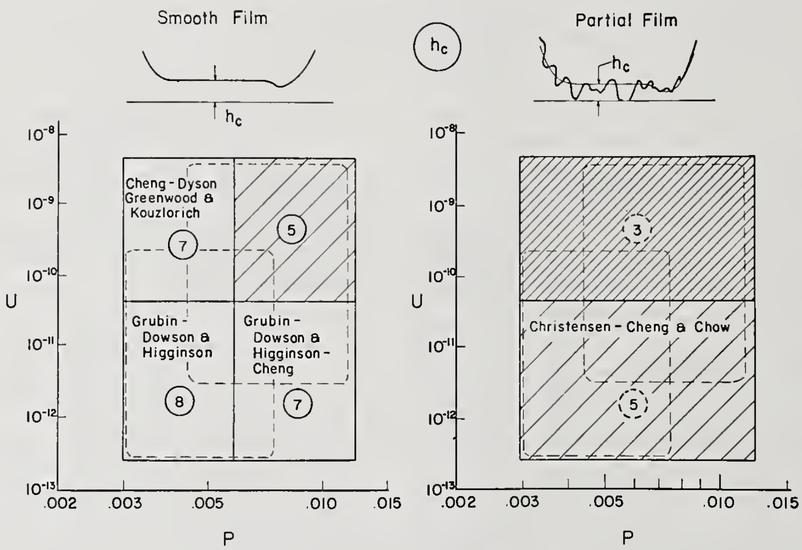


FIGURE 5. *Predictability of* h_c *in EHD line contacts.*

Since the degree of separation is largely governed by h_c, its prediction has been one of the most important tasks in EHD research. This was first achieved by Grubin and Vinogradova [1] for line contacts and confirmed by Dowson and Higginson [4] in their full EHD analysis. For elliptical and circular contacts, similar analyses were also developed by Archard and Cowking [5], Gohar [6], and Cheng [7]. For ultra-high-speed contacts, thermal effects at the inlet region become significant, and center film thicknesses predicted from the isothermal theories can be corrected by thermal correction factors (introduced in [8, 9]). For lubricants with low or moderate pressure-viscosity dependence, theories developed by Herrebrugh [10] and by Cheng [11] may be used.

trailing film thickness h_T in elliptical contacts and the side minimum film thickness h_s in circular contacts. The left map for h_T shows that the only open area, where reasonable accuracy is expected from the Dowson and Higginson theory [4], is the lower left quadrant. For other conditions, one may use the thermal inlet theories for high speed cases, and the recent Jacobson's line contact theory [17] for high loads. However, some of these suggested procedures have not been confirmed by experiment, and, therefore are considered extremely uncertain. Turning to the right portion of figure 6 for the side minimum film h_s in circular contacts, there is very little available analytically. Jacobson [18] recently developed an impressive numerical method for the two dimensional EHD film thickness based on a solidified

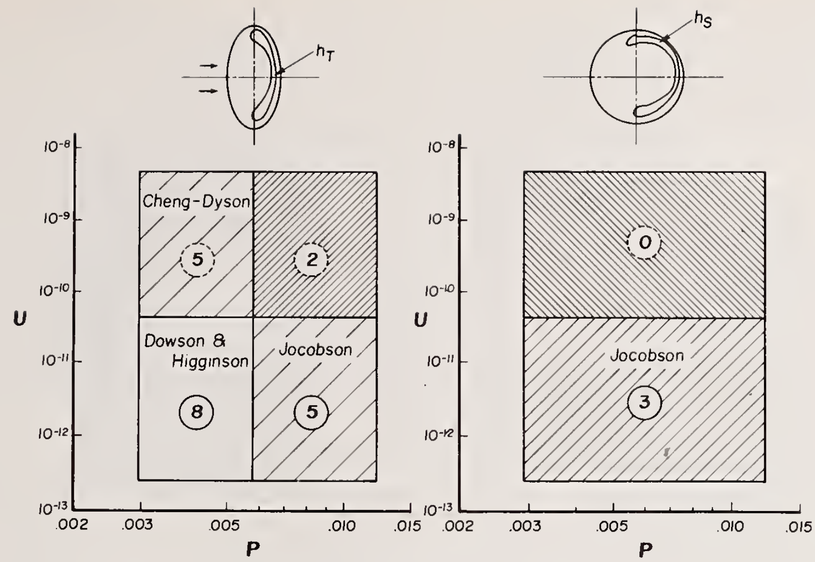


FIGURE 6. *Predictability of* h_T *and* h_s.

lubricant; however, the validity of his solidification model as well as the numerical accuracy is yet to be confirmed. In view of these uncertainties, the entire region is declared to be a dark area.

4.2. Pressure Distribution

The pressure distribution in the contact region of EHD contacts is not altered greatly from the Hertzian distribution. For a line contact or an elliptical contact with very small b/a ratio, the pressure distribution along the center strip of the contact is modified at the inlet and exit edges. In the inlet region, the pressure is built up gradually to the Hertzian distribution, as shown in figure 7. At the exit, EHD analysis for line contacts [4] predicts a sharp secondary pressure rise just prior to the termination of the film. This secondary

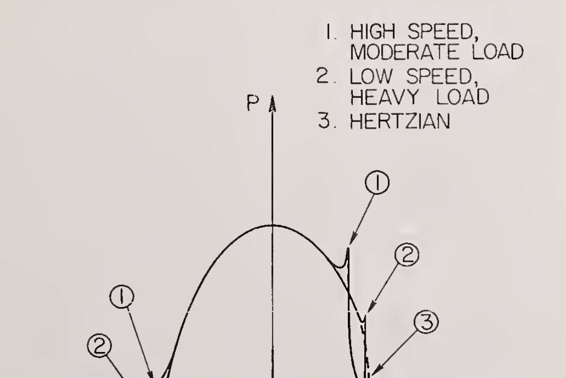


FIGURE 7. *Pressure profiles.*

pressure is caused mainly by the strong pressure-viscosity dependence of the lubricant. However, for very heavily loaded contacts, this pressure spike occurs at the very end of the flattened region, and is not expected to have a strong effect on subsurface fatigue life.

In spite of the extensiveness of EHD theories, the predictability of the magnitude of secondary pressure peaks in EHD contacts is still not fully satisfactory. Present EHD theories can predict an abrupt pressure disturbance and its approximate location, but they cannot predict the degree of its severity. For this reason, only 30 percent confidence is rendered in the entire region in terms of predicting p^* as shown in figure 8. Actually, in terms of relevance to failure, pressure disturbances caused by surface roughness in EHD contacts are far more important compared to the single pressure spike in full-film EHD lubrication. In terms of predicting the severity of these pressure fluctuations, under partial EHD, very little is available. Tallian [19] made an extremely interesting pioneering analysis in assessing the pressure ripples in EHD line contacts. It certainly sheds some light in this otherwise totally dark chamber.

4.3. Temperature and Friction

Temperature and frictional forces in EHD contacts have a very strong mutual dependence, and, therefore, are discussed under a single heading. In general, sliding speed is the most dominant variable governing temperature and friction. Except at very high rolling speeds, temperature and friction in pure rolling contacts usually do not rise to a significant level to be of great concern. For sliding contacts such as gear teeth, both quantities play major roles in affecting failure.

Typical temperature profiles at the center of a lubricant film and on the contacting surfaces are depicted in

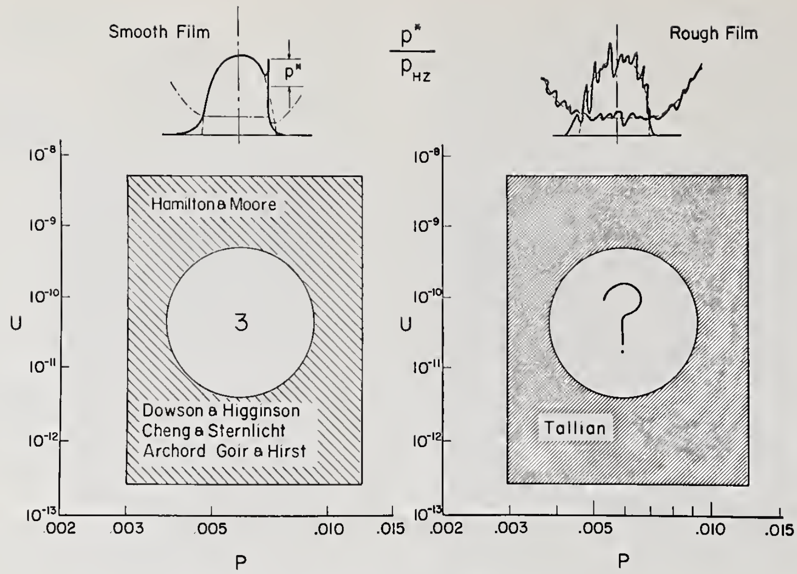


FIGURE 8. *Predictability of pressure perturbation.*

figure 9. In general, the temperature variation corresponds quite well with the pressure distribution, except in the inlet region where the temperature rise occurs much earlier than the pressure curve. Most friction measurements in EHD contacts have exhibited characteristics like that shown in figure 10. For low sliding speeds, the slope is approximately linear, and this slope can be predicted by a transient viscosity model proposed by Harrison and Trachman [20]. With this transient viscosity, the predicted friction and temperature based on an effective viscosity are fairly accurate, and a predictability of 70 percent is given in this low-slip region in figure 11.

As sliding speed increases, the relation between friction and slip becomes non-linear. Analytical predictions in this non-linear region are found to be not reliable,

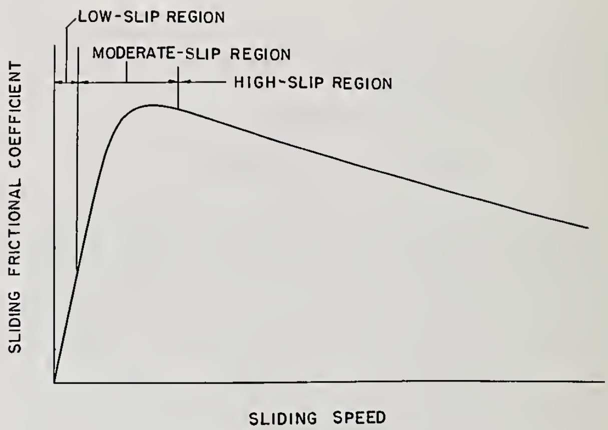


FIGURE 10. *Sliding friction curve.*

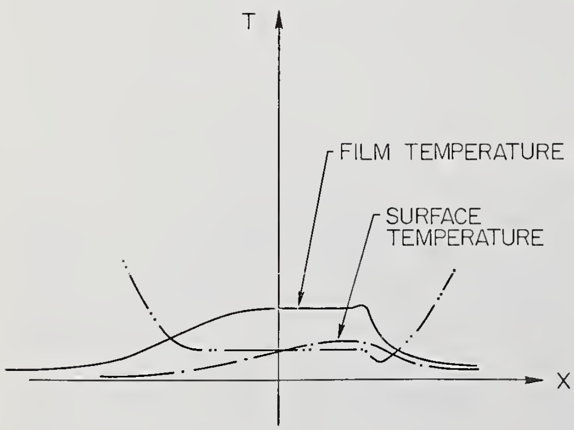


FIGURE 9. *Temperature profiles.*

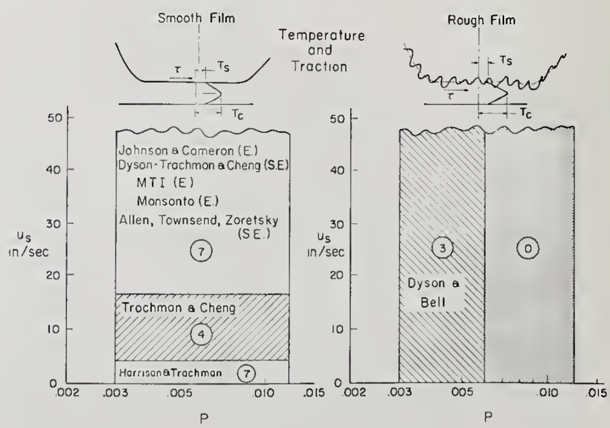


FIGURE 11. *Predictability of temperature and friction.*

and a 40 percent predictability is given. In the region of high sliding speeds, the variation of friction with slip becomes gradual, and reasonably accurate predictions of friction can be made by using empirical friction models [21, 24, 25] and semiempirical models [22, 23, 26, 27]. The expected accuracy may vary somewhat from one friction model to another, but the overall predictability in this region can be regarded as 70 percent.

Analytical developments dealing with friction in partial EHD contacts is still in its very early stage. About the only theory available is the recent contribution by Dyson and Bell [28]. More research is definitely needed here in order to obtain even a marginal predictability of friction and temperature for partial EHD contacts.

5. Effects of EHD on Failure

It has been recognized [29] that the life of concentrated contacts is generally terminated by one of the following modes of failure: subsurface or surface fatigue, scuffing, or mild wear. The advent of EHD has triggered a sustaining effort to determine the extent of EHD influence on each major failure mode. For instance, fatigue experiments with discs under controlled conditions have yielded considerable evidence of the beneficial effect of film thickness on surface fatigue. Other relations between EHD and failure are currently being uncovered. In the following sections, a brief discussion is given to a commonly accepted view of failure mechanisms for each failure mode. In addition, a quantitative assessment is given to the relative influence of the major EHD variables on each failure mode.

5.1. Subsurface Fatigue

Contact fatigue is generally referred to as the loss of contacting material under cyclic, high shear stresses. The material damage can appear in many forms from very small pits to very large spalls. Regardless of the physical appearance of the final failed specimen, contact fatigue can usually be traced from cracks initiated around subsurface inclusions or surface flaws, or from cracks caused by repeated asperity collisions on the surface. This section will be mainly devoted to discussions of fatigue originating from subsurface inclusions. Fatigue originating at the surface will be covered in the next section.

According to Littmann [30], subsurface fatigue originating around large inclusions can be depicted as a sequence of events as shown in figure 12. When a contacting surface is under a repeated normal pressure, small cracks initiate around large inclusions located near layers of large alternating shear stress. These cracks first tend to propagate normally upward until they reach the surface. Once this happens, further propagation can be aided by fluid pressure. Cracks then propagate along a plane slightly inclined to the surface. Spalling finally takes place when some of the branch cracks intersect the surface.

In order to compare the relative influence of the EHD variables with other factors affecting various failure models, an assessment is made in table 2 for each failure model. The relative influence is measured by a number with ten being the most influential and zero being least influential.

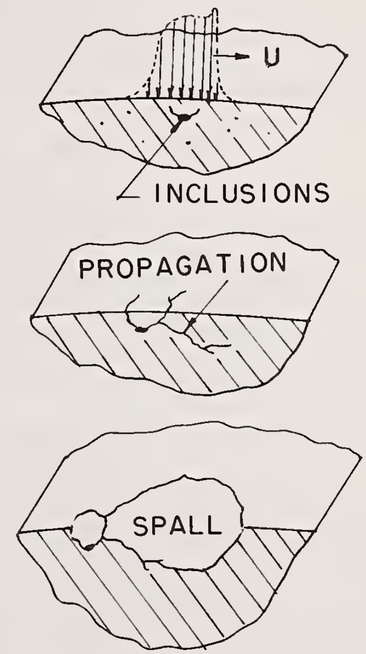


FIGURE 12. *Mechanism of subsurface fatigue.*

TABLE 2. *Relative influence of EHD and other variables on various modes of failure*

	Subsurface fatigue	Surface fatigue	Scuffing	Mild wear
h_c/σ	5	10	8	8
p^*/p_{Hz}	5	8	5	5
T and τ	4	6	10	5
Inclusion density and size	8	—	—	—
Slide to roll ratio	—	8	—	—
Chemical and physical properties of surface film	—	—	10	—

For inclusion originated fatigue, the relative influencing factors are assessed in table 2. The assessment is based on the fact that inclusion originated fatigue is mostly influenced by the subsurface stress distribution, inclusion density and size distribution. Film thickness does not alter subsurface stress distribution, and has little or no influence. The normal pressure as well as surface shear stress can have a significant influence on the failure mode. if they produce a gross redistribution of subsurface shear stress, particularly in the vicinity of maximum alternating shear stress. However, for a heavily loaded EHD contact, subsurface stress deviations resulting from an EHD pressure spike as well as from local sliding friction are not expected to be significant enough to cause a substantial change in fatigue life in inclusion originated fatigue. Thus, in predicting subsurface fatigue life, the effects contributed by EHD are expected to be secondary compared with the influence from inclusion density and size.

5.2. Surface Fatigue

As the material improves in quality, the mode of fatigue is expected to shift from subsurface to that originated on the surface. Littmann [30] called this mode point source origin which is considered to be a major factor in contact fatigue. Surface fatigue can be further divided into two types based on whether cracks are originated around surface defects, such as furrows and nicks, or caused by asperity collisions under marginal or poor EHD lubrication.

The mechanism for surface fatigue from defects is shown in the left diagram of figure 13. The presence of surface flaws tends to interrupt EHD lubrication, and produce large pressure fluctuations and surface distress. This causes cracks to appear around the surface defects, which in turn set up propagation along an inclined plane beneath the surface leading to a fatigue spall.

Figure 13 also depicts the scond type of surface fatigue which is caused mainly by microcracks initiated around asperities when they collide repeatedly with each other under insufficient EHD lubrication. These cracks usually do not penetrate deep into the surface, and propagate along a thin layer just below the surface. Before they have a chance to travel far, they usually intersect with other cracks and break out into small pits. They are also referred to by Littmann [30] as surface peeling, or by Tallian [28] as surface distress.

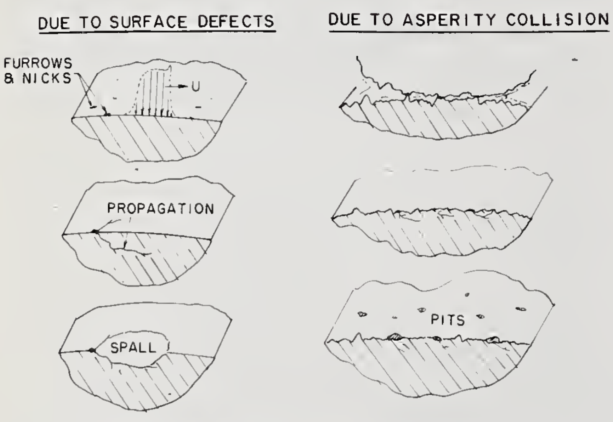


Figure 13. *Mechanism of surface fatigue.*

Since surface fatigue is mainly due to deterioration of lubrication, EHD film thickness is expected to have a strong influence on the failure mode. The relation between EHD film and fatigue life was discovered independently by Dawson [31] with rollers and Tallian et al. [32], with rolling element bearings. After these two pioneering efforts, many fatigue experiments with various contacting geometries have confirmed that fatigue life increases with the specific film thickness Λ, following some power relationship. A typical empirical power relation can be shown in figure 14 based on a series of tests by Lund and Andreason [33]. Thus, EHD is expected to have a strong effect on surface fatigue for contacts which are in the partial film regime and relatively free from inclusions. Qualitatively, the regions separating surface and subsurface fatigue can be shown in figure 15, by plotting fatigue life against the inclusion density and the specific film thickness. It is seen that for a high quality contact material, surface fatigue dominates until Λ exceeds 3 where the inclusion originated subsurface fatigue is likely to terminate the contact life. For material of high inclusion density, subsurface fatigue is expected to dominate the entire range of specific film thickness; EHD lubrication plays a minor role in this regime.

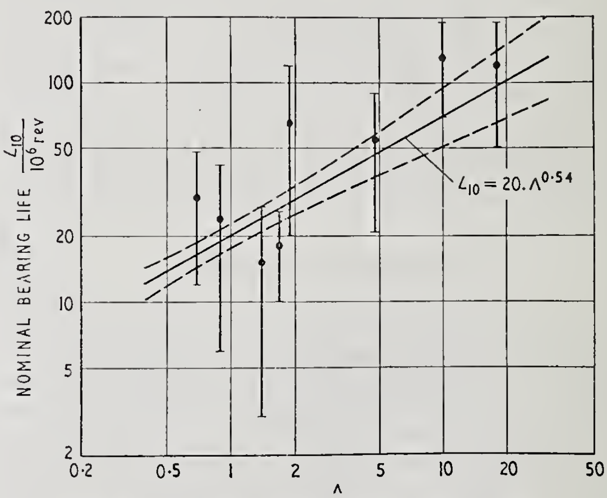


Figure 14. *Relation between Λ and surface fatigue.*

Aside from the film thickness, table 2 also lists several other possible variables influencing surface fatigue. The normal pressure fluctuation and tangential shear stress during asperity collision as well as possible thermal strains may also have some effect on surface fatigue; however, these effects have not been fully explored. A small amount of sliding superimposed on rolling has been shown in past fatigue tests to have a significant effect in reducing fatigue life. This may be explained by the fact that sliding increases the number of collisions for a specific asperity, and hence increases the probability of fatigue failure. There is yet no empirical method to relate life to slide to roll ratios.

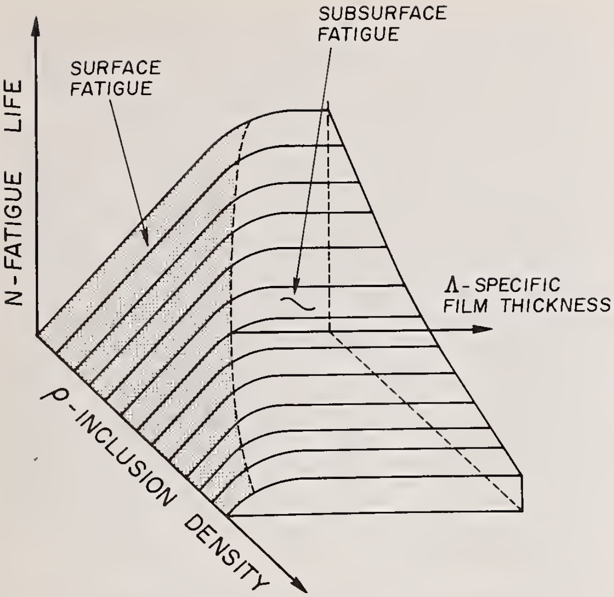


FIGURE 15. *Effect of inclusion density and EHD specific film on fatigue life.*

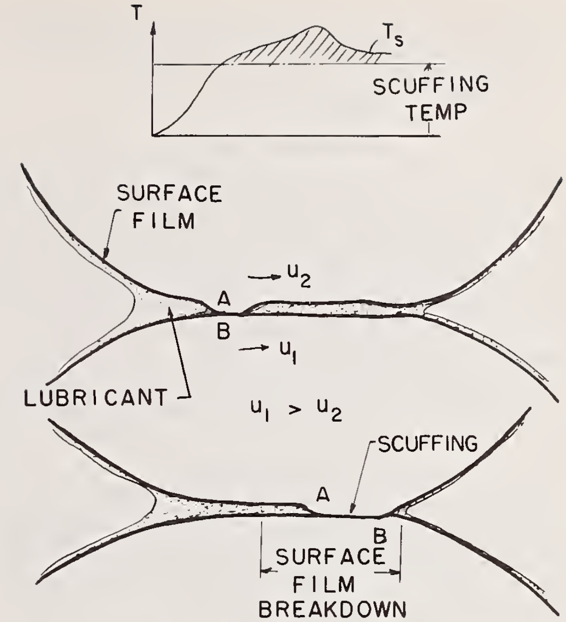


FIGURE 16. *Mechanism of scuffing.*

6. Scuffing

According to Blok [34], scuffing failure in rolling and sliding contacts is characterized by material transfer from one surface to the other through asperity welding which usually results in a rapid deterioration of the surface.

The exact mechanism which leads to scuffing is still not fully understood. Many hypotheses have been proposed, but none has been adopted as an undisputed criterion for scuffing. However, most researchers seem to favor the critical temperature concept which is shown schematically in figure 16. In lubricated contacts, the surfaces are usually covered with an extremely thin surface film which is formed by physical adsorption, chemisorption, or reaction between an additive and the surface material. Such a surface film is regarded as the last line of defense against scuffing. When this surface film is removed, any sliding of an asperity against the opposing surface will lead to a continuous cold welding of the junction causing rapid surface deterioration known as scuffing.

The incipient conditions causing breakdown of these surface films have been under investigation by researchers in boundary lubrication for decades. The problem is extremely difficult, and to this day has not received a fully satisfactory explanation. Nevertheless, the majority of researchers seem to accept the so-called critical temperature concept advanced by Blok for nonadditive mineral oils. He postulates the existence of a constant critical temperature for nonadditive mineral oil (approximately 175 °C) beyond which the surface film would be detached. Others have shown that this critical temperature is not constant, and may be dependent on ambient viscosity.

Before one discusses the possible EHD effects on scuffing failure, one must emphasize that scuffing is only possible if there is asperity contact, i.e., one is dealing with the regime of partial EHD. As such, the determination of the specific film is essential, because it not only shows whether protection from a full EHD film is possible, but it also indicates the severity of asperity contact. For this reason, h_c is interpreted as having a strong effect on scuffing in table 2.

There seems to be little dispute that the maximum surface temperature is the key factor controlling scuffing. To determine this maximum surface temperature, one may use Blok's flash temperature equation [35], or one of the refined flash temperature theories described in [34]. Whatever the method may be, it is essential to know the heat input through the frictional coefficient which still cannot be predicted with accuracy even for full film EHD lubrication. Past predictions of scuffing failure have been based mostly on empirical friction formulas such as those given by Bennedict and Kelly [36]. However, this is not regarded to be entirely satisfactory because there has been little confirmation or correlation of the accuracy of these empirical friction formulas. Needless to say, a comprehensive, accurate, thermal friction analysis in the regime of partial EHD would give a significant lift in this field.

Pressure fluctations around asperities seem to have a more direct influence on fatigue rather than scuffing; however, its indirect influence on local temperature fluctuation could become an important consideration.

While the attention here is focused upon effects of EHD, one should not overlook that prediction of incipient failure based on partial film EHD behavior is only possible if the critical temperature is known beforehand. Keys for determining the accurate critical temperature for a given condition, by and large, still lie in the hands of chemists. The last factor in table 2 for scuffing failure serves as a reminder to us mechanists that EHD is only one facet of the entire problem on scuffing.

6.1. Mild Wear

Peterson [37] suggested that wear in lubricated contacts can be classified into two basic types: metallic contact wear and film wear. In the first mode, wear particles result from the making and breaking of a junction between two sliding asperities, a purely mechanical process. Whereas, in the second mode, wear is a continuous removal of a solid film formed continuously through lubricant-surface interaction. Since EHD is likely to have more influence on metallic wear, subsequent discussions will be confined to this type of wear only.

In figure 17, it is shown that if two asperities slide over each other under partial EHD, a junction is formed. If the level of surface temperature is sufficiently

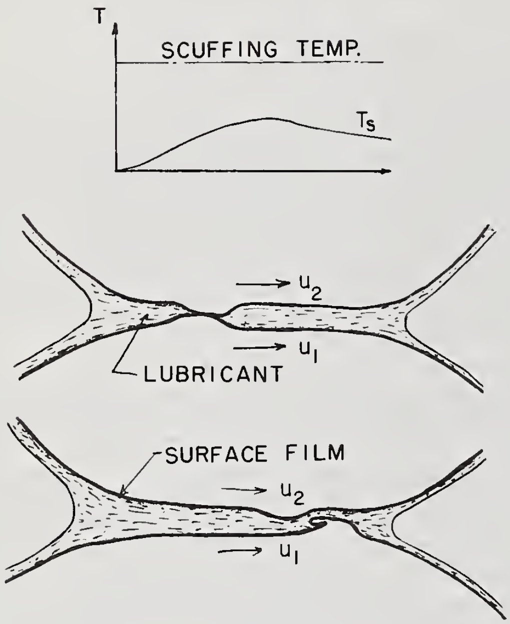


FIGURE 17. *Mechanism of mild wear.*

far below the critical temperature of the surface film, this junction usually does not grow in size, and is broken away as soon as the sliding action is completed. In a large size model, Cocks [38] has shown that such making and breaking of a junction usually produces a wedge of material adhering to one surface. Subsequent sliding usually removes these wedges and produces wear particles. This mechanism is sometimes referred to as adhesive wear.

Besides the adhesive wear mechanism there is still another mechanism for producing wear particles known as fatigue wear. In this mode, wear particles are caused by micropitting around asperities after many collisions with other asperities. Actually, this mechanism can also be interpreted as a surface fatigue failure. Since micropitting, in general, does not terminate life, it is usually referred to as mild wear rather than fatigue failure.

From the description of wear mechanisms, it is not difficult to see a strong relation between wear rate and EHD film thickness, as suggested in table 2. Indeed, experiments with a 4 ball tester [39] have discovered a marked reduction in wear rate as EHD film thickness increases. However, these experiments also showed that the slope between the wear rate and specific film thickness appearing on a log-log plot still varies considerably with lubricants and lubrication condition. It is not certain what factors contribute to the variance of the slope for these cases. There has been speculation that besides the influence of h_e/σ, pressure fluctuations and tangential stresses as affected by different rheological properties of lubricants may also have some influence on the formation of wear particles if fatigue is the wear mechanism. These additional effects may have some influence on the variance in slope observed in [39].

7. EHD and Failure Prediction

As contributions in EHD are increasing at a rapid rate, it is desirable to examine which topics in EHD are more useful in predicting incipient conditions for various modes of failure. In the following, equations for prediction of failure have been suggested. Included in these equations are factors which can account for the influence of EHD. Hopefully, these suggested relations may serve as a guide for establishing a priority for selecting future EHD research topics.

7.1. Subsurface Fatigue Life

In spite of the fact that early important EHD work was partly motivated by its possible influence on subsurface fatigue, all subsequent evidence seems to indicate that EHD has a much stronger influence on surface fatigue than subsurface fatigue. Thus, it is believed that for subsurface fatigue, improvement in life is likely to come from better materials than from better EHD performance.

Past standards for predicting life against subsurface fatigue for rolling element bearings are based mainly upon the classical Lundberg and Palmgren theory [40] which relates the probability of survival to the maximum alternating shear stress, τ_0, the depth at which τ_0 occurs, the size of the contact, and the loading cycle. The variation of inclusion size and density are not accounted for. Recently, two procedures have been made available to include the effects due to changes of inclusion density and size. Both methods propose the use of a life adjustment factor for the standard life predicted from the Lundberg-Palmgren Theory. The first procedure, proposed by Zaretsky et al. [41], is simple, easy to use; the adjustment factors are based on extensive fatigue tests either with full scale bearings or with simulated 4 or 5 ball testers, for bearing materials with impurities controlled by chemical composition and by process variables. The second procedure by Tallian et al. [42], is more basic, but requires much additional input for implementation because one must know beforehand the inclusion density, size distribution, and stress concentration at the inclusion tip.

There seems to be little room for argument that the use of adjustment factors in conjunction with the Lundberg and Palmgren Theory is an effective way to predict subsurface fatigue life. Perhaps what is debatable are the methods by which these factors are determined. In reference [41], these factors are entirely empirical, and require extensive fatigue tests for each new material and process. On the other hand, the factors in reference [42] were obtained purely analytically, and their reliability is somewhat uncertain. In view of these comments, it might be worthwhile to consider a compromised model, which is unrestrictive to the material type or material process, and which seeks an empirical relation between fatigue life and inclusion size and density. This compromised model can be written as

$$\frac{L_{10}}{L'_{10}} = \bar{\rho}^{\,n_1}\ \bar{r}^{\,n_2} \left(\frac{p^*}{p_{Hz}}\right)^{n_3}$$

In the above equation, the ratio of actual fatigue life (90 percent survival) L_{10}, to that based on the Lundberg-Palmgren Theory L'_{10}, is assumed to vary with $\bar{\rho}$ and $\bar{r}$ following a power relation, where $\bar{\rho}$ and $\bar{r}$, are the inclusion density and the inclusion mean radius, respectively, normalized by those found in materials used in Lundberg-Palmgren's tests.

In addition to these two factors, it is also suggested that the secondary pressure peak caused by EHD effects deserves a reexamination regarding its relative influence on fatigue life. A power relation between fatigue life and p^*/p_{Hz} is also suggested, and the exponents n_1, n_2, and n_3 are to be determined by further fatigue experiments with controlled inclusion density and size.

7.2. Surface Fatigue Life

As described earlier, surface fatigue is strongly affected by specific film thickness h_c/σ if the contact operates under partial EHD. The film thickness effect has been incorporated in references [41] and [43] to predict fatigue life against surface fatigue. In reference [41], an adjustment factor is used to account for the effect of h/σ, and this factor is largely based on a selected number of fatigue tests. Whereas in reference [43], an analytical approach has been favored for determining the influence of h_c/σ. However, there have been no efforts to correlate these two approaches.

So far, among the major EHD variables only one identified the specific film thickness as a primary influence on surface fatigue. Other factors such as pressure fluctuations and slide to roll ratios may also have significant influence but are yet to be explored. It is speculated that for microcracks initiated around asperities, the local pressure variation, as influenced by rheological properties of different lubricants, and the frequency of asperity collision, as affected by the slide to roll ratio, must have some influence on its occurrence. If these two effects are included, one might consider the following relation for future predictions of surface fatigue life.

$$L_{10} = C\left(\frac{h}{\sigma}\right)^{m_1}\left(\frac{p^*}{p_{Hz}}\right)^{m_2}\left(\frac{U_s}{U}\right)^{m_3}$$

where L_{10} = surface fatigue life, based on 90 percent survival

C = a bearing material constant

m_1 = exponent for EHD film effect, approximately equal to 0.6

m_2, m_3 = empirical exponents for effects due to pressure fluctuations and slide to roll ratios.

7.3. Scuffing Load

Incipient scuffing for a concentrated contact has been customarily measured by its ultimate load capacity against scuffing when all other variables such as speed and material properties are kept constant. This incipient load is known as the scuffing load, which is a familiar term in gear technology.

As described earlier this incipient scuffing load is regarded as the load at which the flash temperature, as calculated from Blok's theory, or any one of its refined versions, reaches the critical temperature of a lubricant. The functional relation between flash temperature and scuffing load can be expressed symbolically as

$$T_f - T_b = F\ (P, f, \text{Speed, Material, and Lubricant Parameters})$$

where T_f = the flash temperature or the maximum surface temperature

T_b = temperature of the surface before entering the contact region

f = EHD frictional coefficient, which is dependent upon the specific film h_c/σ and the shear modulus of the lubricant G_∞

P = contacting load.

The symbol F stands for Blok's flash temperature function or one of the more refined but more complicated versions.

The inverse form of this function G permits one to compute the scuffing load P^* provided that the critical temperature T^*_f is given,

$$P^* = G\left\{T^*_f - T_B, f\left(\frac{h_c}{\sigma}, G_\infty\right), \text{etc.}\right\}$$

The accuracy in predicting P^* by means of the above relation would depend on:

1. The validity of the flash temperature theory F in calculating T_f.
2. The predictability of the lubricant critical temperature T^*_f.

3. The predictability of the base temperature T_B.
4. The predictability of EHD frictional coefficient f.

It appears that the state-of-the-art of flash temperature theory is overdeveloped compared with the predictabilities of T^*_f, T_B, and f. There are uncertainties in all available methods of predicting these three quantities. While chemists concentrate their efforts in resolving questions, such as whether there exists a single constant critical temperature for a mineral oil or whether the critical temperature is also speed dependent, mechanists should accelerate their efforts on a better heat transfer analysis of T_B and a more accurate analysis of EHD friction. Thus, if these three terms can be determined with reasonable accuracy, there is no reason to doubt that the prediction of scuffing load based on existing flash temperature theories will yield a satisfactory result.

7.4. Wear Rate

Even though there has been some scanty evidence indicating that the wear rate in EHD contacts resulting from the metallic contact mode appears to be amenable to Archard's wear law, more correlations are still needed before it can be applied with full assurance.

According to Archard's wear law, wear rate for rolling and sliding EHD contacts should obey the following relation [44]:

$$w = K\left(\frac{A_R}{A}\right)u_s = K\left(\frac{h_c}{\sigma}\right)^{-n} u_s$$

where w = wear rate

K = a wear constant

$\frac{A_R}{A}$ = ratio of the real area of contact to the apparent area;

$$\frac{A_R}{A} = \left(\frac{h_c}{\sigma}\right)^{-n}$$

$\frac{h_c}{\sigma}$ = specific film

n = slope on log-log plot between w and h_c/σ

u_s = sliding velocity.

It is not certain to what extent pressure fluctuations or tangential shear stresses may influence K and n. One can only hope that their influence will be relatively mild so that the wear rate in EHD contacts can be predicted with the simple relations suggested above.

8. Concluding Remarks

(1) It seems evident that among the major EHD variables, the specific film thickness definitely plays the most important role in influencing failure by surface fatigue, scuffing, and mild wear. Present methods in predicting the specific film h_c/σ for determining partial or full EHD is considered to be quite reliable based on existing EHD analyses. However, the accuracy when $h_c/\sigma << 3$, i.e., at the far end of the partial film regime, is still uncertain because of the lack of knowledge on the effect of surface roughness on EHD film analysis.

(2) For fatigue originating at subsurface inclusions, EHD seems to play a secondary role compared to material effects. The EHD influence is from the redistribution of the subsurface stress field due to secondary pressure spikes and from the tangential shear stress. A quantitative determination of these effects is needed.

(3) For fatigue originating at surface defects and surface asperities, EHD definitely plays a significant role. There are empirical relations now available for predicting the surface fatigue life as influenced by the EHD film thickness; however, other EHD effects such as normal pressure and tangential shear are yet to be explored.

(4) Determination of scuffing failure seems to hinge critically on the predictability of: EHD frictional force and surface temperature in the partial film regime, base surface temperature prior to entering the contact region, and the critical temperature for a given lubricant. Considerably more effort is needed in the study of friction under partial EHD to enhance the reliability in predicting scuffing failure.

(5) Prediction of wear rate, at least for rolling EHD contacts, seems to be amenable to Archard's wear law which relates wear rate and EHD film linearly on log-log grids. However, the slope seems to be dependent on other EHD variables; further explorations are needed.

9. Nomenclature

a, b	=	semimajor and semiminor axis of Hertzian ellipse
A	=	apparent area of contact
A_R	=	real area of contact
C	=	a material constant in the suggested formula for predicting surface fatigue life
E	=	reduced elastic modulus
f	=	sliding frictional coefficient
G	=	the pressure-viscosity parameter for EHD contact; also used as a function for the scuffing load
G_∞	=	shear modulus of the lubricant
h_T	=	trailing minimum film thickness
h_s	=	side minimum film thickness
h_c	=	center film thickness
K	=	wear constant
L_{10}	=	fatigue life based on 90 percent survival
L'_{10}	=	fatigue life based on Lundberg and Palmgren theory for 90 percent survival
m_1, m_2, m_3	=	exponents in the suggested formula for predicting surface fatigue life
n_1, n_2, n_3	=	exponents in the suggested formula for predicting subsurface fatigue life

n = exponent in the formula for predicting wear rate
P = contacting load; also used as the load parameter for EHD contact
P^* = scuffing load
p_{Hz} = maximum Hertzian stress
p^* = pressure deviated from Hertzian profile
$\bar{r}$ = normalized inclusion tip radius
R_{x1},R_{x2} = radius of curvature in the direction of rolling for bodies 1 and 2
R_{y1},R_{y2} = radius of curvature perpendicular to the direction of rolling for bodies 1 and 2
S = slide to roll ratio
T_f = flash temperature or maximum surface temperature
T^*_f = critical temperature
T_B = base temperature
u = rolling velocity
u_1,u_2 = velocity of bodies 1 and 2
u_s = sliding velocity
w = wear rate
α = pressure-viscosity
Λ = specific film
μ_o = ambient viscosity
$\bar{\rho}$ = normalized inclusion density
σ = composite surface roughness

10. References

[1] Grubin, A. N., and Vinogradova, I. E., Central Scientific Research Institute for Technology and Mechanical Engineering, Book No. 30 (Moscow); D.S.I.R. No. 337.
[2] Dowson, D., Elastohydrodynamic lubrication: An introduction and review of theoretical studies, Lubrication and Wear Group Symposium on EHL, Paper R1.
[3] McGrew, J., Gu, A., Cheng, H. S., and Murray, F., Elastohydrodynamic Lubrication, Phase I, Technical Documentary Report AFAPL–TR–70–27.
[4] Dowson, D. and Higginson, G. R., 'The effect of material properties on the lubrication of elastic rollers,' J. Mech. Eng. Sci. **2** (3), 188.
[5] Archard, J. F., and Cowking, E. W., A simplified treatment of elastohydrodynamic lubrication theory for a point contact, Lubrication and Wear Group Symposium on Elastohydrodynamic Lubrication Paper 3 (Inst. Mech. Eng., London).
[6] Cameron, A., and Gohar, R., Theoretical and experimental studies of the oil film in lubricated point contact. Pro. R. Soc., London, Series A, Vol. **291**, pp. 520–536.
[7] Cheng, H. S., A numerical solution of the elastohydrodynamic film thickness in an elliptical contact, Journal of Lubrication Technology, Trans. ASME, Vol. **92**, Series F, No. 1.
[8] Greenwood, J., and Kauzlarich, J., Inlet shear heating in elastohydrodynamic lubrication, Journal of Lubrication Technology, Trans. ASME, Vol. **95**, Series F, No. 4, pp. 417–426.
[9] Cheng, H. S., Calculation of Elastohydrodynamic Film Thickness in High-Speed Rolling and Sliding Contacts, Mechanical Technology Technical Report MTI–67TR24.
[10] Herrebrugh, K., Solving the incompressible and isothermal problem in elastohydrodynamic lubrication through an integral equation, Journal of Lubrication Technology, Trans. ASME, Vol. **90**, Series F, No. 1, pp. 262–270.
[11] Cheng, H. S., Isothermal elastohydrodynamic theory for the full range of pressure-viscosity coefficient, Journal of Lubrication Technology, Trans. ASME, Vol. **94**, Series F, No. 1, pp. 35.
[12] Dyson, A., Naylor, H., and Wilson, A. R., The measurement of oil film thickness in elastohydrodynamic contacts. Lubrication and Wear Group Symposium on Elastohydrodynamic Lubrication, Paper 10 (Inst. Mech. Eng., London).
[13] Sibley, L. B., and Orcutt, F. K., Elastohydrodynamic lubrication of rolling contact surfaces, Am. Soc. Lubr. Eng. **4** (2), 234.
[14] Gohar, R., Oil film thickness and rolling friction in elastohydrodynamic point contact, Paper No. 70–Lub–2, Presented in the ASME–ASLE Joint Conference in Cincinnati, Ohio.
[15] Chow, L. S. H., and Cheng, H. S., The effect of surface roughness on the nominal film thickness of elastohydrodynamic contacts, to appear as a NASA Contractor Report in 1974.
[16] Christensen, H., Stochastic models for hydrodynamic lubrication of rough surfaces, Pro. Inst. Mech. Eng., Tribology Group, Vol. **184**, Pt. 1, No. 55, p. 1013.
[17] Jacobson, B., On the lubrication of heavily loaded cylindrical surfaces considering surfaces deformations and solidification of the lubricant, Journal of Lubrication Technology, Trans. ASME, Vol. **95**, Series F, No. 3.
[18] Jacobson, B., On the lubrication of heavily loaded spherical surfaces considering surface deformations and solidification of the lubricant, ACTA Polytechnica Scandinavia, Mech. Eng. Series **54**, Stockholm.
[19] Tallian, T. E., Pressure and traction rippling in EHD contact of rough surfaces, ASME Paper No. 73–Lub–18.
[20] Harrson, G., and Trachman, E. G., The role of compressional viscoelasticity in the lubrication of rolling contacts, J. of Lubrication Technology, Trans. ASME, Vol. **95**. Series F, No. 4, pp. 306–312.
[21] Johnson, K. L., and Cameron, R., Shear behavior of elastohydrodynamic oil film at high rolling contact pressure, Proceedings of Inst. of Mech. Eng., Vol. **182**, p. 307.
[22] Dyson, A., Frictional traction and lubricant rheology in elastohydrodynamic lubrication, Philosophical Trans., of the Royal Society of London, Vol. **266**, No. 1170, p. 1–33.
[23] Trachman, E. G., and Cheng, H. S., Thermal and non-Newtonian effects on traction in elastohydrodynamic contacts, Submitted for presentation and publication at the Second Internat'l Conference on Elastohydrodynamic Lubrication, Leeds, England.
[24] Smith, R., et. al., Research on EHD Lubrication of High Speed Rolling-Sliding Contacts, Technical Report AFAPL–TR–56.
[25] Hammann, W. C. et. al., Synthetic Fluids for High Capacity Traction Drives, ASLE Paper 69–LC–10.
[26] Allen, C. W., Townsend, D., and Zaretsky, E., Elastohydrodynamic lubrication of a spinning ball in a nonconforming groove, J. Lubr. Technol., Trans. ASME, Series F, Vol. **92**, No. 1. pp. 89–96.
[27] McCool, J. I., et. al., Influence of EHD Lubrication and Operation of Turbine Engine Ball Bearings, SKF Technical Report AL73PO14.
[28] Bell, J. C., and Dyson, A., The Effect of Some Operating Factors on the Scuffing of Hardened Steel Discs. Second Symposium on EHL, Leeds, England, Paper C11/72.
[29] Tallian, T., On competing failure modes in rolling contact, ASLE paper no. 67AM1c–3 presented at the 1967 ASLE Annual Meeting in Toronto, Canada.
[30] Littmann, W., The Mechanism of Contact Fatigue, Interdisciplinary Approach to the Lubrication of Concentrated Contacts, NASA SP–237, pp. 309–377.
[31] Dawson, P. H., Further experiments on the effect of metallic contact on the pitting of lubricated rolling surfaces, Lubrication and Wear Group Symposium on EHL, Paper No. 7.
[32] Tallian, T. E., Sibley, L. B., and Valori, R. R., Elastohydrodynamic Film Effects on The Load-Life Behavior of Rolling Contacts, Paper 65–Lubs–11, ASME, New York, N.Y.
[33] Andreason, S. and Lund, T., Ball-bearing endurance testing considering EHL, Presented at the 1972 EHL Symposium held at Leeds University, Leeds, England, Paper no. C36/72.

[34] Blok, H., The postulate about the constancy of scoring temperature, NASA Special Publication–237. Interdisciplinary Approach to the Lubrication of Concentrated Contacts, pp. 153–249.
[35] Blok, H., Surface temperatures under extreme pressure conditions, Congr. Mondiale Petrole Assoc., Franc. des Techniciens du Petrole, Paris.
[36] Bennedict, G. H. and Kelley, B. W., Instantaneous coefficients of gear tooth friction, Trans. Amer. Soc. Lubr. Eng. Vol. **4**, 59.
[37] Peterson, M., Mechanisms of wear, boundary lubrication—an appraisal of world literature, edited by Ling, Klaus, and Fein, an ASME Publication, pp. 19–34.
[38] Cocks, M., Interaction of sliding metal surfaces, JAP 33, Vol. **7**, p. 2152.
[39] Tallian, T. E., McCool, J. I., and Sibley, L. B., Partial elastohydrodynamic lubrication in rolling contact, Proceeding of Institution of Mechanical Engineers, Paper 14, p. 169, London.
[40] Lundberg, G. and Palmgren, A., Dynamic capacity of rolling bearings, ACTA Polytechnica, Mechanical Engineering Series, Vol. **1**, No. 3.
[41] Zaretsky, E. V. et al., Life Adjustment Factors for Ball and Roller Bearings—An Engineering Design Guide, ASME Publication.
[42] Tallian, T. E., and McCool, J. I., An engineering model of Spalling failure in rolling contact—the surface model, Wear **17**, pp. 447–461.
[43] Tallian, T. E., An engineering model of spalling fatigue failure in rolling contact—engineering discussion and illustrative examples, Wear **17**, pp. 463–480.
[44] Archard, J. F., A Review of Wear Studies, NASA SP–181, Interdisciplinary Approach to Friction and Wear.

11. Discussion

J. E. Stern, NASA, Goddard Space Flight Center: I am a little confused. Would you please explain the difference in the mechanism between wear and scuffing?

H. S. Cheng: In scuffing, the surface temperature exceeds the temperature at which the surface film breaks down. There is no protection of the surface and there is basically local welding. Material is transferred from one surface to the other resulting in the junction growth phenomenon. In the wear mechanism, surface temperatures are not high enough for the surface film to break down. Asperities can slide over each other and, in so doing, produce wear debris. There is no continuous propagation of plastic deformation.

N. Glassman, Naval Ship Research & Development Center: There are possibilities in what you call lubricant parameters for chemical effects, particularly in the partial lubrication regime and even in the full film regime. We have data from rolling contact fatigue experiments with Λ factors larger than 3. With small amounts of water added to the lubricant, there are early fatigue failures. In other work, possibly with Λ less than 3, the use of different additives results in different load carrying capacities of gears. A large part of the reduced gear load carrying capacity which might be attributed to temperature effects might be due instead to the chemical reactivity of additives.

H. S. Cheng: I am aware of the effect of water content and other additives on fatigue life. There has been some work at Battelle which shows that additives have an influence on the film thickness. The actual film thickness is not the same as the film thickness calculated on the basis of the bulk viscosity.

N. Glassman: As I say it's possible. Both physical and chemical effects may be operative.

H. S. Cheng: Yes.

D. C. Drucker, University of Illinois: I am puzzled that you added the density of the inclusions to the list of fatigue variables. I would have thought that the size of the inclusions, rather than the density, was more important. In a line contact problem, for example, the density enters in as related to location. Except for that, I would have thought the size would be more important.

H. S. Cheng: I agree that the size of the inclusions also plays an important role in contact fatigue, and it should have been included in the list of fatigue variables.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Failure in Gears and Related Machine Components

R. A. Wilde

Eaton Corporation
26201 Northwestern Highway
Southfield, Michigan 48076

Tooth breakage and abrasive wear are gear failure mechanisms that are well understood. Adhesive wear (scuffing and scouring) and surface pitting are more complex failure mechanisms.

When sliding is introduced into the rolling condition the rubbing between partially lubricated surfaces introduce high surface tensile stresses which result in adhesive wear and surface pitting. Neither of these failures will occur if a continuous oil film separates the metal surfaces.

Because these failures are the result of the interaction of many variables, a systems approach is required to understand the nature of the failure. The following variables have a direct effect on the thickness of the oil film and therefore on the amount of metal to metal contact: Surface roughness, direction of scratches relative to the sliding direction, oil viscosity, conjunction temperature, rolling speed, relative sliding speed, oil composition and load (Hertz stress). These variables will have significantly different effects when sliding is introduced to the rolling condition as occurs in gear applications.

Key words: Abrasive wear; adhesive wear; gear failure; machine component failure; mechanical failure; surface pitting.

1. Introduction

This paper will describe the types of failures that gears might experience. The general principles that are advanced will also apply to clutches, brakes and other components. Tooth breakage and abrasive wear are gear failure mechanisms that are rather well understood so they will be reviewed only briefly. Most of the discussion will concentrate on failures that are caused by adhesive wear and surface pitting which are more complex failure mechanisms. Neither of these failure mechanisms will occur if a continuous oil film separates the metal surfaces.

The factor that distinguishes this paper from that of Dr. Walter Littmann's is that the two mating surfaces experience substantial relative sliding in addition to rolling.

2. Tooth Breakage

Gear teeth under load act like cantilever beams. Bending loads on gear teeth usually cause the highest stresses to occur at root fillets. Whole gear teeth or substantial portions of the teeth can be broken off as the result of high overloads, often of an impact nature. Also, fatigue failures can occur as a result of high cyclic stresses.

Gear tooth fractures may be averted by proper consideration of design, metallurgical and machining factors [1,2].[1]

[1] Figures in brackets indicate the literature references at the end of this presentation.

3. Abrasive Wear

Abrasive wear takes place whenever hard, foreign particles, such as metal grit, metallic oxides and dirt from the environment are present between the rolling and sliding gear surfaces.

These hard particles first penetrate the metal, adhere temporarily to one surface, and, in turn, tear metallic particles from the second surface. Abrasive wear normally does not occur in a gear drive unless the lubricating oil becomes contaminated with foreign particles that are larger than 30 micrometres [3].

The best abrasive wear resistance is generally found in an alloy steel with a structure of uniformly distributed fine carbide in a martensite matrix containing some residual austenite. The bearing alloy 52100 properly heat treated has this type of structure. Residual austenite is thought to add toughness to the matrix, provide support for the carbide and be capable of undergoing transformation to martensite locally during wear [4].

The abrasive wear resistance of an alloy steel will increase with its carbide content but only if it is finely distributed. A carburized steel with 0.09/1.00 percent surface carbon heat treated properly to obtain a good distribution of the excess carbides would be a reasonably good structure (figure 1).

4. Scuffing Failure

First, the failures caused by extreme conditions of adhesive wear, which will be called scuffing failures,

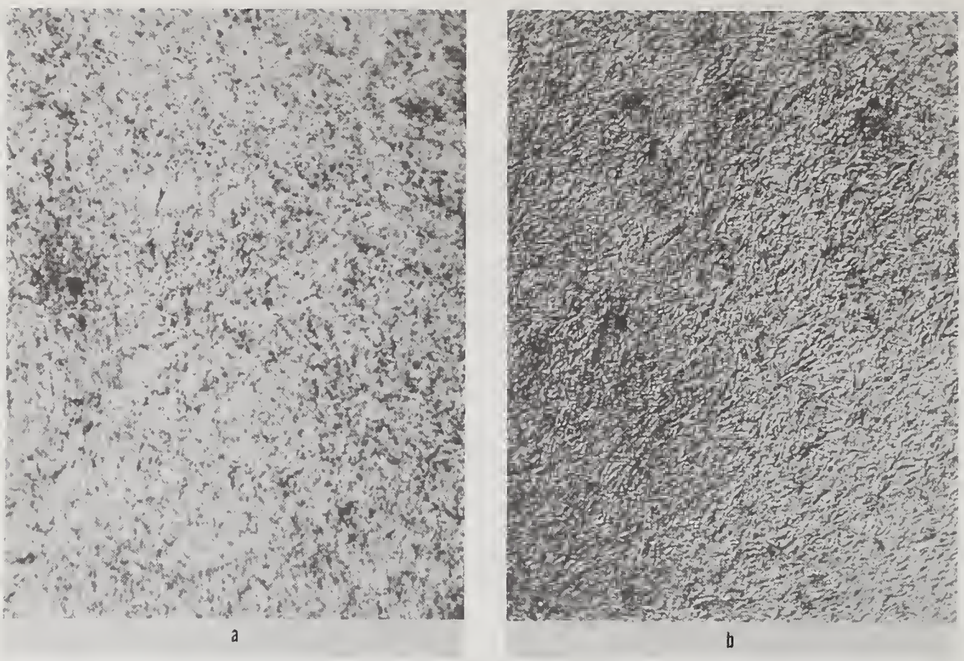


FIGURE 1. *Surface microstructure of carburized (0.90/1.0 percent C) 8620 steel with (a) fine spheroidized carbides and (b) fine martensitic needles.*

will be examined. These failures occur under rolling and sliding conditions when a critical area of adhesive contact takes place through the oil film which generates enough heat so that the oil is rolled away from the surface ahead of the contact. This enables the contact area to spread in an uncontrolled manner causing catastrophic scuffing. The way to prevent adhesive wear is to prevent all metal to metal contact through the oil film.

What are the factors that influence the oil film between the surfaces? To simplify the discussion, we will consider either gears or test rollers that are rolling and sliding against one another at load pressures in the range of 50 000 psi to 300 000 psi mean Hertz stress. In the case of gears at these loads, several things take place. First, both surfaces elastically deform. This increases the load carrying contact area, and consequently the pressure distribution over the contact area is changed. Second, the viscosity of the oil increases due to the rise of the pressure. Third, the shear rate has a permanent or semi-permanent effect on the viscosity of various oils. The higher the shear rate, the lower the apparent viscosity will be. This shear rate increases rapidly with the increase in relative sliding velocities of the gear faces. Finally, the temperature of the gear surface is raised by the elastic flexing of the material under load and by asperity contact through the oil film in heavily loaded gears which causes a temperature rise in the oil and has the effect of lowering the viscosity [5].

There is another way of looking at two metal surfaces that are held apart by a film of oil. If any load is applied the oil film will gradually be squeezed out from between the metal surfaces. The rate at which the oil is squeezed out is dependent on a number of factors. The higher the load the quicker the oil is squeezed out. The lower the oil viscosity the quicker the oil is squeezed out. The higher the temperature of the oil the lower the viscosity and the quicker the oil is squeezed out. The shorter the distance the oil has to travel, i.e., the smaller the area of contact, the quicker the oil is squeezed out. The oil viscosity is increased somewhat by load and decreased by the shearing action in the oil as the relative sliding velocity is increased. As two rollers roll against each other the oil is wedged into the contact. The higher the speed the more oil is wedged between the rollers and the less time there is to squeeze the oil out.

So far we can translate these statements into these relationships: Resistance to adhesive wear and scuffing is increased by:

Higher oil viscosity
Higher roller or gear speeds
Lower operating temperatures
Lower unit load on surface
Lower relative sliding velocity
Larger areas of contact

This brings us to the Critical Temperature Equation of Blok. The Critical Temperature Theory proposes that every oil (i.e., oil mixture) has a temperature beyond which scuffing will occur. One of the basic problems of trying to develop mathematical formulas for predicting scuffing is that scuffing initiates on a microscopic scale and evidence is being developed on the basis of macroscopic or bulk measurements.

This critical temperature is the sum of two temperatures as follows:

$$T_{cr} = T_r + \Delta T_{cr}$$

where T_{cr} = Temperature at surface for scuffing

T_r = Bulk roller (gear) temperature

ΔT = Momentary rise in temperature between rollers (gears) from frictional heating (from Blok equation).

This rise in temperature is calculated as follows:

$$\Delta T = \frac{0.79\, fWe\, (\sqrt{V_1} - \sqrt{V_2})}{a\sqrt{b}}$$

where f = Coefficient of friction

W_e = Equivalent unit load, lb/in

V_1 = Surface speed of upper roller, in/s

V_2 = Surface speed of lower roller, in/s

a = A material constant, $\mathrm{lb/in^{-1}\, s^{-1/2}\, {}^{\circ}F^{-1}}$

b = Semiwidth of contact ellipse (minor axis), in

a = Block thermal coefficient

= $(kpc)^{1/2} = 41.8\ \mathrm{lb/{}^{\circ}F\ in\ s^{1/2}}$

k = thermal conductivity

p = density

c = specific heat

This equation shows that ΔT is increased and therefore scuffing is adversely affected by an increase in sliding velocity, load, and coefficient of friction. The coefficient of friction represents the energy losses when the two surfaces are brought close together. The major loss would be adhesion and separation of asperity peaks between the two surfaces. Minor losses would be from resistance to oil shear and overcoming other attractive forces between the metal surfaces of a noncontact nature.

The major deficiency of the critical temperature hypothesis is that it does not consider surface roughness, which has a large effect on T_{cr}. Some data developed at the Southwest Research Institute will give some idea of the magnitude of this effect. The tests were run in a Caterpillar roller test rig using 3 in diam rollers with a 14 in crown radius and phasing gears were used to develop the desired amount of relative sliding. The rollers were ground in the circumferential direction. The surface roughness was measured in the circumferential and transverse (axial) directions and averaged together to obtain the composite roughness, in μin CLA.

Figure 2 shows typical surfaces as ground and after break-in. At this higher surface finish there was little difference in the surface roughness number. Figure 3 shows another roller with a rougher surface which was substantially improved by break-in (composite 12.5 to 9.5–25 per cent improvement). During break-in the projecting asperities on the surface are worn away as the rolls are step loaded to a high preload level. It is obvious that the broken-in surface at 13.0/6.0 μin is different than a newly ground surface of the equivalent surface roughness. The pointed asperities have become flats and the volume of the valleys into which the oil can be squeezed is considerably reduced.

FIGURE 2. *Scanning electron micrographs of ground roller XX173L.*

(a) As ground—× 540. Surface finish in tranverse and circumferential directions 5.0/4.5 RMS.
(b) As ground—× 1620.
(c) As ground plus break-in—× 540. Surface in transverse and circumferential directions 4.5/4.0 RMS.
(d) As ground plus break-in—× 1620.

The surface roughness in the range of 4 to 30 has a significant effect on the critical temperature T_{cr}. An increase of 1 μin δ_i reduces the critical temperature by 5 °F. In the program δ_i varied by 12 uin so this effect could change the T_{cr} by 60 °F. An increase of 1 μin in δ_f reduces the critical temperature by 20 °F. In the program, δ_f varied by 8 μin so this effect could change the T_{cr} by 160 °F. which is substantial.

The purpose of this example was to show that surface roughness of gears and rollers or other devices will have a significant effect on the load at which scuffing occurs.

The direction of the grinding marks also has a big effect on loads to failure. SWRI had some rollers ground transverse to the direction of sliding and rolling and

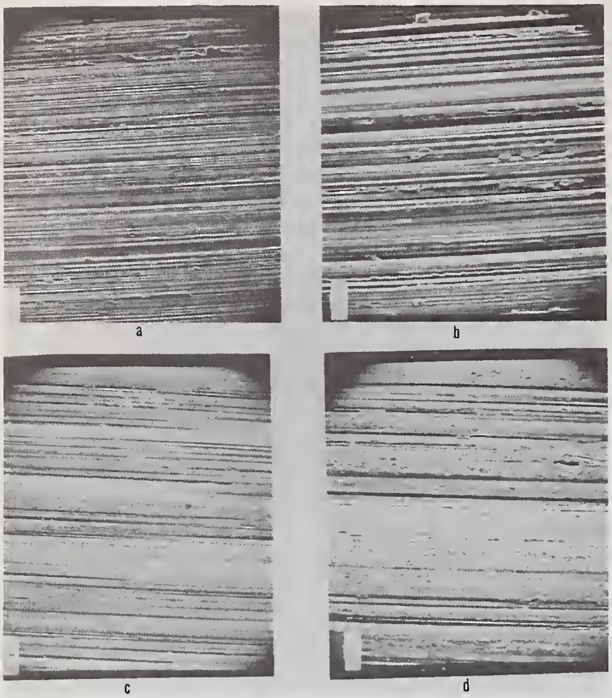


FIGURE 3. *Scanning electron micrographs of ground roller XX95U.*

(a) As ground—X 160. Surface finish in tranverse and circumferential directions 17.5/7.5 RMS.
(b) As ground—X 530.
(c) As ground plus break-in—X 160. Surface finish in transverse and circumferential directions 13.0/6.0 RMS.
(d) As ground plus break-in—X 530.

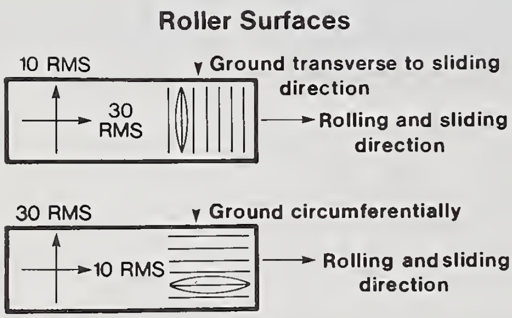


FIGURE 4. *Wear scars in rollers ground in transverse and circumferential directions.*

loads to failure were compared with rollers ground circumferentially (fig. 4). The rollers were run at a high load with no break-in. The rollers ground circumferentially failed immediately with long circumferential scuffing scars.

The rollers ground in the transverse direction developed wear scars, transverse to the rolling direction, at substantial loads. The interrupted nature of the surface prevented the wear scar from spreading in a catastrophic manner in the rolling and sliding direction. This shows that the composite surface roughness values, which in this experiment would be the same, are not enough to describe the critical surface condition. This also suggests that cutting gears transverse to the rolling and sliding direction is an aid in resisting scuffing but was not done for this reason, and that the spline machine marks are in the wrong direction to minimize scuffing during axial movement.

In the case of heavy-duty gears such as truck ring and pinion gears in the differential, the loads are in the range of 250 000/300 000 psi Hertz stress which put the gears into boundary lubrication, i.e., intermittent metal to metal contact is occurring. To prevent scuffing, additives such as sulphur and phosphorus are added to the oil. At the temperature range that the gears will see in service (250/300 °F), these additives react chemically with the steel surface to form sulphides and phosphides which prevent the two gear surfaces from adhering to each other. This allows the differential to operate successfully in the boundary lubrication regime.

Another way of minimizing metal adhesion in heavy-duty gearing operating in the boundary lubrication regime is to introduce some ammonia (5 cfm) in the last stages of carburizing, especially in those cases where the gears are cooled to 1575/1600 °F before quenching. The steel surface will accept nitrogen readily at these temperatures and the concentration of iron nitrides in the surface will significantly reduce adhesion.

Scuffing type failures can be prevented by the following means:

(1) Reduce the conjunction temperature.
(2) Use as smooth surfaces as practical. If there are machining marks, have them perpendicular to the rolling and sliding direction.
(3) Use lots of oil—direct it on the most heavily loaded areas.
(4) Cool the oil.
(5) In those cases where the gears are operating in the boundary lubrication regime, use E.P. additives in the oil such as sulphur and phosphorus to form chemical films on the surface.

5. Surface Pitting

How does sliding affect the surface stresses and how does this influence failure by surface pitting (fig. 5)? What happens in the case of gears has been simulated by means of two rollers, by numerous investigators, in which the rollers are rolling together in the direction as shown by the arrows, but not at the same peripheral speed. The relative motion of the rollers is such that the top roller (the faster one) appears to be sliding in the direction of rolling (positive sliding) and the lower roller (the slower one) appears to be sliding in the opposite direction of rolling (negative sliding). In the case where the rollers are forced hard together radially so metal to metal contact can occur, tangential frictional forces are developed at the surface which are tension stresses in the areas marked t–t and compression stresses in the areas marked c–c. This will occur under conditions of boundary lubrication where there is intermittent metal to metal contact. When there is no lubrication, adhesion would occur and catastrophic failure would follow quickly.

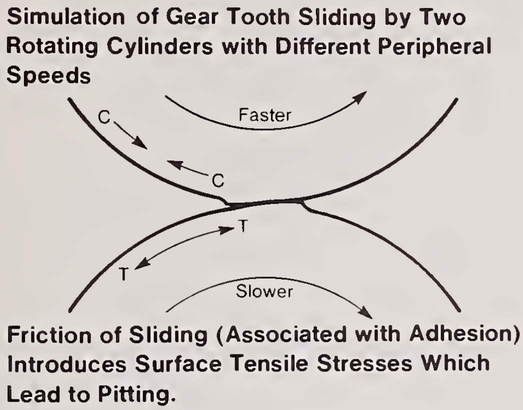


FIGURE 5. *Stresses developed at the surfaces of rollers sliding and rolling together.*

Figure 6 is an enlarged view of what is taking place on the surface. In this particular case, the oil is shown being fed in advance of the rolling point. Any minute cracks or imperfections on the surface on the lower roller would be opened up and correspondingly closed up on the upper roller. The oil would, of course, then be forced into these cracks and raised to very high pressures by its passage under the point of contact, with resulting progressive fatigue failure which would finally lift small pieces out of the surface. Observations show that the pits invariably form on the roller where the friction of the sliding would produce tensile stresses (negative sliding) and do not form on the roller where the sliding would produce compressive stresses. This is so marked that even when the upper roller, the high speed one which is in positive sliding, is made one half the hardness of the lower roller, the upper roller develops no pits [6].

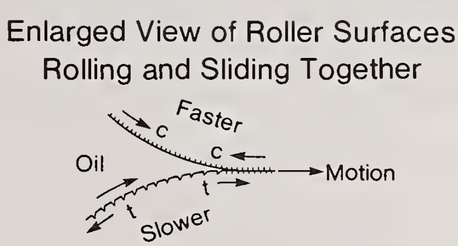


FIGURE 6.

Right away this would suggest that, if the surface could be put in compression to such an extent as to more than offset the tensile stresses which would be caused by the sliding friction, then it might be possible to preclude the formation of this type of pitting on gear teeth.

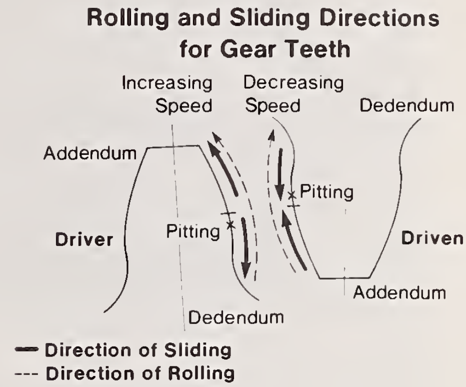


FIGURE 7. *Action and stresses on surface of gear teeth.*

What happens in the case of a gear set? On a driving-tooth face (fig. 7) the point of contact moves outwards, the tooth rolling from the root towards the tip at an increasing speed. On a driven-tooth face, the point of contact moves inwards from the tip to the root at a decreasing speed. The sliding velocity, i.e., the tangential velocity between the teeth, is zero at the pitch circle; on a driving-tooth face the sliding direction is always away from the pitch line; on a driven-tooth face the sliding direction is towards the pitch line. The consequence is that on the addendum, the rolling and sliding motions have the same direction; on the dedendum, the directions of rolling and sliding motion are opposed to each other. This would suggest that the place pitting failures would initiate on gear teeth is just on the dedendum side of the pitch line. Figure 8 shows a typical V-shaped surface pit on a gear tooth and a cross-sectional view (fig. 9) shows that it progresses under the surface at an oblique angle in the direction of rolling. In other words, the V-shaped surface pit points in the direction of sliding and in the opposite direction to the rolling direction. A typical surface pit on a test roller which failed in an industry sponsored program at IIT Research Institute is shown in figure 10. Figures 11, 12, and 13 show how the pits develop on the surface of a roller sample in a program at IITRI.

The next thing to consider are the factors that affect the development of the surface pits. The results from an industry sponsored program conducted at IITRI entitled "Surface Fatigue of Gear Steels" will be discussed. The purpose of the investigation was to determine the effect of metallurgical treatments and processing of gears on the surface durability of carburized and hardened gear steels under the high contact loads. The pitting fatigue tests were carried out on cylindrical specimens in laboratory fatigue machines of the Caterpillar design. The geometry of the contacting specimen surfaces and the testing parameters were carefully designed to simulate conditions existing between meshing involute gear teeth, namely, loads of 275 ksi to 420 ksi and a relative sliding velocity of 24 in/s.

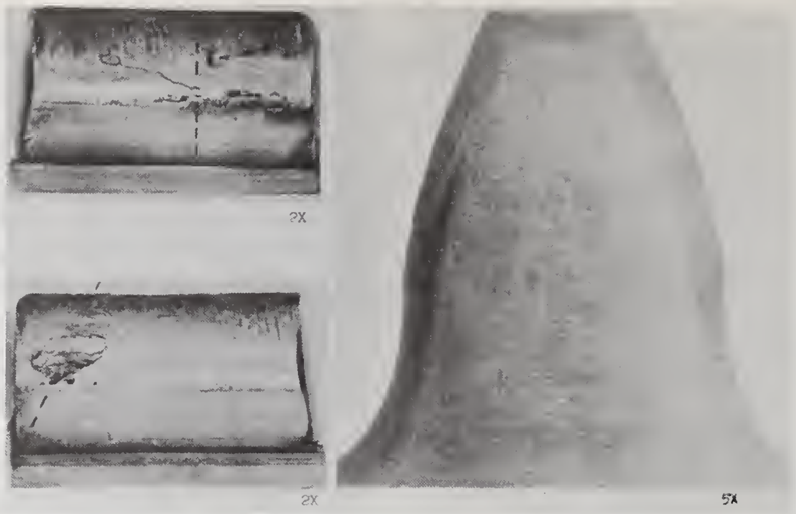


Figure 8. *V-shaped surface pit on gear tooth showing angle of crack propagation.*
This figure reduced 58 percent in reproduction.

Figure 9. *Micrograph of typical cracks originating at surface of gear teeth.*

FIGURE 10. *Typical surface pit on roller.* × 2.

FIGURE 12. *Micrograph of surface of test specimen no 209, 13.6 megacycles at 375 ksi, showing advanced stage in development of macro-pit.*
Apex has been spalled away, and crack has encircled a larger area. × 40.

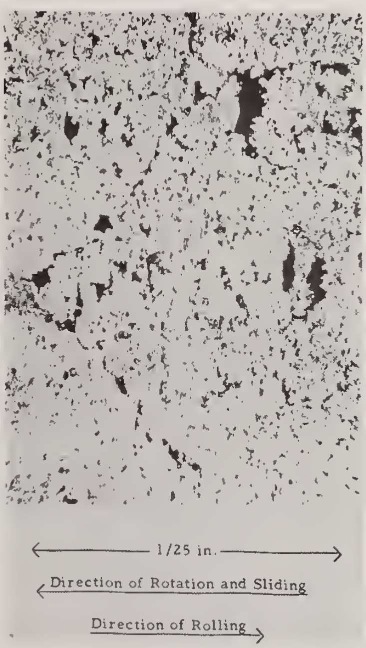


FIGURE 11. *Micrograph of surface of test specimen no. 209, 13.6 megacycles at 375 ksi, showing incipient development of surface macro-pit.*
Apex of pit, at bottom of photo, encircled by crack. × 40.

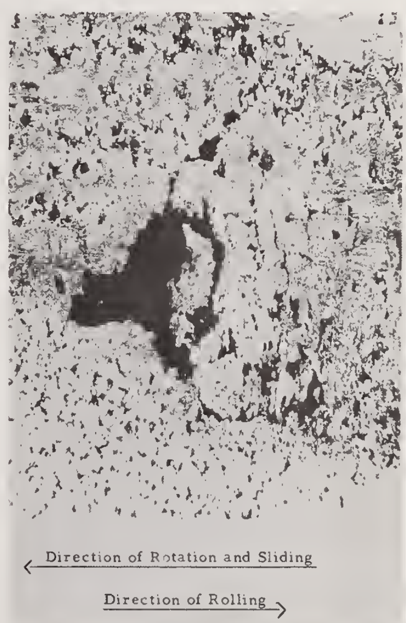


FIGURE 13. *Micrograph of surface of test specimen no. 209, 13.6 megacycles at 375 ksi, showing enlargement of damage area in the development of a macro-pit.* × 40.

The results can be summarized as follows:

No Significant Effect

(1) Grade of Steel
No significant differences in pitting fatigue durability have been found among seven different grades of alloy steel commonly used for gearing. These results suggest that the chemical composition of the steel probably does not have a strong effect on pitting fatigue provided that the surface microstructure of the steel is free from soft non-martensitic transformation products.

(2) Heat Treatment
Contact fatigue tests conducted on a number of different steels hardened after carburizing by direct quenching and by reheat treatment revealed little, if any, difference in pitting resistance for specimens heat treated by both methods.

(3) Retained Austenite
Widely varying amounts of retained austenite in the range of 5 percent to 60 percent have shown no difference in pitting fatigue properties. One of the problems associated with the testing of specimens having large amounts of retained austenite is the plastic deformation that occurs on the surface due to the relatively soft microstructure.

(4) Excess Carbides
In several steels that were carburized to quite high carbon levels there was a tendency for massive carbide formation at the carburizing temperature or during furnace cooling from 1700 to the quenching temperature of 1550 °F. These colonies of carbides persisted during direct quenching and reheat treatment, but there was no strong evidence in the test results to show that they influenced the pitting fatigue resistance.

Significant Effect

(1) Carbon Content of the Carburized Case
The pitting resistance of 8620 steel increases significantly when the surface carbon content is raised from 0.7 percent to about 0.95 percent. but no further gain in fatigue properties is achieved with higher carbon contents. This increase is thought to be due to two factors: (1) the hardenability effect of carbon in minimizing pearlite transformation in the surface, and (2) a strengthening of the martensitic-austenitic matrix with increasing carbon.

(2) Hardness
The pitting resistance of carburized steels falls off rapidly below 58 R_c.

(3) Oxidized Grain Boundaries
The pitting resistance of the as-heat-treated surfaces containing grain boundary oxides was significantly higher than surfaces with the surface layer ground off (fig. 14). This effect has been a baffling condition and seems to defy explanation. However, the following explanation is proposed. Surfaces with grain boundary oxides tend to fail in a manner shown in figures 11, 12 and 13. The shallow surface cracks tend to be oriented transverse to the rolling and sliding direction and therefore act as little oil reserves which provide more lubricant to the surface than if the surface were smooth or ground in the circumferential direction.

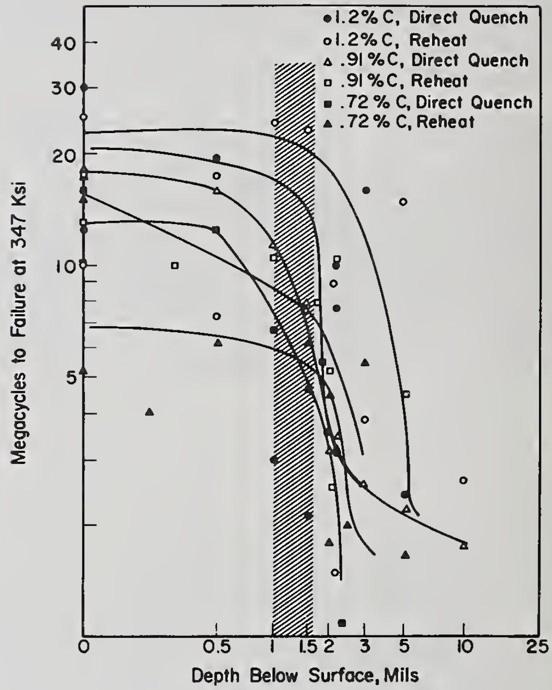


FIGURE 14. *Fatigue life as a function of the amount of material removed from the surface before testing.*
Carburized SAE 8620 steel tested at 347 ksi.

As happens quite frequently in laboratory work, the variables that start out being the most suspected in influencing pitting resistance exert minor effects and other variables that weren't the main object of the investigation have substantial influence on the results.

Such was the case for the lubricant and surface roughness. In each case they were supposed to be held constant and therefore they should not influence the results.

Lubricant

Although lubrication was not intended to be a variable in this study, it inadvertently became one through the process of aging of the oil. The lubricant used for the tests was an SAE 80 containing sulfur and phosphorus E. P. additives. If a batch of this oil was used for a long period of time, it was noted that the specimens tested in the oil experienced a gradual but significant increase in pitting fatigue life. The reason for this increase in fatigue life with aging of the oil must be associated with a change in the concentration of the E. P. agents.

Other tests run on the same base stock with and without the E. P. additive showed significant interactions between the type of oil and the surface of the test specimen, and significantly influenced fatigue life. This work, together with that of other investigators, has demonstrated the importance of the role of lubricants in pitting fatigue. There appears to be a stress corrosion effect acting at the bottom of the crack.

6. Surface Roughness

During the progress of this investigation the exact technique for preparing the samples was inadvertently changed because it was necessary to obtain the rollers from a different machining source. The changes in surface finish that occurred caused the results to be radically altered. This led to specific experiments to determine the effect of surface finish on surface pitting.

The surface finishes of the rollers were as follows:

Test Roller—1 in diameter
- (1) As heat treated 10 μin RMS
- (2) Ground (2 mils removed) 10μin RMS (fig. 16)
- (3) Ground and Polished—2 μin RMS—Axially (fig. 17)
- (4) Ground and Honed—2 μin RMS—Circumferentially

Load Roller—5 in diameter
- A. Ground—10/15 μin RMS
- B. Polished—5/9 μin RMS

Lubricants

EP Oil—8 percent SP Additives

Proficorder traces of the surface of the best rollers are shown in figure 15. The data are summarized in figure 18.

Curves (1) and (2) are for the polished load roller and the load to pitting is substantially higher than in the case of the ground load roller, curves (4) and (5).

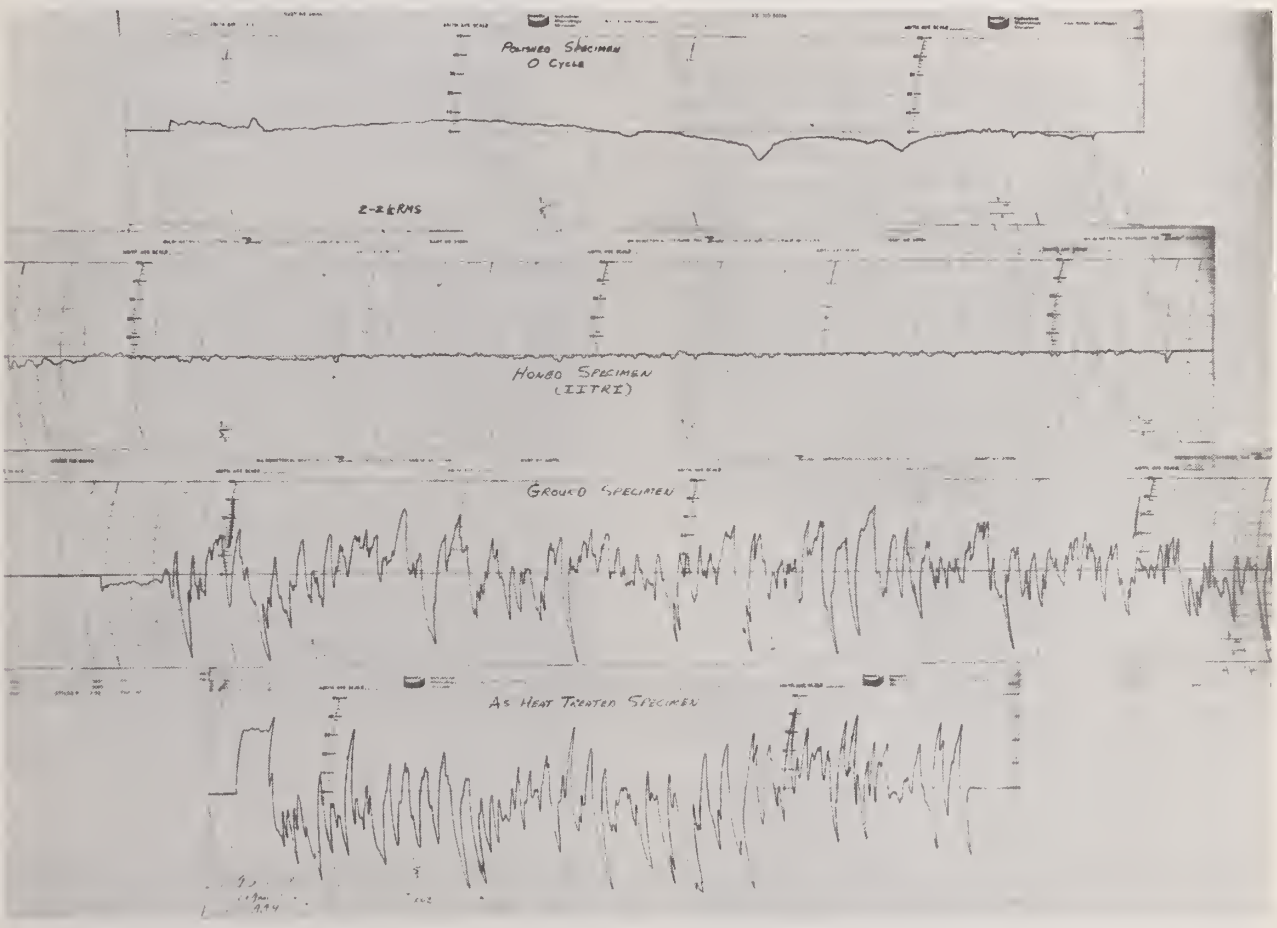


FIGURE 15. *Proficorder traces showing surface roughness of polished, honed, ground, and as-heat treated specimens.*

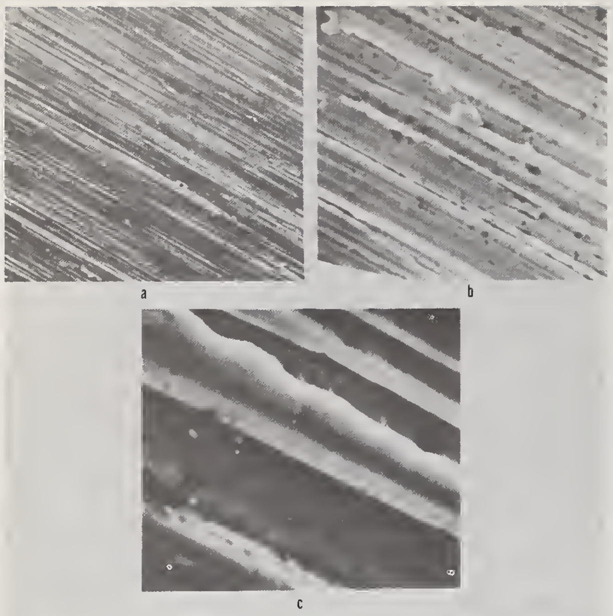


FIGURE 16. *SEM of ground specimen at (a) × 515, (b) × 1030, and (c) × 2575.*

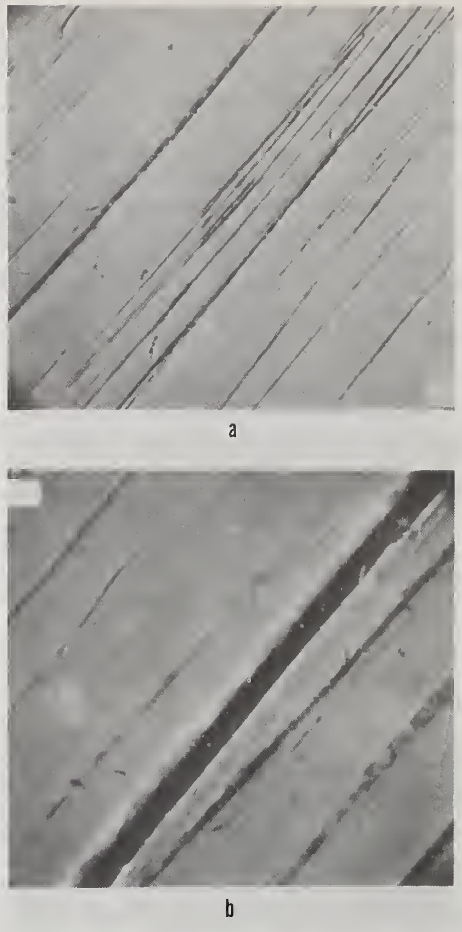


FIGURE 17. *SEM of polished specimen at (a) × 450 and (b) × 1350.*

The surface condition of the test roller has a lesser effect on load to pitting. In the case of the ground load rollers, (4) and (5), there is practically no difference due to surface finish of the test roller. In the case of the polished load roller, (1) and (2), the surface finish of the test roller has an effect—curve (2) for the ground test roller is appreciably below that of the polished test roller, curve (1). When the roughness of the loading roller (positive slip element) was reduced from about 15 μin to about 8μin RMS and the test specimen (negative slip element) was reduced from about 10 μin to 2 μin RMS, there was a gain of approximately 150 percent in load-carrying capacity.

From this figure it can be concluded that the roughness of the surfaces, and particularly the surface of the load roller which is in positive sliding, has a dominant effect on pitting fatigue.

Specimens with as-heat-treated surfaces, curve (3), have significantly better pitting resistance than ground or polished test specimens when ground load rollers are used in the tests; however, they are not as good as specimens tested against polished rollers.

The reason for these differences in service life are quite apparent in figure 19. These show three combinations of surface finish on the test specimen and load roller after 100 000 cycles at 363 ksi in non EP oil.

The ground roller (fig. 19a) against the polished test specimen showed a very distinct contact track and evidence of heavy metallic contact between the two rollers. The polished roller (fig. 19b) against the ground test specimen showed a very faint but full contact track with considerably less severe metallic contact. The polished roller (fig. 19c) against the polished test specimen showed only a random array of small contact track.

In figure 20a at 100X the contact track is quite smooth as the result of the polishing action of the ground roller eliminating all transverse polishing marks but there are many transverse surface cracks even after a short testing period of 100 000 cycles. This roller failed at 7½ million cycles. (See fig. 20b.) At 100X very little polishing has occurred from the polished load roller and only a few, very small, cracks are present (fig. 20c). At 100X the transverse polishing lines are still present and the random spots of metallic contact seem to have occurred between high points on the load roller and the tops of the polishing ridges in the test specimen. Total life for this roller was 32 million cycles with no pitting.

This series of tests can be summarized as follows:

Surface finish measurements show that the change in surface finish of the specimen, which is an indication of the amount of metallic contact, is a function of the roughness of the loading roller (the surface in positive

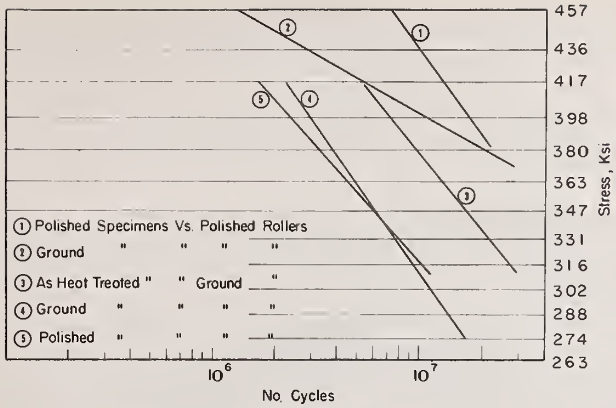


FIGURE 18. *Composite of S–N curves for the various surface finish combinations for tests conducted in the E. P. lubricant.*

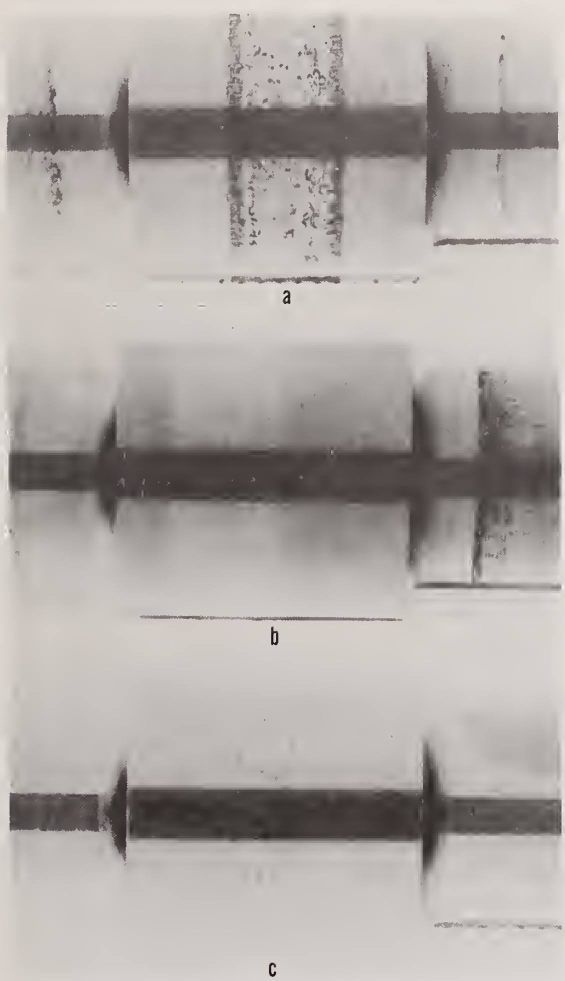


FIGURE 19. *Illustration of severity of contact between various specimen-roller combinations after 10^5 cycles of fatigue at 363 ksi in non-E. P. oil.*

(a) Polished specimen against ground roller.
(b) Ground specimen against polished roller.
(c) Polished specimen against polished roller.

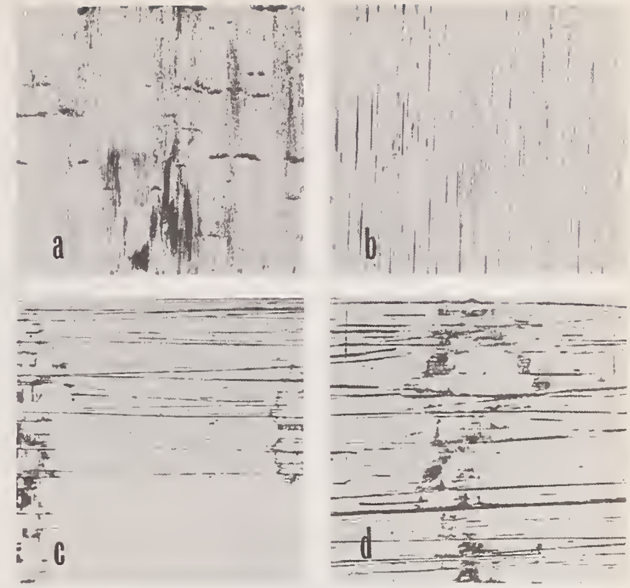


FIGURE 20. *Higher magnification illustrations of severity of damage in contact tracks for same combinations shown in figure 19.*

(a) Polished specimen against ground roller. × 60.
(b) Ground specimen against polished roller. × 60.
(c) Polished specimen against polished roller. × 30.
(d) Polished specimen against polished roller. × 60.

sliding). When a rough loading roller is used, an initially rough test specimen will tend to become smoother during the test and an initially smooth test specimen will become rougher, and both will approach about the same degree of surface roughness before failure occurs. When a smoother loading roller is used, there is only a small change in the surface roughness of a ground test specimen and practically none in a polished specimen.

This program has been continued at IIT Research Institute with a change in direction under the sponsorship of the ASME Research Committee on Lubrication. While quantitative information is not available at this time, a general indication of the results can be given:

(1) Fatigue life is lowered by an increase in the amount of slip between the rollers.
(2) Additives in the oil will lower the fatigue life. This has been shown in other investigations and tends to suggest that there is a stress corrosion effect at the root of the surface cracks which would increase the rate of crack growth.
(3) Polished surfaces have better fatigue lives than ground surfaces.
(4) High viscosity oils and higher roller speeds tend to improve fatigue life. This is a reflection of better lubrication between the rollers. As the speed of the rollers is increased there is less time to squeeze out the oil between the rollers and as the viscosity of the oil increases, its resistance to being squeezed out increases.
(5) Surface temperature adversely affects fatigue life—this reduces oil viscosity and enables it to be squeezed out faster.

7. References

[1] Wulpi, Donald J., How components fail, Am. Soc. Met., p. 48–51.
[2] Shipley, Eugene E., Failure modes in gears, p. 107, Gear Manufacture and Performance, Materials/Metal Working Technology Series, ASM.
[3] Lorick, Robert R., Lubricating gears, Machine Design, July 9, 1970, p. 108.
[4] Hurricks, P. L., Some metallurgical factors controlling the adhesive and abrasive wear resistance of steels (A Review), Swansea Tribology Entr., G. B. Wear Volume **26**, No. 3, (December, 1973), p. 285.
[5] Ku, P. M., Private Communication, Southwest Research Institute, San Antonio, Texas.
[6] Buckingham, Earle, Prof. M. E., M.I.T., Surface fatigue of plastic materials, Trans. of ASME, May, 1944, p. 297–310, Discussion by G. B. Warren, Design Eng., Turbine Engine., Div. G.E.
[7] Sheehan, J. P., and Homes, M. A. H., The effect of case carbon content and heat treatment on the pitting fatigue of 8620 Steel, SAE Preprint No. 720268.

8. Discussion

A. J. Babecki, NASA, Goddard Space Flight Center: I think you stated that in your roller tests where you had differential speeds between 2 rollers you found that the one that moved the slower was the one that got the pits. And you said that hardness didn't seem to make any difference if the roller was in compression. Yet, later on in your talk you said that hardness is one of the significant factors.

R. A. Wilde: Yes, but that would be only for the one of negative sliding.

R. Lenich, Caterpillar Tractor Company: Do you have any comments on a solid additive like molybdenum disulfide on scuffing?

R. A. Wilde: No, I have never done any work in that area. Have you had any kind of experience with it?

R. Lenich: We have done some work on oscillating shafts, but little on gears.

A. Beerbower, Esso Research & Engineering Company: In reference to the smoothness of the roller surfaces, you said that too smooth gets rougher and too rough gets smoother. I had postulated this a couple of years ago. Is there any way of predicting what the equilibrium will be? It's well known that used oils have populations of particles that are 50 to 100 times the diameter, or at least as big in diameter, as the gaps that you are talking about, assuming that you are talking 1 to 1 film thickness to roughness. What happens when one of these particles goes through the gap? In other words, have you any indication as to the effect of oil cleanliness?

R. A. Wilde: In all of our experiments we very carefully filtered out all of the particles possible so that these would not become part of the problem. If there were large metal particles going through the gap, it would be the same as having high asperities on either side and it would lead to adhesive wear.

J. E. Stern, NASA, Goddard Space Flight Center: Has any attempt been made to determine the effect of the concentration of the additive in the oil on the degree of pitting of gears and rollers? In a recent study of the effects of additives in oils on bearings, it was found that there was an optimum additive concentration for a particular case.

R. A. Wilde: We have done some work in this area, but all we have is general quantitative data. We don't have good data in that area.

C. E. Vest, NASA, Goddard Space Flight Center: Were all your components carburized?

R. A. Wilde: Yes, they were carburized by very careful practice because we recognized this to be very important.

C. E. Vest: What did the microstructure look like when the ammonia was added to the carburized structure?

R. A. Wilde: Usually there isn't enough added to produce a real strong iron nitride structure. The purpose of this is to prevent deterioration of the surface as a result of grain boundary oxidation and the reduction in hardenability of the surface. There will be a very minor iron nitride structure at the surface.

C. E. Vest: Is the structure changed when the 0.002 inch was machined off?

R. A. Wilde: No, the 0.002 inch was taken off simply to get below the grain boundary oxides and the lower products of transformation that might be at the surface.

V. C. Westcott, Trans-Sonics, Inc.: The particles from gear wear that we are seeing on Ferrograms seem to confirm the mechanisms that you have described here. In particular the particles tend to be chunky and have dimensions of thickness, width, and length, which are comparable. Have you photographed any micropits on the surfaces of these gears at a magnification of the order of 1000 times?

R. A. Wilde: No.

J. V. Deller, Naval Ship Engineering Center: Mr. Beerbower commented on the surface roughness of the test roller and the load roller. Can you elaborate on that?

R. A. Wilde: Apparently the load roller that's in positive sliding acts much more heavily on the test roller itself because the load roller is in positive sliding and the test roller is in negative sliding. You tend to take more surface off the test roller by the action of the load roller. The load roller itself does not seem to be changed much. Apparently this is the nature of the relationship between positive sliding and negative sliding. The positive sliding surface tends to retain its surface and the one in negative sliding is either smoothed out a little bit if it's real rough, or is roughed up if it's real smooth to start with.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Bearing Damage Analysis

R. L. Widner and W. E. Littmann

The Timken Company
Canton, Ohio 48076

Failure analysis of machine components is described with the major emphasis on tapered roller bearings and their interactions with other components of a system. Included in the discussion are the effects of loading, deflections, misalignment, lubrication, temperature, debris, water and electrical current on the type of damage and life of tapered roller bearings.

Key words: Bearing failure; bearing lubrication; bearing loading; bearing misalignment; bearing temperature; mechanical failure.

1. Introduction

Most mechanical systems have a complex combination of components that transmit forces across concentrated contacts. These contacts vary from hydrodynamic film junctions to metal to metal junctions and from essentially pure rolling to pure sliding motion. Such components are journal bearings, gears, splines, rolling element bearings, clutches, cams, and seals.

As a result of repeated loads of varying magnitude in service, high stresses in these concentrated contacts give rise to changes in the contact surfaces and in the region just beneath the surface. When these changes are of sufficient magnitude as to adversely influence the subsequent performance of the contact surfaces, one says they are damaged. When the damage is sufficient to threaten the essential function of the components, one says they are failed. Continued operation would threaten the performance of the entire mechanical system.

2. Control of Bearing Life

In a given system the ultimate life of a rolling element bearing is determined by factors controlled by (1) the manufacturer of the bearing, by (2) the designer and builder of the machine in which the bearing is applied, and (3) the user of the finished machine. Knowledge of these factors can be coupled with an effective method of bearing damage analysis to identify and reduce some of the causes for mechanical system failure.

3. Role of the Manufacturer

The manufacturer has the control of the internal design, materials, processing and inspection used in making the bearings. He also establishes load-life ratings for each bearing and the revelant instructions and specifications for application, assembly, and adjustment of bearings.

3.1. Internal Design

The load capacity, torque, heat generation, and vibration characteristics of a given bearing are characteristics largely controlled by the producer in the design and manufacture of bearings. The internal and external geometry of bearing components, the materials from which they are made, and surface integrity of critical surfaces are the essential elements used by the producer to control the performance capability of bearings.

3.2. Material

Material requirements for rolling element bearings have been comprehensively described in relation to usual and unusual operating environments. They are briefly summarized as follows:

(1) Maximum hardness at the operating temperature consistent with the fracture toughness required for a given design and application environment.
(2) Corrosion resistance sufficient to maintain surface integrity under prevailing environmental conditions.
(3) A low level of integral discontinuities such as nonmetallic inclusions, porosity, or microcracks.

These requirements can be met by a variety of alloy steels of sufficient carbon content and hardenability to develop needed properties with appropriate processing. Two major types of steels are used, low carbon bearing steels which are case carburized and hardened or high carbon bearing steels which are through hardened or case hardened by surface heating methods. Bearings to be applied in high temperature environments require materials that retain their hardness upon sustained exposure to expected temperatures. A summary of suitable steels is given in reference [1].[1]

[1] Figures in brackets indicate the literature references at the end of this presentation.

3.3. Processing

Bearing quality steels may be manufactured by several steel making processes as summarized in table 1. The frequency-severity distribution of oxide inclusions resulting from the method of steel making used to produce the bearing steel largely governs the load capacity of bearing life as limited by the inclusion origin mode of contact fatigue.

TABLE 1. *Processing methods for manufacture of bearings*

Melting and Casting
Electric Furnace Melting and Casting
Consumable Electrode Vacuum Melting and Casting
Electroslag Remelting and Casting

Plastic Deformation
Hot Rolling of Bars and Billets
Hot Piercing, Rolling, or Extension of Seamless Tubing
Hot Forging of Rings and Rolling Elements
Cold Forging of Rings and Rolling Elements

Rough Machining
Conventional Chip Removal Machinery

Heat Treatment
Carburizing, Hardening, Tempering (High Carbon Steel)
Hardening, Tempering (High Carbon Steel)
Flame or Induction Surface Hardening (High Carbon Steel)

Finish Machining
Grinding
Honing
Polishing

Also, table 1 shows the various ways in which the steel composition and processing can be varied to produce the required material characteristics in the finished bearing. This summary describes the methods which are most widely used for processing of bearings. A comprehensive summary of the influence of materials and processing is given in references [1] and [2].

3.4. Inspection

Inspection assures that parts meet the design, material, and processing specifications.

3.5 Establishment of Load-Life Rating

Using the relation of bearing internal geometry and load to contact stresses, a load rating and methods for computing expected fatigue life are supplied by the bearing manufacturer. The factor representing material fatigue strength is determined from bearing life tests under simulated service conditions.

For one tapered roller bearing producer [3], an audit sample of bearings from world-wide operations are life tested under a set of reference operating conditions to monitor the life rating of "standard" bearings. Additional testing has been used to determine the influence of various environmental factors, such as temperature, speed, misalignment [4], lubricant chemisty, lubricant film thickness [5, 6, 7, and 8], etc., on bearing life. Adjustment factors derived from such tests can be used to modify the expected life in applications that are known to place bearings in an environment significantly different from the reference operating conditions.

4. The Role of the Designer and Builder

A bearing is just one of many components that a machine designer and builder incorporates into a useful piece of equipment. Therefore, the manner in which a bearing is applied to a machine application and its interaction with other components become significant factors which influence the life of the bearing. Important considerations for the designer are the load, speed, misalignment, and/or deflection under service loads, bearing adjustment, heat generation [9] and heat flow within the system, the type of lubrication system [10], lubricant composition, and ambient temperature. In addition, he must consider stray electrical current, and related requirements for shunting, debris, and the filter system to control debris recirculation, water entry, and sealing problems.

If the designer can accurately predict the items above with the help of the bearing application engineer, then there can be greater confidence in the estimate of the bearing life in a given system.

The builder must consider correct bearing housing alignment and the accurate machining and/or grinding of the shafts and housings for proper shapes and fits. Before assembly, adequate cleaning is needed to assure removal of machining chips, abrasive debris from sand castings and all foreign materials which could contaminate the lubricant. Finally, the careful assembly of bearings with the other components must be made to avoid damage to critical surfaces and assure proper bearing adjustment settings [11, 12]. It is important that the designer and the builder jointly understand how the cold setting of the bearing lateral (internal clearance) is affected by heat generation in the system during start-up and the run-in of gears and bearings. The internal load distribution among the rollers in a tapered roller bearing is adversely affected by excessive lateral. On the other hand, heat generation and contact stresses at raceway and rib-roller end contacts increase rapidly if lateral goes to zero and into a preload condition. This is critical for new bearings and at start-up when heat flow has not established thermal equilibrium in the system.

5. The Role of the User

If the manufacturer, designer, and builder have done their work correctly, then the life of the individual components and the machine will be as-predicted unless something is altered by the user. The user can modify the expected life of the bearings in a given system by: (1) loads and speeds beyond the design ranges, (2) improper maintenance, or (3) by using the machine in an environment in which it was not designed to operate.

Machine life and component life are statistical in nature. The contact fatigue life of individual bearings has been found to be well represented by a Weibull distribution. Early incidence of fatigue damage is usually the result of a bearing from the portion of the population having the lowest intrinsic fatigue life being coincident with problems in machine design, building, and/or use. In other words, the lowest life of a machine or component is usually the result of a chance coincidence of factors which were caused inadvertently in the manufacture of the component, by the designer or builder of the machine, and by the user of the machine.

6. Modes of Damage

Whenever forces are transmitted across concentrated contacts as between the elements of a tapered roller bearing there will be an eventual breakdown of the surfaces by wear, fatigue, or some other process depending on the load, speed, temperature, lubrication, and other environmental factors. Over the years, there have been several papers [13, 14, 15, 16, 17, 18, 19, and 20] dealing with the classification of the various types of damage that have been observed. The classification schemes have been based on (1) location of the damage (surface or subsurface), (2) type of stress concentration (surface flaws or nonmetallic inclusions), (3) type of contaminant (debris or water in the lubricant), (4) type of propagation (hydraulic pressure or transverse cracking), etc.

The damage classification scheme in table 2 is proposed. The scheme varies from that previously used [18] as a result of discussion at a Wear Mode Terminology Meeting in conjunction with an oil analysis program, currently underway [21].

Table 2. *Damage classification*

- Wear
  - Adhesive Wear
    - "Normal," Mild, Smooth Wear
    - Scuffing, Scoring, Smearing, Galling, Seizure
  - Abrasive Wear
  - Corrosion Wear
  - Fretting Damage
- Plastic Flow
  - Debris Denting
  - Brinelling
  - Burn-up
- Electrical Damage
  - Fluting
  - Arcing
- Fatigue
  - Contact Fatigue—Subsurface Origin
    - Inclusion Origin
    - Geometric Stress Concentration
    - Subcase Fatigue
  - Contact Fatigue—Surface Origin
    - Point Surface Origin
    - Microspalling (Peeling)
- Section Fracture

A classification scheme with pictorial examples of the various types of damage is the most important tool at the disposal of someone engaged in bearing damage analysis. It is valuable because it provides a ready reference to past experience and it enables the investigator to communicate effectively with those who will use the results of his investigation.

The remainder of this paper will be concerned with a description of the various modes of bearing damage using examples of these modes from tapered roller bearings from service and laboratory applications. Although the examples given are specific to tapered roller bearings, most of them are applicable to other types of rolling contact bearings, gears and cams.

7. Wear

Although contact fatigue, especially surface fatigue, is considered by some to be a form of wear [22], the authors choose to separate the contact fatigue and wear modes of damage. There are two basic mechanisms of wear [22] in rolling element bearings: adhesive wear and abrasive wear. Corrosive wear also may be found under some circumstances in chemically aggressive environments. Cavitation wear has been hypothesized as a possible mechanism for some rolling contact surface damage. As with the competitive modes of contact fatigue, any one or all of them may be acting in a given application. The cage (retainer), roller ends, and the large end rib (thrust rib) of a tapered roller bearing are the sites of most wear in tapered roller bearings. However, wear can and does occur on the rolling contact surfaces as well. The small amounts of wear normally encountered do not interfere with bearing operation. The amount of wear that can be tolerated in a given bearing is dependent upon the particular application. Excessive adhesive wear is an indication of inadequate lubrication. Abrasive wear occurs when an abrasive material contaminates the lubricant.

7.1. Adhesive Wear

The most serious adhesive wear damage in tapered roller bearings has been called anything from scuffing, scoring, seizing, to galling. The terms used for adhesive wear are somewhat indicative of the degree of damage. Scuffing is least severe; scoring is a similar stage of damage often applied to gears; seizure and galling indicate more extensive plastic flow and metal transfer. Real surface contact is limited to asperity interactions under effective lubrication. When operating conditions and surface characteristics permit metal to metal contact at asperity contacts, a welded junction is formed. With relative motion between the contacting surfaces, junction growth occurs by plastic deformation. Fracture ultimately takes place within one of the two surfaces or at the original interface. The resulting metal transfer and plastic deformations produce loose wear debris and modify the original surface texture and microstructure of the contact surfaces. At high loads and speeds, the local temperature rise can be sufficient to cause tempering, rehardening and melting of the surface. Mechanical activation, together with the high local temperatures, cause decomposition of the lubricant. Chemical reactions occur between the surfaces and the lubricant and with the gaseous environment. The resulting modifications of surface texture, microstructure, and composition take place at astonishingly rapid rates and have persistent effects on the subsequent performance of the contact surfaces.

Because scuffing often results from transient conditions of inadequate lubrication, the loss of material by adhesive wear is often of negligible importance compared to the changes in surface texture, composition, and microstructure. These changes sometimes cause one of the contact surfaces to become harder than the other. Since this change is often coupled with greater surface roughness due to scuffing, the wear rate in subsequent operations is greatly increased, even if adequate lubrication prevents further scuffing damage.

The degree of damage caused by adhesive wear is dependent upon the material properties, surface texture, EHD film thickness, surface films, lubricant additives, rolling and sliding velocities, surface and lubricant temperatures, speed, and load.

In tapered roller bearings operating under so-called "normal" conditions essentially no adhesive wear occurs because the surfaces of the rolling contact and the rolling sliding contacts (rib-roller ends) are well lubricated and separated by an EHD and/or boundary lubricant film.

Under more severe operating conditions or when the lubricant supply is interrupted in a "normal" application the adhesive mode of wear can usually be observed on the rib and roller end surfaces and the cage pockets as shown in figures 1 and 2. If effective lubrication is restored or the severe condition is intermittent, the surfaces may have a chance to heal themselves. However,

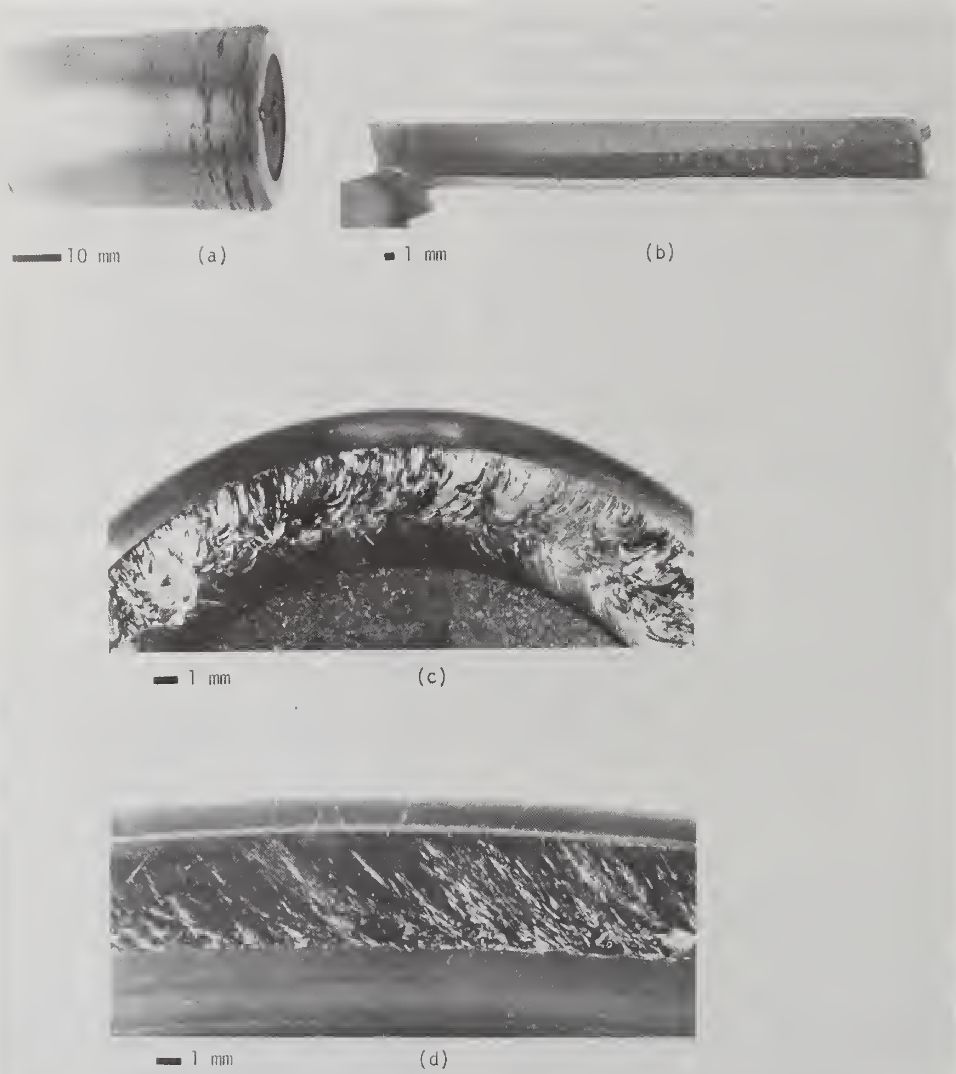


FIGURE 1. *Examples of adhesive type wear caused by inadequate lubrication.*
(a) Metal pickup on a roller O. D. from the sliding contact on the cage shown in (b). (c) The end of the same roller showing the scoring damage from the rolling-sliding contact of the L. E. rib or thrust rib shown in (d).

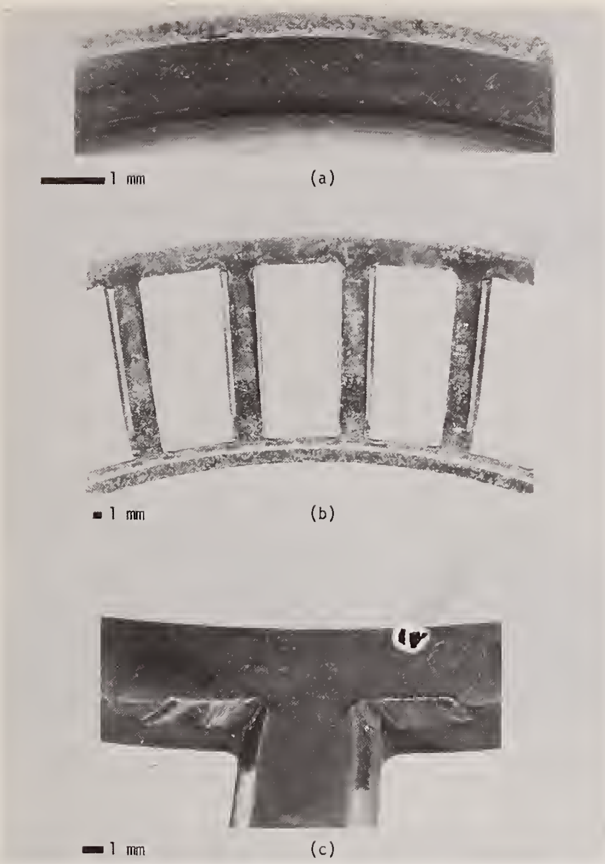


FIGURE 2. *Scuffing (a mild form of adhesive wear) damage on a cone L. E. rib or thrust rib in (a).*
In (b) and (c) the cage or retainer shows abrasive type wear where the hard surface of tbe rollers cut the softer cage material.

if effective lubrication is not restored, the adhesive wear can propagate from a mild form to a gross form (burn-up). Progressive scuffing damage can generate wear debris which will cause abrasive wear, similar to that in figure 3, when the lubricant is restored to the contact.

Transfer of cage material, by adhesive wear, to the roller body produces asperities which can lead to micro-spalling contact fatigue of the cone or cup raceway. An example is shown in figure 1a.

7.2. Abrasive Wear

Abrasive or cutting wear occurs when a "hard" surface abrades or cuts another "softer" surface. Another type of abrasive wear results when a hard third-particle between two surfaces abrades or cuts one or both surfaces as shown in figure 3.

Cage pocket wear, shown in figure 2b and 2c in tapered roller bearings, is mostly abrasive type wear. Wear at the rib-roller end under "normal" conditions can also be abrasive wear of a mild form. If conditions are severe enough that adhesive wear (scuffing) occurs at the rib-roller end contact, or on the raceway in high speed applications, and conditions may improve somewhat, abrasive wear of either the roller ends or the rib occurs because one of the altered surfaces is now harder or the other is softer. Examples of adhesive followed by abrasive wear are shown in figures 4, 5, and 6.

Third-particle wear occurs when debris from wear or fatigue damage on other components, from external sources penetrating seals, or from improperly cleaned components, circulates through the bearing in the lubricant.

Cornish has described the abrasive wear damage in tapered roller bearings resulting from gear wear debris and from other abrasive contaminants in automotive

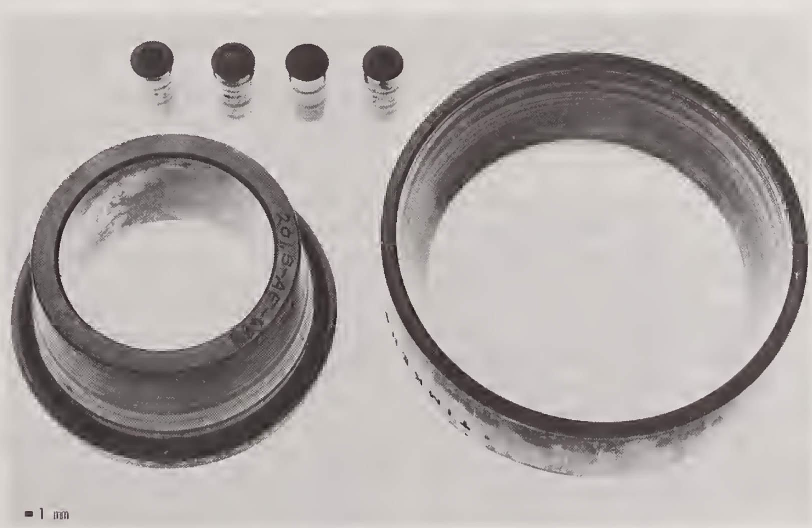


FIGURE 3. *Abrasive wear caused by third particles.*
Natural diamond dust (0-5 μm size) was deliberately introduced into the lubricant in a controlled test. Damage consisted of grooving of the raceway contacts and excessive wear of tbe rib and roller ends.

FIGURE 4. *Adhesive wear caused by high speed skidding on a cone race in (a) and (c).*

The roller in (b) first experienced adhesive wear and subsequently abrasive wear when the very hard raised areas on the cone caused cutting wear on the rollers as shown in figure 5. The arrows indicate the direction of rotation.

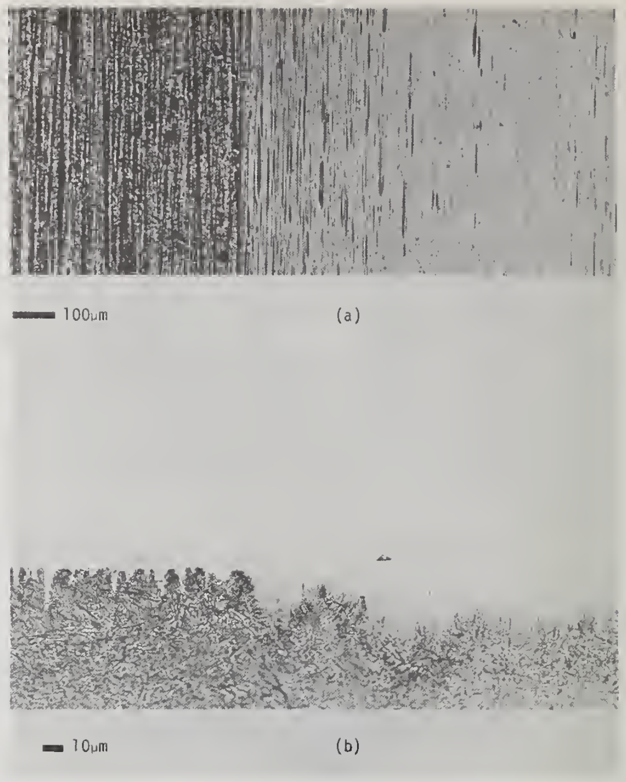


FIGURE 5. *The roller from figure 4 showing the surface at high magnification in (a) with the "normal" wear at left and the abrasive wear at the right.*

(b) is a nickel plated chord section of the roller showing the depth of wear (1.6 µm) and the rehardening and tempering of the microstructure accompanying the damage. The vertical magnification of the chord is about 10 times the horizontal magnification.

rear axle assemblies [23, 24]. An extreme example of abrasive third-particle wear is shown in Figure 3 where the bearing lubricant was deliberately contaminated with diamond dust [25]. The obvious preventive measures against abrasive wear are prevention of gear wear, cleaning of bearing housing and enclosures, effective sealing and adequate filtration of circulating lubricants.

7.3. Corrosion

Corrosion often occurs on bearing components as a result of improper handling or storage of bearings after the protective film of lubricant is removed, as in Figure 7a. Fingerprint deposits are a common cause of corrosion when bearings are handled after removal of oil films. Subsequent bearing operation usually results in removal of the brittle corrosion product by fragmentation. The fine abrasive debris can then cause abrasive wear, sometimes producing a smooth, lapped appearance.

Perhaps the most common contaminant in a bearing environment is water. It can enter through defective or inadequate seals or from condensation under certain conditions. The corrosion that occurs usually forms a black oxide commonly called "water etch." Corrosion usually occurs when the parts are stationary and the lubricant film has been lost, or the grease channeled creating areas with little or no protective film. Under such conditions, water can cause corrosion, especially when "EP" additives are present. In mild cases of "water etch" a trace of black oxide with minimum penetration into the metal occurs. In severe cases, the raceway surface is badly etched away and the bearing runs very rough and noisy. The pitted areas may become enlarged by fatigue spalling. Examples of "water etch" are shown in figures 7 and 8.

Contact fatigue life is substantially decreased by "water etch" corrosion damage. Modification of surface geometry and texture by corrosion causes stress concentrations which lead to early fatigue damage. Hydrogen embrittlement from the corrosion can also contribute to more rapid fatigue spalling at corrosion pitted regions.

7.4. Fretting

Fretting is a form of wear which occurs when two surfaces are subject to slight relative motion under high contact pressures. The fretting process occurs in 3 stages which include adhesive metal transfer, oxidation, and steady-state wear [26].

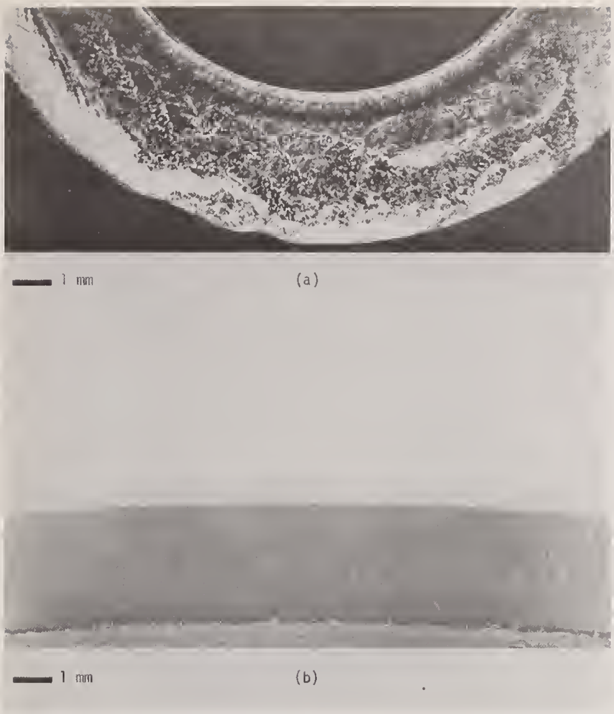


FIGURE 6. *(a) is the spherical end of a roller from the same bearing as in figures 4 and 5 showing the high speed skidding damage (adhesive wear) that occurred between the roller and the L. E. rib.*

In this instance the very hard raised areas caused abrasive wear to the L. E. rib shown in (b) which was first damaged by the adhesive wear process.

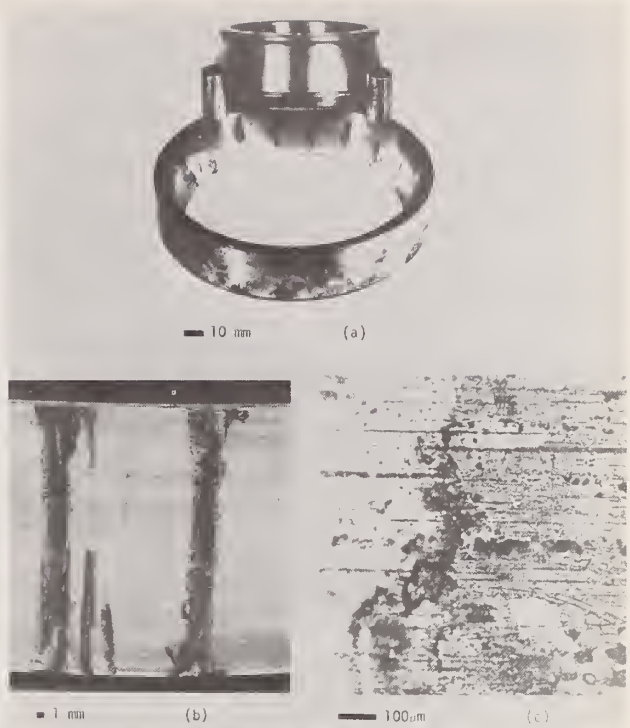


FIGURE 7. *Corrosion or "water etch" on bearing components resulting from improper storage or inadequate seals in the application.*

(a) This type of corrosion damage is caused by the removal of the oil or grease from the bearing and allowing the bearing to lie around unprotected.
(b, c) Corrosion damage typical of that found when seals did not keep water out of the lubricating grease.

In tapered roller bearings and other rolling element bearings, fretting occurs on the raceway of inadequately lubricated bearings that are subjected to vibrations without rotation or to oscillating motion. Fretting damage of this type [27] (also called false brinelling because of the grooved appearance of the raceways) results in noisy bearings that can accelerate fatigue in subsequent operation. Unlike true brinelling, the fretting damage on a raceway will remove material from the surface and alter the surface texture. Examples of fretting of raceway contacts are shown in figures 9 and 10. The fretting damage in figure 10 can be confused with fluting damage caused by electrical current because of a similar appearance. The damage in figure 10 was caused by vibration in the system when the bearing was operating under very thin EHD film conditions.

Fretting damage also occurs on press fitted surfaces and, in fact, is more common there. Fretting occurs between the cone bore and shaft, the cup O.D. and cup seat, and between the bearing faces and spacers or shoulders. Fretting can damage the shaft to the extent that the fit and location of the bearing is destroyed. Or, the fretting may be severe enough to initiate cracks in the shaft and/or the cone which propagate by fatigue and break the shaft or the cone. Fretting of the bearing faces can result in loss of the original bearing lateral adjustment. Examples of fretting of fitted surfaces are shown in figure 11.

8. Plastic Flow

There are several modes of damage that are directly related to the environment in which a bearing is operating and which impose extensive damage. Some are caused by foreign materials entering the bearing system, by overloads, or by failure of the lubrication system.

8.1. Denting

Denting of the contact surfaces is the result of particles much larger than the lubricant film thickness entering the bearing as shown in figure 12. The sources of these particles may be debris from spalling of other bearings or gears in the system, or from an external source through defective seals. The result might be noisy operation or eventual fatigue because of the asperities caused by the dents. Singular debris dents often lead to Point Surface Origin Contact Fatigue damage when EHD film thickness is low.

8.2. Brinelling

Brinelling of bearing races by extremely high static or impact loads is usually seen as grooves located at the roller spacing in tapered roller bearings as shown in

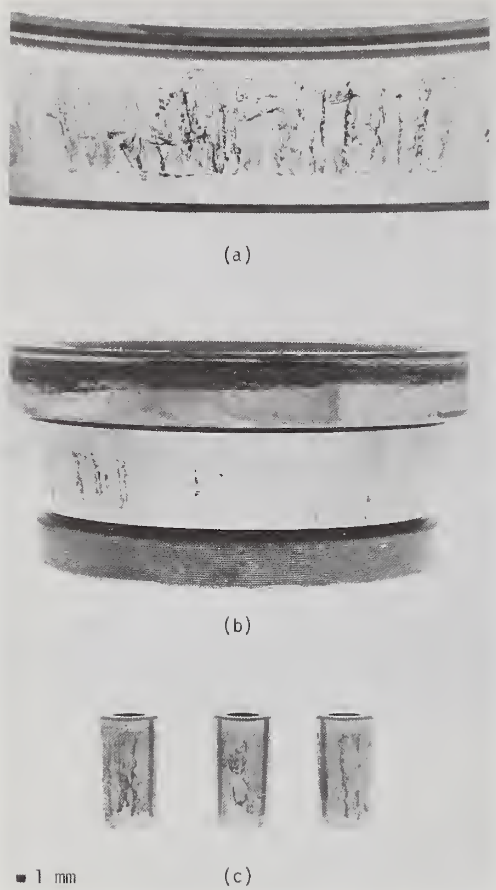


FIGURE 8. *Corrosion or "water etch" caused by excessive water in the lubricant that entered the bearing through defective seals.*
Considerable spalling of the surface can result from damage of this type.

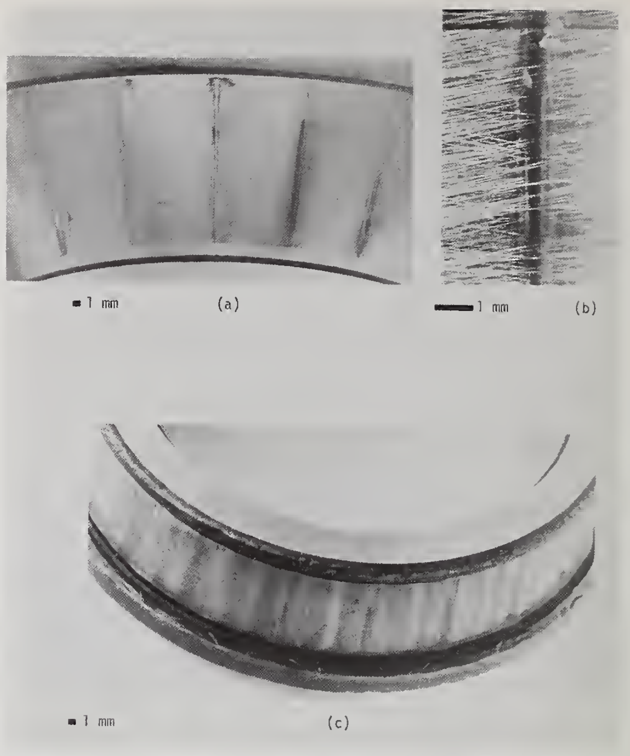


FIGURE 9. *Fretting damage on a cup in (a) and (b) and severe fretting and corrosion on a cone in (c).*
In (c) the fretting damage is evident on the L. E. rib as well as the raceway.

figure 13. True brinelling is characterized by retention of the surface texture in the bottom of the brinell marks because the plastic flow or yielding of the material occurs below the contact surface. A brinelled bearing is usually very noisy and fatigue may initiate from brinell damage after subsequent operation.

8.3. Burn-up

This mode of damage is probably the one most familiar to the average bearing user because it is the single most frequent cause of bearing damage. The causes of the burn-up type damage are (1) an inadequate supply of lubricant because of plugged lubrication lines, inadequate seals, too infrequent maintenance, etc.; (2) overloading the system; (3) improperly adjusted bearings which result in excessive preloading; (4) excessive speed of operation for prevailing load and lubrication. Loss of effective lubrication causes heat generation, and progressive scuffing causes severe metal transfer from one bearing surface to another. The absence of adequate lubrication causes the rate of heat energy input to the contact surfaces to exceed the rate of heat transfer from the region of scuffing damage. The thermal expansion associated with the resulting temperature rise intensifies the contact stresses at the damaged portion of the contact surfaces which causes an accelerating increase in surface temperature. Finally, the material gets hot enough to flow plastically and the bearing geometry is destroyed as shown in figure 14. This mode of damage can be prevented by positive assurance of effective lubrication for the prevalent operating conditions.

9. Electrical Damage

9.1. Fluting

Grounding of electrical current through rolling element bearings causes considerable damage. Fluting, which results from continuous passage of electrical current, whether AC or DC [28], is a washboard effect on the raceway which eventually destroys the raceway geometry and results in generation of noise and fatigue of the damaged surfaces. An example is shown in figure 15.

Figure 10. *Fretting damage on a roller that closely resembles fluting damage caused by electrical current.*
The mechanism suggested is excessive vibration under thin EHD film conditions.

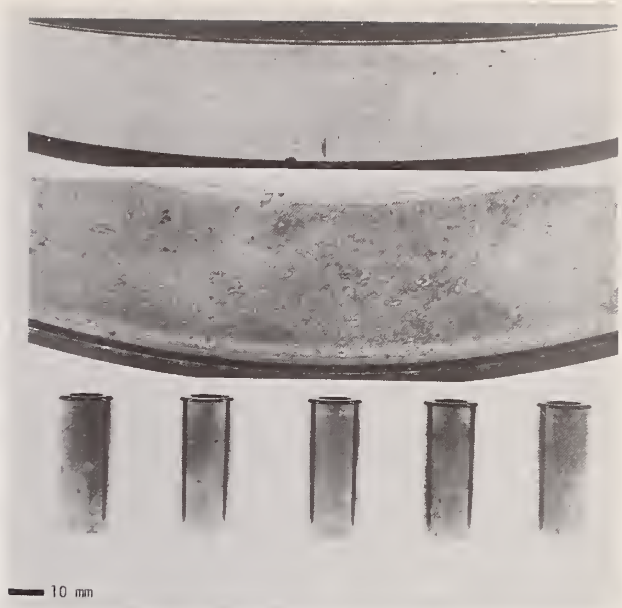


Figure 12. *Denting damage to cone, cup and rollers from debris entering the bearing from spalling in another component, through defective seals, or from improperly cleaned housings.*

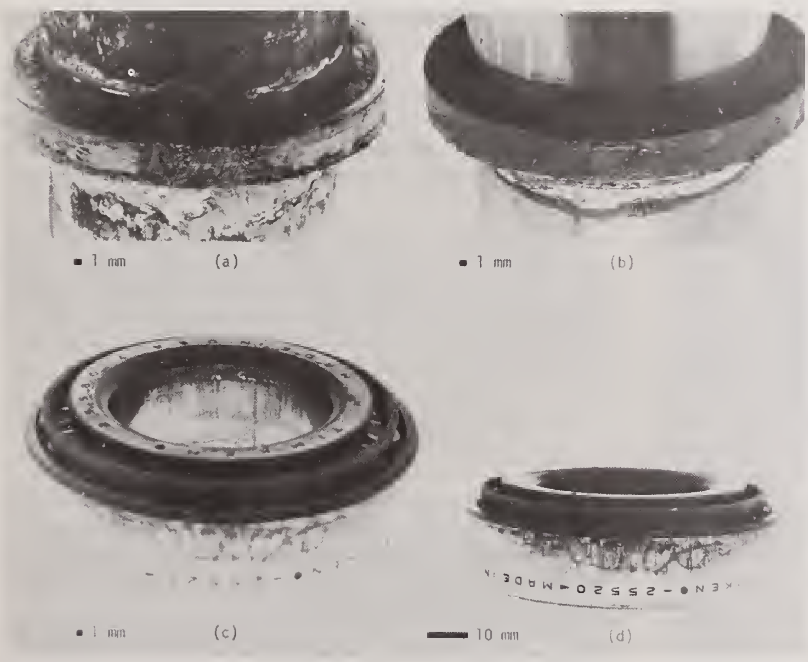


Figure 11. *Fretting damage of fitted surfaces. (a) and (b) are fretted shafts from heavily loaded tests.*
(a) had become worn from repeated use and the cones which are mounted on either side of the integral spacer had very little press-fit. (b) shows one seat that was damaged and one that is still OK. (c) shows a cone bore with fretting damage and (d) shows a cup O. D. with slight fretting damage.

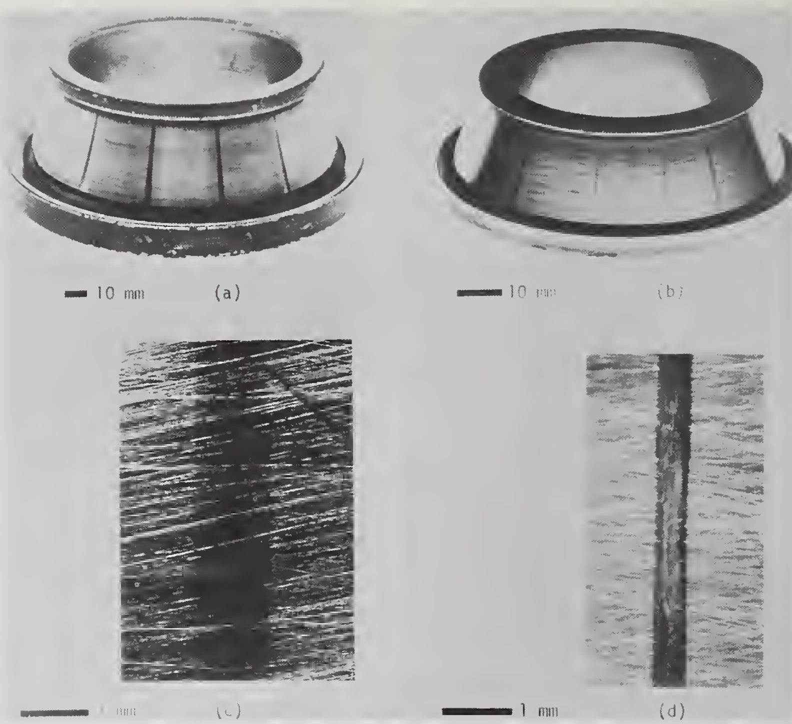


Figure 13. *Brinelling damage on cones caused by high static or impact loads in (a) and (b).*
Brinelling at high magnification in (c) is compared to fretting damage (false brinelling) in (d).

Figure 14. *An example of burn-up with plastic flow which is typical of a bearing that was inadequately lubricated, and/or overloaded.*
The damage starts out as scoring or adhesive wear on the rib and, if the heat cannot be dissipated as fast as it is being generated, then seizure takes place followed by hot plastic flow.

9.2. Arcing

Another type of damage resulting from electrical current is arcing (shown in fig. 16). This occurs by the same mechanism as fluting except that it occurs when the bearing is stationary or rotating slowly and the current passes through only a very few asperity contacts. The individual arc-affected areas are damaged more severely, but the bearing surfaces are not changed as in fluting. Localized melting and vaporization occurs at the arc and the nearby material is rehardened and tempered.

One source of arcing damage is from grounding of arc welding through bearings when repairs are made on railway or rapid transit cars. The other major source of arcing damage is from drive motor current passing through the bearing to the ground.

10. Contact Fatigue

If a bearing is not damaged by one of the modes of damage described above, then the mode of damage ultimately seen will usually be contact fatigue. The load-life ratings of bearings are usually based on contact fatigue as the expected mode of damage.

Contact fatigue involves the initiation and propagation of cracks from surface and/or subsurface origins. The sites for crack initiation are usually at stress concentrations such as non-metallic inclusions, grinding

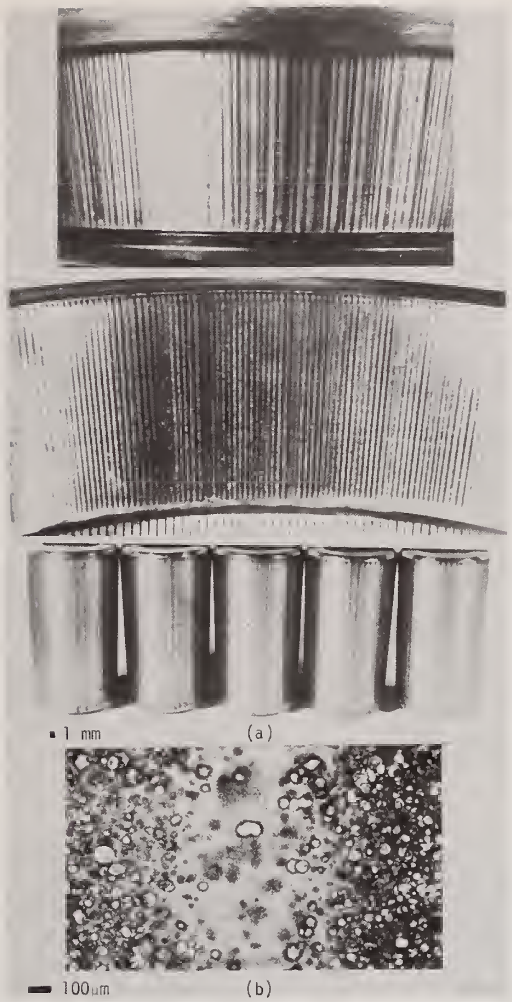


FIGURE 15. *(a) Fluting damage caused by continuous passage of electrical current through a bearing.*
In (b) a roller was polished directly on the O. D. and etched in nital to show the many individual arc marks which are responsible for destroying the geometry of the components.

furrows, edges of contact, and dents or nicks on the surface. Environmental factors such as lubrication, temperature, speed, load, and deflections have a strong influence on the mode of damage and the propagation rate of fatigue cracks. One or more modes of contact fatigue damage initiate and propagate simultaneously at different rates. The characteristics of each mode, the causes and factors which affect the propagation of each, and examples of each are as follows.

11. Subsurface Origin Spalling

11.1. Inclusion Origin

The classical inclusion origin spall is the semi-ellipsoidal spall, shown in figure 17, having its origin located at a depth corresponding to the region of maximum cyclic shear stress. Non-metallic inclusions of the aluminum or silicon oxide types (or hard alloy carbides in some high temperature bearing steels) provide stress concentrations at which fatigue cracks initiate and propagate laterally, circumferentially and upward to the surface.

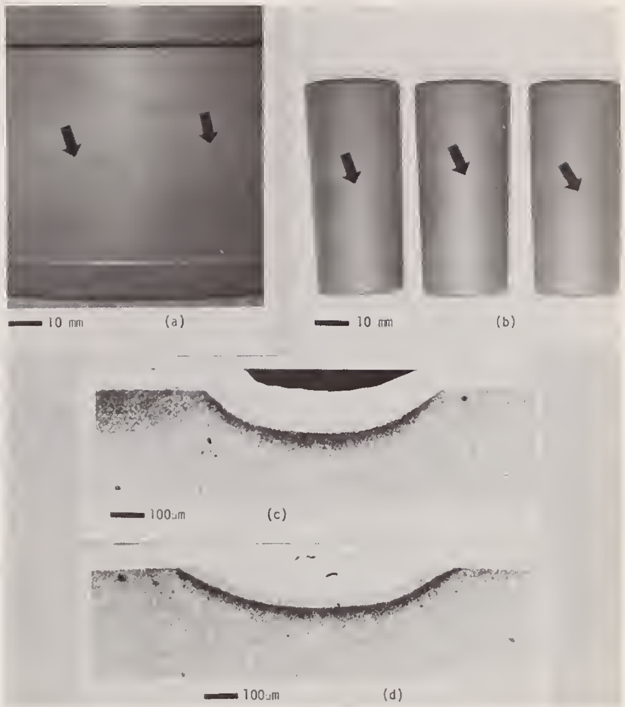


FIGURE 16. *Electrical arcing damage to a cone and rollers caused by grounding of electrical current through the bearing.*
Here the current goes through only a few points and the individual points are more severely damaged but the geometry is not affected as much and the consequences are not as severe as in fluting. (c) and (d) are transverse sections of the cone and a roller, respectively. Nital etch.

In tapered roller bearings operating under relatively thick EHD film conditions, the inclusion origin spall will propagate slowly or not at all beyond the original spall. If propagation does take place under these conditions it will be primarily in the lateral direction because of the contact geometry [5, 16].

When conditions of low EHD film thickness are present the visual appearance of inclusion origin spalls varies from the typical semi-ellipsoidal spall to spalls resembling the point surface origin mode, with extensive circumferential and lateral propagation [5]. Bearings operated under thin film conditions behave this way because the lower viscosity and/or thinner lubricant film alters the initial depth and propagation rate of fatigue cracks. It seems likely that presence of the thinner film increases the tractive force on surface asperities which penetrate the film, causing the depth of the maximum shearing stress to approach the surface. There is also a likelihood, when the thinner film is due to a low viscosity lubricant, that the lubricant penetrates the crack and increases the crack propagation rate in a manner which is not fully understood.

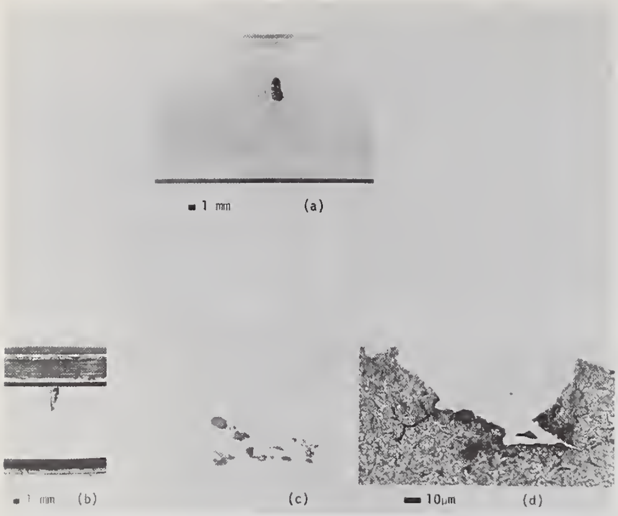


FIGURE 17. *The inclusion origin mode of contact fatigue.*
(a) A typical semi-ellipsoidal inclusion origin spall on a tapered roller bearing cup or outer race. (b) An inclusion origin spall located at the edge of the race where a geometric stress concentration and an oxide inclusion caused the spalling. The vertical line shows the plane of metallographic sectioning for finding the inclusion shown in (c) and (d). (d) was etched in 4 percent nital.

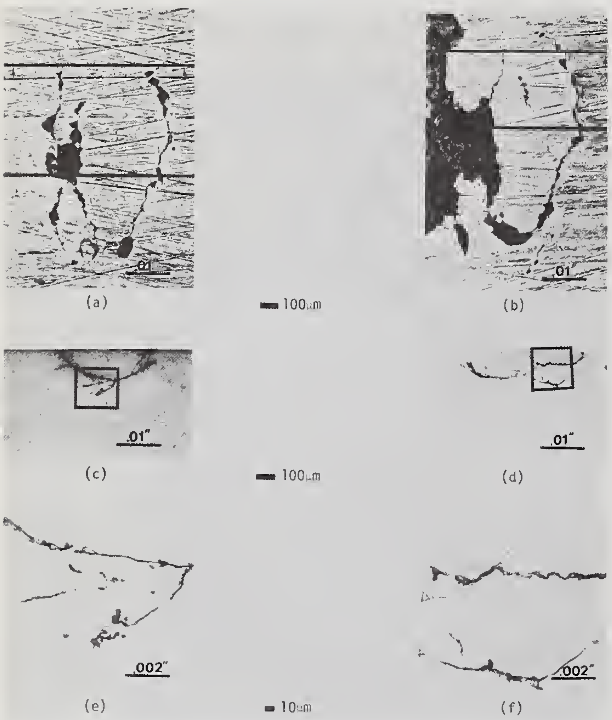


FIGURE 18. *Inclusion origin fatigue on a cone that ran in SAE 10 + 6.5 percent EP additive at 150 °F.*
hg=7.6 microinches (193 nm). (a) Incipient inclusion origin fatigue. Micros in (c) and (e) were taken on the upper black line which is a transverse plane. (b) Inclusion origin similar to (a) but this one propagated extensively for over 100 degrees. (d) and (f) are on the transverse plane at the top of (b). An additional plane is shown in figure 19.

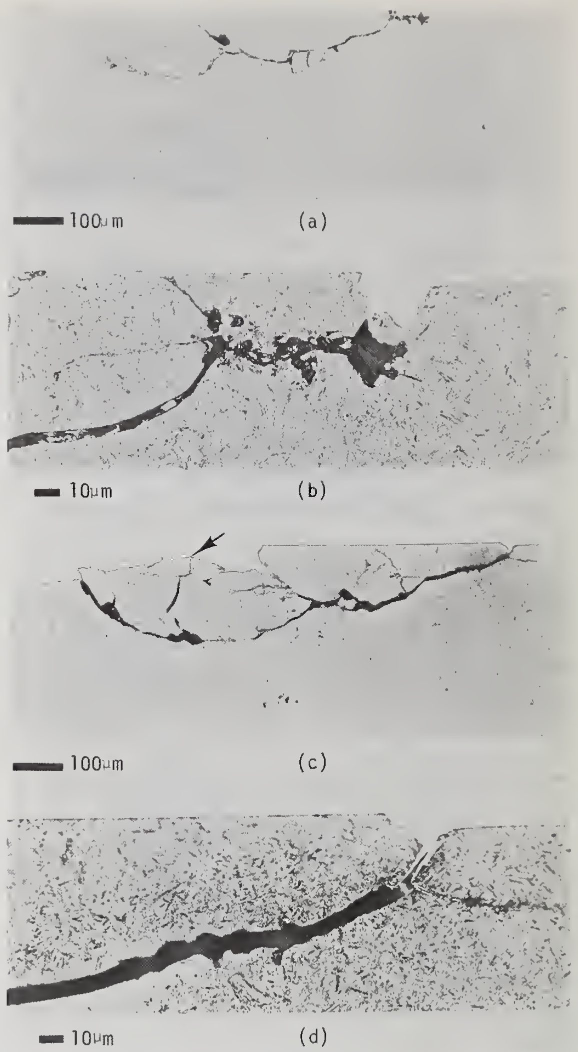


FIGURE 19. *Additional planes of polish [bottom lines of figure 18 (a) and (b)] showing the subsurface cracking associated with the inclusion mode of contact fatigue.*
(a, b) Incipient spall that did not propagate; (c, d) incipient spall that propagated extensively. Note the shallow angle of the propagating crack in (c) that starts at the arrow.

Once a fatigue crack is formed under thin film conditions, the propagation is rapid and the life of the bearing is shorter than under thick film conditions. Sometimes a semi-ellipsoidal spall is formed followed by rapid circumferential propagation. Fatigue cracks can also initiate subsurface and, when they reach the surface, the rapid circumferential propagation occurs as in figures 18 and 19 without formation of the semi-ellipsoidal spall. In some instances the fatigue will start at or very near the surface, where an inclusion intersects the surface, without any of the usual crack pattern of a typical inclusion origin spall.

11.2. Geometric Stress Concentration (GSC)

Tapered roller bearings, straight roller bearings and some other types of components have cylindrical shapes of finite length which means there can be a corner to create a stress concentration [29]. In addition, any deflection or misalignment in the bearing will result in further concentration of the load at the end of contact [13]. The coincidence of a nonmetallic inclusion at the point of geometric stress concentration will also intensify the stress concentration on the material. The depth of the fatigue crack origin is a function of the load, the geometry of the corner, and the lubricant film thickness.

Geometric stress concentration spalls vary greatly in appearance, from very shallow microspalling (peeling) with little lateral propagation, to deep subsurface origin spalling with extensive lateral propagation as in figure 20, or to a single semi-ellipsoidal spall caused by an inclusion coincident with the stress concentration as in figure 17b. Under thin EHD film thickness conditions, the geometric stress concentration can produce point surface origin spalls that propagate extensively in the circumferential and lateral directions (see fig. 23).

Geometric stress concentration fatigue is generally not a problem at or below the rated load of a bearing unless there is a substantial misalignment or large deflection that causes increased loading at one end of contact as shown in figure 20.

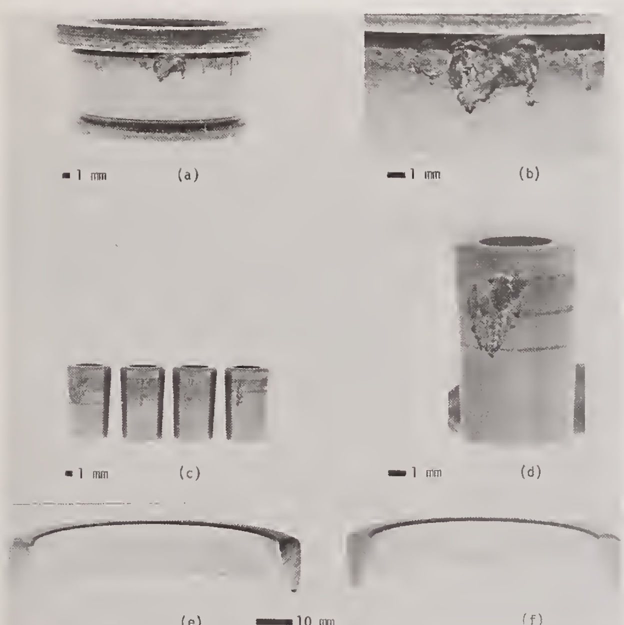


FIGURE 20. *Geometric stress concentration fatigue.*
(a, b) Cone (inner race) with GSC fatigue at the undercut. (c, d) Rollers from the cone in (a, b) with GSC where the roller contacted the edge of the cone race. (e) Cup or outer race with load zone typical of misaligned cup with microspalling at the edge of the raceway. (f) Another cup with the same amount of misalignment (0.007 in/in) but with twice the load as in (e).

11.3. Subcase Fatigue (Case Crushing)

Subcase fatigue is a mode of fatigue associated only with case hardened components [30]. It is caused by high overloads for a given case depth condition. It has been observed most frequently in contact fatigue rig tests where contact stresses are many times higher than those encountered in most bearing applications. An example of subcase fatigue is shown in figure 21 from a 0.5″ (12.7mm) diameter rig test specimen loaded "nut cracker" style with crowned 8″ (203mm) diameter loading rings.

12. Surface Origin Spalling

In tapered roller bearings where the roughness of the contacting surfaces are only of microinch (nanometer) magnitudes, surface origin cracks are usually associated with low viscosity lubricant, or other conditions such as high temperatures or low speeds which cause low EHD film thickness. Extraordinary asperities of large magnitude may cause surface origin cracking with relatively thick EHD films. Combined rolling and sliding will also promote surface origin spalling because the maximum shear stress range is nearer the surface than for pure rolling contact.

12.1. Point Surface Origin (PSO)

The point surface origin spall has a characteristic "arrowhead" shaped origin pointing opposite the direction of load approach as shown in figure 22. The crack is initiated at the surface and propagates rapidly, circumferentially and laterally. The depth of the crack increases as it propagates circumferentially.

The stress concentrations, that can lead to PSO spalling include prominent grinding or honing scratches that lead to microspalling under thin film conditions. Handling nicks and bruises from debris can cause large asperities that act as stress concentrations leading to PSO spalling. Examples of these are shown in figures 22 and 23.

Another stress concentration, already mentioned, is the geometric stress concentration at the ends of roller contact which can result in PSO fatigue under thin EHD film conditions, as shown in figure 23.

Under thin EHD film thickness conditions, there is an increase in asperity contacts in the contact area which causes the maximum shear stresses to occur nearer the surfaces depending upon the magnitude of the tractive-force [31]. Leibensperger and Brittain [32] showed that there is also a micro-Hertzian stress distribution at surface asperities which is superimposed on the macro-Hertzian stress distribution. These factors result in fatigue cracks initiated at the contact surface.

A distinctive characteristic of PSO fatigue is the rapid propagation and the direction of propagation of the cracks from a variety of origins. Low lubricant viscosity at the bearing operating temperature promotes the rapid propagation associated with PSO fatigue. Since the lubricant has immediate access to the surface cracks, physical and chemical effects influence the propagation behavior earlier than when the origin is sub-

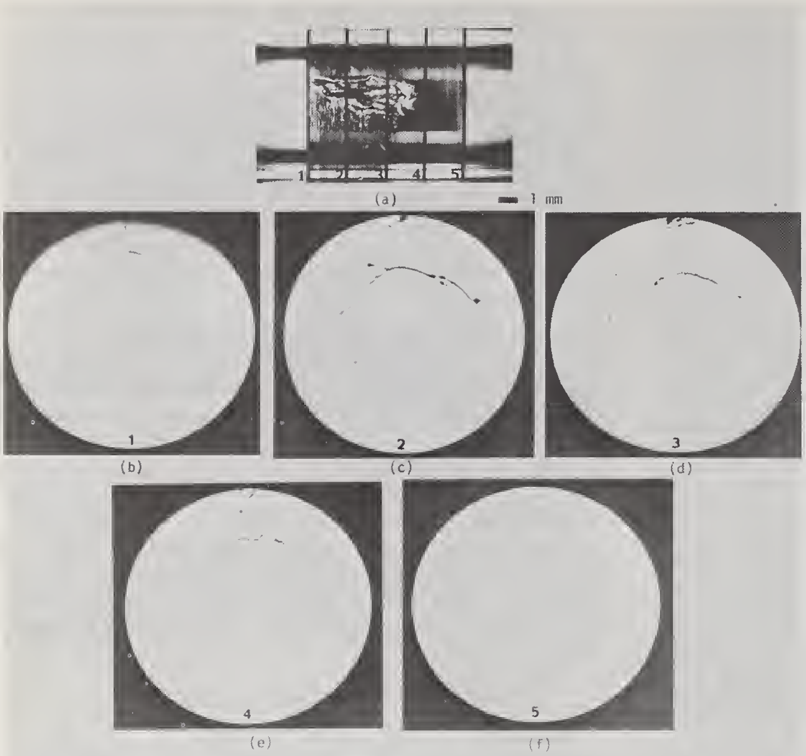


Figure 21. *Subcase fatigue. The example shown here is from a heavily loaded contact rig specimen which is where this mode is usually found.*

The external appearance of this mode of fatigue is not usually indicative of what lies beneath the surface as is shown above in (a). The specimen was loaded to an equivalent bearing load of 750 percent which made the calculated shear stress depth at 0.012 in (0.305 mm). The total case depth was 0.0625 in (1.59 mm). The loading wheels were crowned which concentrated the load at the center of the contact where the subcase cracking was deepest as in (d).

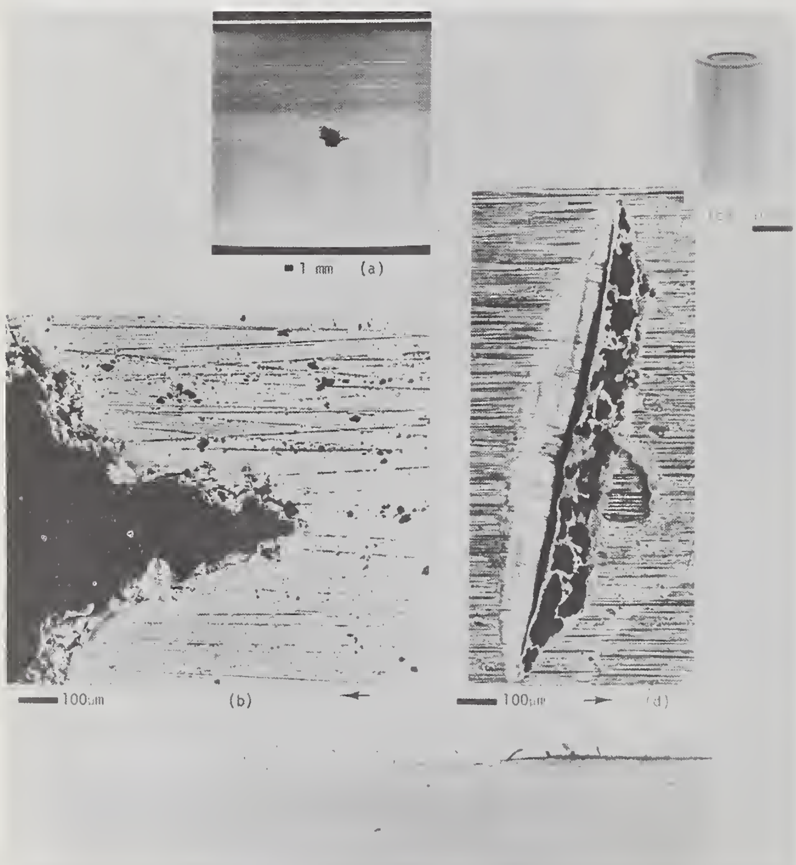


Figure 22. *(a, b) PSO spall on a cone with the characteristic arrowhead shape.*

(c, d) Handling nick on a roller responsible for the stress concentration that caused the spalling in (a, b). (e) Transverse section through the PSO spall in (a, b) showing the shallow entry of the fatigue crack. Nital etch. Arrows show direction of load approach.

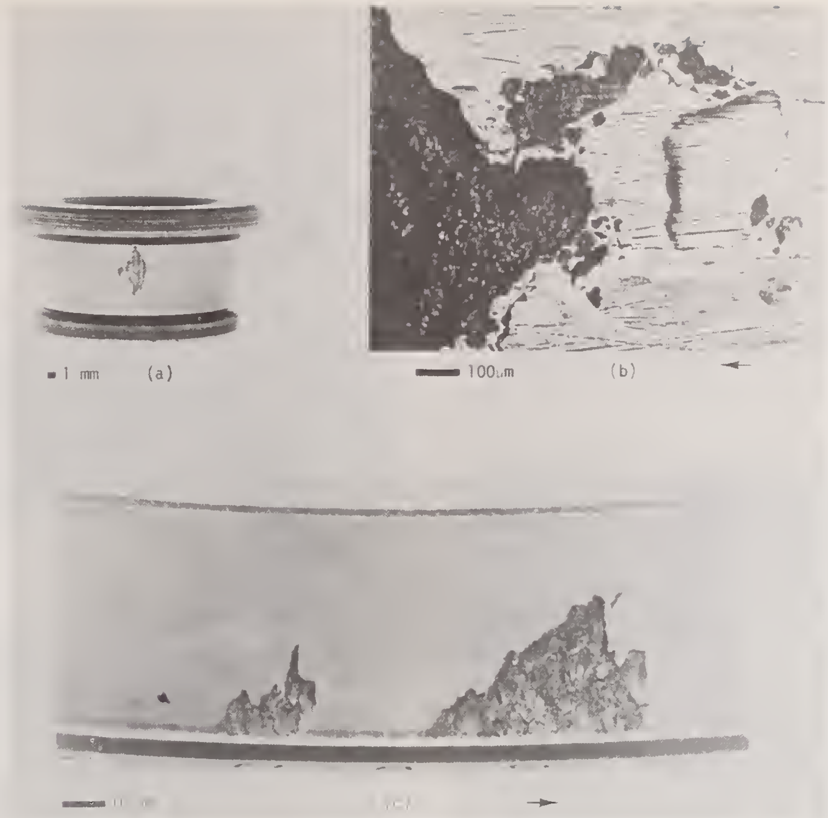


FIGURE 23. *Point surface origin fatigue.*
(a) PSO fatigue spalling on a cone caused by an asperity adjacent to a debris bruise as shown in (b).
(c) Microspalling and PSO fatigue on a cup caused by a geometric stress concentration at the amall end of the raceway. Arrows indicate the direction of load approach.

surface. Whether the origin is an inclusion, geometric stress concentration, a spot of micro-spalling, an asperity at a handling nick or propagating cracks from a subsurface inclusion, the arrowhead shape and rapid propagation are also invariably present.

12.2. Microspalling

Microspalling is a very shallow fatigue spalling of the surface that occurs under low EHD film thickness. Sites of crack initiation are at peaks of the surface finish and adjacent to prominent grooves from finishing operations, debris bruises, handling nicks, other fatigue spalls, and edges of contact. This mode of fatigue was formerly called peeling by the authors [14, 16, 18, 20] and examples are shown in figure 24.

The conditions that prevail for PSO spalling are identical for microspalling. The two modes often occur together. On occasions microspalling develops into PSO spalling. The cracking and microspalling seldom extend below 0.001″ (25.4μm) in depth. Tallian, et al. [33] proposed a mathematical surface model and Leibensperger and Brittain [32] presented experimental evidence of a micro-Hertzian stress distribution at surface asperities that explain the probable mechanism of the microspalling mode of fatigue.

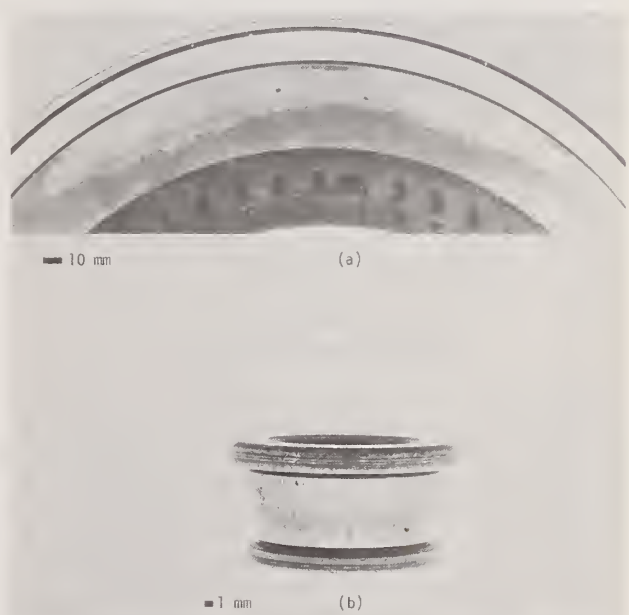


FIGURE 24. *Miscrospalling (peeling) on a 12 in (305 mm) O.D. double cup in (a) and a ¾ in (19 mm) bore cone in (b).*

When two surfaces of different roughness are run under thin EHD film conditions, the smoother of the two usually develops a general microspalling condition while the rougher of the two will have only the peaks of the roughness spalled away as in figure 25. Microspalling appears to relieve the stress concentration that caused it and may not propagate beyond this stage in some applications. Microspalling does generate a new edge of contact if the spalling does not extend over the entire length of contact. Such a discontinuity can become a site for GSC or PSO fatigue under certain conditions. If other damage modes do not accompany microspalling, the useful life of the bearing or component will probably not be affected in most applications.

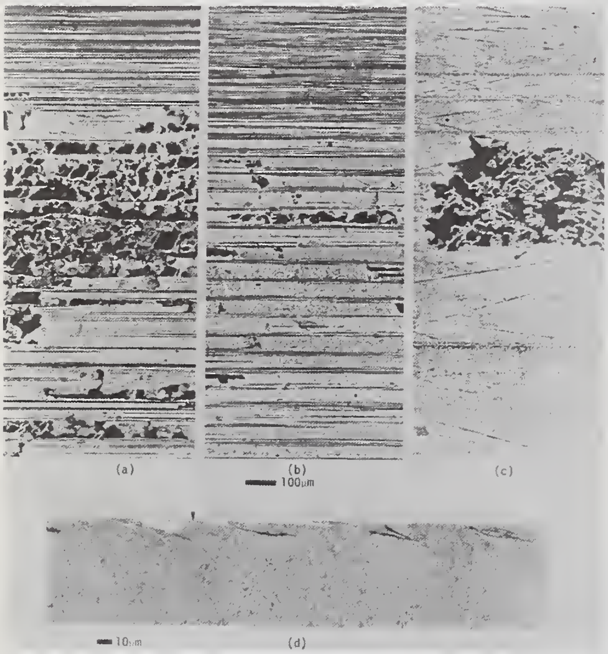


FIGURE 25. *Microspalling (peeling) on a tapered roller bearing caused by a thin lubricant film compared to the composite surface roughness.*
(a) Cup showing fatigue on the peaks of surface texture. (b) Cone showing fatigue on the peaks of surface texture. (c) Roller with a general spalled area. (d) Transverse section through a slightly spalled area on a roller showing typical shallow cracking. Nital etch.

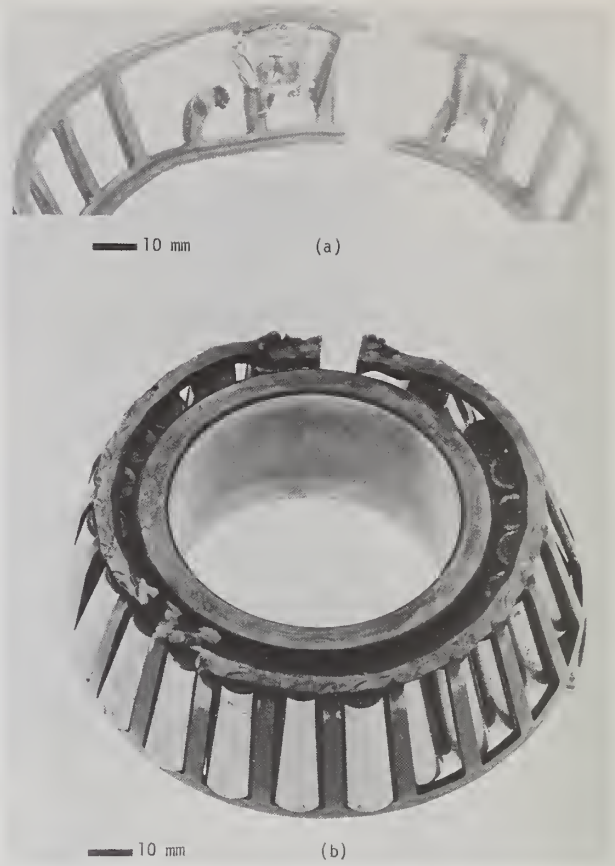


FIGURE 26. *Section fracture of cages. In (a) the cage has fractured because of inadequate lubrication.*
Plastic flow of the cage has also taken place. In (b) the bearing was mounted improperly so that the cage was rubbing against something which eventually resulted in cage fracture.

13. Section Fracture

Occasionally, the operating conditions in a tapered roller bearing are severe enough to cause fracture of some components. Cage fracture often occurs under inadequate lubrication conditions because of the increased forces exerted on the cage by the rollers as shown in figure 26a. An improper system design may result in the cage rubbing against some component of the system causing cage breakage as shown in figure 26b.

Cone fracture (formally called Transverse Cracking by the authors) may result from excessive press fit from fatigue propagated cracks generated by fretting between the cone and shaft, from repeated thermal stresses in burn-up damage, or from extensive fatigue spalling on the cone raceway.

Cup frature may result from foreign material entrapped between the cup O.D. and the cup seat, from extensive spalling on the cup raceway propagating through the section when it is not effectively supported or from cracks initiated by fretting of the cup O.D. propagating through the section. In one example, foreign material was entrapped between the cup O.D. and the cup seat. The distortion of raceway geometry caused early contact fatigue of the raceway and bending fatigue fracture which propagated partially through the cup section. In another unusual situation, fretting damage on a cup O.D. caused a loss of support which subjected a cup to repeated bending stresses in a region where there was extensive propagation of fatigue damage on the raceway. Cracking from the contact fatigue finally propagated through the section by bending fatigue.

When a bearing inner race made from through hardened steels is press fitted onto a shaft, the initial subsurface or surface origin fatigue crack sometimes propagates radially through the entire cross section.

Examples of this mode of damage are shown in figure 27. Propagation through the section may be so rapid that no evidence of contact fatigue damage is seen on the raceway surface, as in figure 27a. The low fracture toughness of the through hardened section, the residual stress and press fit stress combine to cause this mode of fatigue fracture.

In balls and rollers of all types, section fracture can sometimes occur under very high overload by fatigue propagation from inclusion origin fatigue cracks.

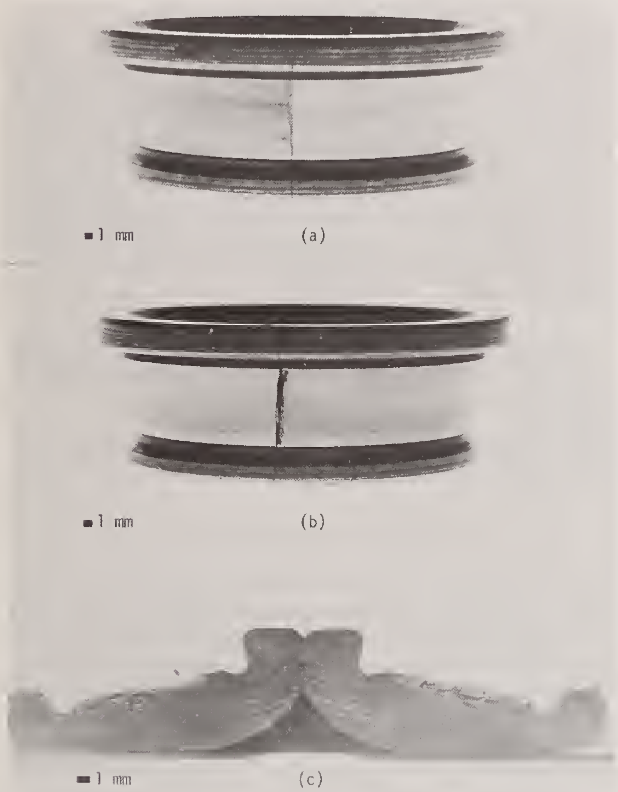


FIGURE 27. *Transverse cracking damage on tapered roller bearings made of through hardening steel.*

(a) One crack has propagated through the section and there are at least 4 earlier stages of cracks showing. (b) A second cone with spalling at the transverse crack. (c) The fracture surfaces of the transverse crack shown in (b).

14. Summary

Modes of damage to tapered roller bearings, many of which are seen in all types of rolling element bearings, have been discussed. Classification of the various modes of damage, according to a scheme which relates the characteristics of each mode to its most probable cause, has obvious diagnostic value when service damage is encountered. In a broader sense the reliability of machines and components are dependent upon the cognizance by the manufacturer, designer-builder, and user of the type of service damage normally encountered in a machine or component. Awareness of (1) the consequences of manufacturing variables affecting components, of (2) design and building variables in machines, and of (3) the user's application of the machine can be used to increase the reliability of mechanical systems. When used in conjunction with an effective method of bearing damage analysis, such knowledge permits one to diagnose and correct the cause of bearing damage seen in prototypes and thus greatly improve the chances for producing design life in a new machine.

15. References

[1] Zaretsky, C. V., and Anderson, W. J., Effect of materials—General background, NASA SP–237 Interdisciplinary Approach to the Lubrication of Concentrated Contacts (July 1969), p. 379.

[2] Bamberger, E. N., Effect of materials—Metallurgy viewpoint, NASA SP–237 Interdisciplinary Approach to the Lubrication of Concentrated Contacts (July 1969), p. 409.

[3] Neifert, H. R., and Wren, F. J., The rating of tapered roller bearings through continuing worldwide quality audit, Preprint 690577, Soc. Automo. Eng. (September 18, 1969).

[4] Zantopulos, H., The effect of misalignment on the fatigue life of tapered roller bearings, Journal of Lubrication Technology, Trans. ASME, Series F, Volume **94**, No. 2 (April 1972), pp. 181–187.

[5] Littmann, W. E., Widner, R. L., Stover, J., and Wolfe, J. O., The role of lubrication in the propagation of contact fatigue, Journal of Lubrication Technology, Trans. ASME, Series F, Vol. **90**, No. 1 (January, 1968), pp. 89–100.

[6] Danner, C. H., Fatigue life of tapered roller bearings under minimal lubricant films, ASLE Trans. Vol. **13** (1970).

[7] Moyer, C. A., The use of elastohydrodynamic lubrication in understanding bearing performance, SAE Preprint 710733 (September, 1971).

[8] Harris, T. A., Predicting bearing performance under non-standard operating conditions, Mach. Des. (August 17, 1967), pp. 158–162.

[9] Witte, D. C., Operating torque of tapered roller bearings, ASLE Translations, Vol. **16**, No. 1 (January, 1973), pp. 61–67.

[10] Leibensperger, R. L., An analysis of the flow of oil through a tapered roller bearing, Journal of Lubrication Technology, Trans. ASME, Series F, Vol. **94**, No. 2 (April 1972), pp. 125–130.

[11] Cornish, R. F., Acro-Set: A new method for installation of tapered roller bearings with accurate end-play setting, Soc. Automo. Eng. (April, 1966).

[12] Cameron, R. W., and Jenkins, J. A., Design parameters for automated bearing settings, Preprint No. 700730, Soc. Automo. Eng. (September, 1970).

[13] Kaufman, H. N., and Walp, H. O., Interpreting service damage in rolling type bearings—A Manual on Ball and Roller Bearing Damage, ASLE (1953).

[14] Littmann, W. E., and Moyer, C. A., Competitive modes of failure in rolling contact fatigue, Preprint 620A, Soc. Automo. Eng. (January 1963).

[15] Davies, W. J., and Day, K. L., Surface fatigue in ball bearings, roller bearings and gears in aircraft engines, fatigue in rolling contact, Inst. of Mech. Engineers, London (1963), pp. 23–40.

[16] Littmann, W. E., and Widner, R. L., Propagation of contact fatigue from surface and subsurface origins, Journal of Basic Engineering, Trans. ASME, Series D, Vol. **88**, No. 3 (September 1966), pp. 624–636.

[17] Simpson, F. F., Failure of rolling contact bearings, Lubrication and Wear, Second Convention, Inst. Mech. Eng., Vol. **179**, Part 3D, (1965), p. 248.

[18] Widner, R. L., and Wolfe, J. O., Analysis of tapered roller bearing damage, Technical Report No. C7.11.1, Am. Soc. Met. (October 1967).

[19] Tallian, T. E., On competing failure modes in rolling contact, ASLE Trans. Vol. **10** (1967), pp. 418–439.

[20] Littmann, W. E., The mechanism of contact fatigue, NASA SP–237, Interdisciplinary Approach to the Lubrication of Concentrated Contacts, July, 1969.
[21] Belasco, S., Oil Analysis Program Quarterly Summary, A/T A3400000/051B/4F53537401, WU #01, Naval Air Engineering Center (February 8, 1974).
[22] Burwell, J. T. Jr., Survey of possible wear mechanisms, Wear Volume 1, (1957/58), pp. 119–141.
[23] Cornish, R. F., A critical look at tapered roller bearing wear in hypoid axles, Preprint 650440, Am. Soc. Automo. Eng. (May, 1965).
[24] Cornish, R. F., Wear of Tapered Roller Bearings, American Society for Metals 1969 Wear Conference.
[25] Fitzsimmons, B., and Clevenger, H. D., The effect of contaminated lubricants upon tapered roller bearing wear, To be presented at the 8th Annual Fluid Power Research Conference at Oklahoma State University, October, 1974.
[26] Hurricks, P. L., The mechanisms of fretting—A review, Wear, Vol. 15 (1970), pp. 389–409.
[27] Culp, D. V., and Lieser, J. E., Fretting corrision in automobile wheel bearings, Lubrication Engineering (October, 1971).
[28] Simpson, F. F., and Crump, W. J. J., Effect of electrical currents on the life of rolling-contact Bearings, Lubrication and Wear Convention, 1963, Inst. Mech. Engr. (1963), pp. 299–304.
[29] Moyer, C. A., and Neifert, H. R., A first order solution for the stress concentration present at the end of roller contact, ASLE Trans. Volume 6 (1963), p. 324.
[30] Pederson, R., and Rice, S. L., Case crushing of case carburized and hardened gears, SAE Trans. Vol. 68 (1960), p. 187.
[31] Smith, J. O., and Liu, C. K., Stresses due to tangential and normal loads on an elastic solid with application to some contact stress problems, J. Appl. Mech. ASME (June, 1953), pp. 157–165.
[32] Leibensperger, R. L., and Brittain, T. M., Shear stresses below asperities in Hertzian contact as measured by photoelasticity, Journal of Lubrication Technology, ASME (July, 1973), pp. 277–286.
[33] Tallian, T. E., and McCool, J., I. An engineering model of spalling fatigue failure in rolling contact, II. The surface model, Wear, Vol. 17 (1971), pp. 447–461.

16. Discussion

W. Gaylord, United States Postal Service: Do you have any information which would show how to correlate loading in the case of a steady load as compared to high frequency, intermittent loading?

W. E. Littmann: You're talking about how to do a life prediction when the loading is not going to be constant, but going through some spectrum. In other words you have a periodically varying load.

W. Gaylord: It's constant but it's periodically applied. It goes from zero to the maximum in the case of a letter machine where the shaft has to flex to compensate for various thicknesses of letters.

W. E. Littmann: Many real mechanical systems operate over a range of loads and one uses a weighted average summing up of damage together with the load-life relationship, that is the life is inversely proportional to the load, to empirically determine power. This accumulation of damage can be used by simply integrating over the significant range of loading and ignoring the lower load levels to come up with a weighted average equivalent load and then compute the expected life from that load.

C. E. Vest, NASA, Goddard Space Flight Center: In a starvation situation, did you have a good lubricant at all times?

W. E. Littmann: Yes, except for the adhesive wear and burn-up types of fatigue.

C. E. Vest: Were most of the lubricants oils or greases?

W. E. Littmann: The majority were circulating oil systems, although, for example, the AP freight car bearing that had suffered corrosion damage when water got in it was a grease lubricated bearing. This is a case where channeling of the grease will tend to develop a dry condition on the raceway and make it susceptible to corrosion. So with grease lubrication, if there is a tendency for channeling I would say that it's important that the seals be kept in condition to avoid ingress of water.

MODES OF FAILURE

Session III

Chairman: J. R. Strang
Air Force Materials Laboratory

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Anti-Friction Bearing Design, Inspection and Recycling

Kay M. Glaesner

St. John's Evangelical Lutheran Church
Wittenberg Avenue at Columbia Street
Springfield, Ohio 45502

The paper will describe the ramification of failures of antifriction bearings related to aircraft applications, including engine bearings. The paper will endeavor to describe the problems of antifriction bearings in four categories:

I. Design of application. Here, the paper will describe the inherent design failures as noted in applications. The paper will also include parameters of bearing design which will best provide operational success and life expectancy.

II. Failure modes in bearing applications will be described and an outline of the most popular types of failure as detected in bearings, will be discussed.

III. A recovery process for the reclamation of bearings will be outlined and again limits for the acceptability of defects on bearings will be presented.

IV. New inspection techniques now under study will be outlined and the method of inspection will detail failure prevention once the bearing is mounted in the application.

Key words: Aircraft engine bearings; antifriction bearings; bearing failure; bearing failure modes; bearing inspection; bearing reclamation; bearing recycling.

1. Introduction

The performance of an aircraft engine is directly proportional to the precision and quality of the antifriction bearings mounted in that engine. There is no single part in engine design that plays such a vital role and determines performance more than that of a bearing. For many years in the military service, we have had an unusual opportunity to study thousands of bearings, not only as they are originally designed into the aircraft application, but also through their service life and reuse, after engine overhauls.

Through the years, we have noted a longer life expectancy in our heavy loaded bearings and this has been accomplished because of a number of factors, some of which will be outlined in this paper. In a recent life expectancy study on ten bearings picked at random from a number of positions in a number of engines, we discovered that most of the bearings had accumulated between 5000 to 7000 serviceable hours and that all of the bearings still had usable hours so that they could be remounted into the engine. This is a long way from ten years ago when the service life of these bearings was running from 500 to 2000 h.

The bearing in an aircraft engine is critical for several reasons: (1) It is subject to all the rotational stresses of the engine; (2) it carries all the radial thrust and torque loads imposed upon the engine assembly; and (3) because of its sensitivity of design it is most vulnerable to any variations in normal engine operation, such as: heat transfers, misalignment, improper load distribution, shaft deflections, shock loads, or improper assembly operations. As soon as one or more of these discrepancies occurs the antifriction bearing is the first to pick up the signal and sooner or later, will relate to itself such discrepancies in the form of some bearing distortion, fatigue or failure. Many of the improvements in bearing aircraft design have resulted from our analysis of bearing failures. As indicated, failures of bearings are not only attributed to the original design for the particular application, but also to components associated with the system. By noting these factors, we have been able to improve the life expectancy of aircraft engine bearings over the last 25 years.

2. Design of Application

The following design factors for bearing application are of great importance to insure precision operation and maximum life.

(a)—Fit-up between Bore and Shaft. The fit-up between the bore of the bearing and the shaft is extremely important. If an out-of-roundness or taper exists in a particular bearing bore, then you can be assured of a reduction in the life of the bearing. In bearing design there is a great effort made to control the tolerances of the bearings, but often little attention is given to the matching dimension of the shaft. An improper fit-up or a taper will place undue stress upon the rings as well as upon the rolling elements as they pass through the loading zone. Oftentimes, as a result of an improper bore mounting, a fracture will occur or a fatigue failure will develop in the raceway proper. Sufficient equipment is available today to determine proper bore diameters and fits.

(b)—Parallelism of Sides. One of the most outstanding causes for misalignment of the bearing itself is at-

tributed to non-parallel sides of the bearing. Here again the bearing will assume a cocked position in the installation and cause undue stress on the rings and rolling element.

(c)—Eccentricity. Here is one design factor which is often overlooked, especially in high speed applications. Inherent in the bearings is this built-in problem which during operation is prone to develop vibratory loads. A good way to measure eccentricity is simply to mount the bearing on an arbor, having a very slight taper (0.0001 to 0.0002 inch in diameter per inch of length), and apply an indicator on the center of the stationary outer ring. The eccentricity is the difference between the minimum and maximum readings when rotating the arbor one revolution. With this inherent problem, considerable out-of-balance may be experienced.

(d)—Side Runout. Another important factor is the side runout of the bearing which creates undue thrust loads on the bearing itself, and if such loads exceed the thrust load capacity, it is not long before fatigue will set in. For measuring this, use the same arbor technique as described above. Apply the indicator against the side of the inner ring. The side runout is the difference between the maximum and minimum reading when rotating the arbor one revolution.

(e)—Groove Parallelism with Side. One traditional way to create undue stress on the rings and rolling elements is to have some discrepancy in the parallelism between the groove of the raceway and the side of the ring. During operation, a counter axis is developed and the stress centers between the balls and the raceways, resulting in a shifting of race curvature.

(f)—Outside Diameter. As with the bore, so often we overlook the necessity of a proper fit between the outside diameter of the bearing and the housing. Depending on design, this fit should be prescribed. If the outer ring is to be fixed, especially for cross loading, then the fit-up should follow. If the outer ring is to float to provide for thermal expansion, then the fit-up for this type of application should follow. Although it is very difficult to require exacting tolerances in housings, such as are required in antifriction bearings, the solution for a proper mounting may be accomplished by the technique of "selective fits."

Other design factors related to engine bearing problems are also associated with such problems as the constant flow of lubricant to reduce the effect of heat, shaft deflections which may occur after an engine is assembled which causes misalignment of the bearings, and damage to the bearing parts due to contamination and improper handling.

Another important factor in designing a bearing for an application is the "internal fit-up within the bearing." Except for special bearings, the internal fit-up in antifriction bearings does not have a uniform standard dimension established by the manufacturers. Generally speaking, they are classified into four groups:

(a)—Standard Fit-up. In this class, internal unmounted clearances are to be of such magnitude that, as a result of the bearing's internal design, when mounted with standard ABEC fits for revolving shafts, the average radial clearances will be best suited for single row annular bearing service.

(b)—Loose Fit-up. Here the bearings can accommodate heavier press fits with a resultant small amount of radial clearance.

(c)—Tight Fit-up. This type of design will actually create a preload condition within the bearing.

(d)—Selective Fit-up. This is the type of internal fit-up that best describes the high precision, heavy load or high rpm bearing. This type of internal clearance is a pre-calculated one in which consideration is given to the thermal flow, the lubrication, torque loading, and the resultant ability of the bearing to make compensation for variations in operation.

The above factors have been the most important design needs in bearing manufacturing for aircraft engines. Of course, the internal design of the bearing such as race curvature, angle of contact, retainer drag, and type of retainer, are also important in choosing a bearing, but these factors are usually taken up by the bearing manufacturer, rather than the application designer.

3. Failure Modes in Aircraft Engine Bearings

Visual inspection for bearing defects has been, up until recent years, the main technique for determining the serviceability of bearings, especially where such bearings have been in a previous application. Visual inspection is now passing out of the picture and more sophicated techniques are being employed to determine the serviceability of the bearings. Later in this paper an outline of some of these new techniques will be presented. The purpose now is to provide descriptions and service limitations of defects most commonly found in aircraft engines.

(a)—Indenting. This indenting may occur any place on the bearing rings, raceways, or rolling elements. Usually indentations will appear as cavities with smooth bottoms and sides. The cavities are depressions in the operating surface where the metal is displaced but not removed. This condition is caused by the passing of foreign or loose material between the races and the rolling elements. Usually any dent up to 0.060 in long which cannot be felt with a 0.030 in radius scribe, is considered acceptable unless there is a cluster of four or more dents within a quarter in diam area.

(b)—Fatigue Pitting. This kind of pit may be recognized by the occurrence of irregular, relatively deep cavities on the rolling surfaces. These cavities are usually the result of subsurface, molecular breakdown or voids and can be further distinguished by a characteristic fractured appearance on the side. Fatigue pitting on a bearing renders it unserviceable. Geometric stress concentrations may often generate this defect.

(c)—Corrosion Pits. Abrasive wear results from small droplets of corrosive material attacking the exposed surface of the bearing. They are distinguished by irregular shaped, ragged edged pits which have a reddish color. Usually such corrosion is widely distributed and if such pits cannot be felt with a 0.030 in scribe,

the bearing can be serviceable if the cluster is less than four within a quarter-in diam area.

(d)—Nicks. Nicks on bearing surfaces result from bearing components striking together either in shipment or during application. Normally, this type of defect is acceptable and not cause for rejection unless it exceeds 0.060 in in length.

(e)—Scratches. These are linear abrasions of the surface which again occur from improper handling and are not usually a cause for bearing rejection unless they can be felt with a 0.030 in scribe.

(f)—Adhesive Wear. Circumferential bands and grooves on rolling elements result from foreign material which becomes trapped or embedded between rollers and retainer pockets. It can also be produced by breakdown of lubrication, scuffing, smearing or galling. Axial scoring occurs often during assembly or disassembly of bearings. When the bearing has two or more such defects, the bearing is considered unserviceable.

(g)—Axial Galling. This results from an ineffective lubrication at the sliding contact surfaces. Galling may be recognized by the presence of metal from one part remaining attached to another. Often it is also caused by metal removed from the retainer by the rolling elements and deposited on the races. No form of galling is acceptable.

(h)—True Brinelling. This results from extremely high shock loads which leave permanent impressions at each rolling element position on the loaded side of the bearing. Often this is the case when a shock load occurs while the bearing is in a stationary position and even while it is in transit. Visual inspection does not often detect true brinelling, but it can be seen by reflected light measurements. No form of true brinelling is acceptable.

(i)—False Brinelling. This is a specialized form of fretting occurring only at the rolling contact surfaces of the bearing raceways. It may be recognized by a series of shallow indentations in the race at each rolling element position and is also accompanied by a red iron oxide substance. This brinelling does not occur when a bearing rotates.

(j)—Bearing Discoloration. Discoloration may come from heat or stain due to oil or grease and is not considered harmful unless it has advanced to an etching or oxidation process on the race.

Of course, the obvious failures of bearings such as fractures, broken retainers and dimensional changes are often seen in engine bearing applications. The above outline of failure modes has provided bearing engineering with the background for the development of more advanced bearing inspection equipment and techniques. In 25 years of research, we would say that the subsurface, as well as the surface, of the bearing plays as vital a role as does the geometry of the bearing itself. Of course, unusual temperatures, shock and operating conditions vitally affect the various bearing defects and accelerate bearing failures.

4. Recovery Process

One of the techniques developed by the Air Force and currently used to analyze bearings arises from the Air Force program to recover used antifriction bearings. About 75 percent of engine bearings are now being recycled in a program of bearing reclamation. The new technology now being developed will increase the number of bearings which can be reused in aircraft installations. The Air Force is at the present time considering an enlarged facility to reclaim antifriction bearings. An inspection at one of our bearing shops would include the following:

- Initial inspection—to remove obviously damaged bearings
- Cleaning—to remove foreign elements
- Polishing—to restore surface finish
- Rework—to accomplish retainer repair
- Measurement—to determine internal clearance, race curvature and dimensions
- Calibration—matching angular contact and preload
- Inspection—to determine metallurgical and operational condition
- Preservation—to protect against corrosion
- Packaging—to provide for adequate protection during transportation.

The recovery process is timely because of the energy crisis, as well as the scarcity of man hours and raw materials. The dollar savings in this program is estimated at 50 percent over the cost of new bearings.

5. New Inspection Techniques

A considerable amount of study on causes and effects of bearing failures has indicated that many of these failures occur beyond the visual concept and are oftentimes related to subsurface structure and to the bearing race finish per se. The classical methods for nondestructive evaluation have proved to be too limited in meeting the new requirements of the space age. The limits of sensitivity have now been defined to a point where conventional analysis for bearing serviceability is inadequate. New modern techniques for bearing inspection are well underway and will not only prove to be very efficient in detecting bearing failures but also will provide assurance for bearing life predictability.

The purpose here is to call attention to some of these advances, and while some of the techniques are not new, we do believe that only recently have they been applied to bearings. Some of these new techniques are as follows:

(a)—Magnetic Field Perturbation. We have known for many years that the performance of high speed and heavily loaded bearings is directly related to the condition of the subsurface structure of the bearing metal. Subsurface voids and contaminants provide flaws which will in turn affect the smooth operation of the bearing and eventually become the actual cause of bearing failure as the subsurface deteriorates. Many experiments have proved that most all outward signs of fatigue and failure are related initially to subsurface flaws.

Magnetic perturbation is a system to measure these subsurface imperfections. Using the Hall effect, this system is capable, through sensitizing a probe signature, of measuring the subsurface defects and voids when a uniform magnetic field is induced into the bearing raceways. The signature yield from the subsurface inclusion or void becomes a measurable value. The density of such inclusions or voids, their location, and the depth from the surface can all be detected in a second, measured mechanically, and thus used to determine actual serviceability as well as life prediction.

(b)—Stress Measurements by Means of the Barkhausen Effect. This is a system which has objectives as follows:

(1) To prepare Barkhausen stress measurements on the reference bearing at the beginning and end of its normal life's service.

(2) To analyze the records obtained and from this analysis evaluate the effectiveness of the Barkhausen noise stress measurement method for determining the uniformity of the residual surface stresses in new bearings and measuring the change of the stress caused by service life and environment.

(3) Laser optical inspection. This is a system which aims to develop a laser optical technique for the automatic inspection of bearing parts which eventually will eliminate the need for visual inspection. The design of this system is based on the principle that the directional characteristics of light reflected by a surface contain information pertaining to the profile of the surface. The photodetectors are used to measure certain aspects of the directional distribution of the reflected light and the analyzer, which might be a simple electrical analog computer, processes the photodetector response data. The work being accomplished by this system will help us to ascertain surface geometry as well as being a method to determine defects not noticeable by the naked eye. This system will provide automatic characterization of bearing surfaces in terms of defect parameters.

(4) Automated bearing life cycles. Considerable work has been done to establish a data storage system through the use of automation. In order to provide a profile and statistical analysis on critical bearings, such a collection of data would include:

Number of hours of life
Population statistics
Magnetic perturbation signature
Barkhausen signature
Optical-metallographic examination
Physical dimensions and tolerances.

Work in this area is enabling us to predict life expectancy, performance ability, and premature failure. In the space age, flights require a precision and determinability beyond anything of the past. There is no reason why bearings cannot be included in the present science which allows for automated data. Automation now makes it possible for us to measure operational conditions of a bearing under a given load so that simultaneously we can empirically measure such factors as film thickness, retainer drag, ball contact, temperature variations, and torques at any given instant and at any position while the bearing is under mission operation.

The author wishes to acknowledge the following for certain concepts included in this presentation: J. R. Barton, Southwest Research Institute, Pat Hampton, Robbins Air Force Base, Warner Robbins, Georgia, Charles Eddy, Hdqrs. AFLC, WPAFB, Richard L. Jentgen, Battelle Columbus Laboratories, Columbus, Ohio.

6. Discussion

C. E. Vest, NASA, Goddard Space Flight Center: Do you use grease or oils in most of your applications?

K. M. Glaesner: Oil.

H. Ferris, Hughes Helicopters: I am interested in your predictions for the future for replacing aircraft engines, transmissions, and bearings, which as you say, is costly. Will it be possible to predict on an "on condition" basis rather than a TBO basis by using diagnostic means like acoustics so you don't have to overhaul something at some prescribed period?

K. M. Glaesner: Yes, that is what we are working for. If we can analyze a bearing or require of the bearing manufacturer certain parameters which we know are important in ascertaining the life expectancy of a bearing, then we can be reasonably sure that that bearing is going to operate for the maximum life that we will need to perform our particular mission. That will, of course, reduce a lot of the rework and inspection. Some of the factors involved in our program are the scarcity of bearings, the delay oftentimes experienced in obtaining them, and having the technical knowledge to repair a bearing in the event that we would have a national emergency. The biggest factor has been that of endeavoring to establish a quality control on all of our bearings. This is one of the things that we are hoping will be accomplished when we develop this technique of examining the subsurface of the bearing, the proper condition of the bearing, the surface condition of the bearing, and inner stress of the bearing. In instrument bearings right now I think it's safe to say we are using only 30 out of every 100 bearings we receive from the manufacturer. We don't blame the manufacturer for this. It's just that our requirements have risen so greatly that by the time the manufacturer gets down to fulfilling our requirements the bearing gets so costly that it's almost prohibitive for us to buy. By inspecting the bearings ourselves, we find that we get better deliveries and better production requirements.

H. Ferris: Why overhaul it at all if it is running good?

K. M. Glaesner: We don't. If the bearing is satisfactory when we take it out, we don't do anything with it except clean it.

H. Ferris: Why take it out at all, why not leave it on the aircraft and let it run?

K. M. Glaesner: Usually a bearing is taken off an aircraft when the aircraft is being overhauled for other purposes. When I first started working in bearings at Wright Field it was mandatory to take the bearings off at regular time intervals such as 300 hours, 500 hours. Now we'll operate an aircraft engine for as many hours as that engine will operate showing no defects or defaults and we won't do anything with the bearing.

H. Ferris: You don't have TBO's in the aircraft engines in the Air Force?

K. M. Glaesner: Not in all of them.

J. E. Stern, NASA Goddard Flight Center: You mentioned detergents as being deleterious to the surface of a bearing. Can you name names and tell us what the effects are?

K. M. Glaesner: Most any of the standard detergents used for cleaning raceways of bearings will leave some sediment on the bearing. Some of the new systems of alcoholic ether spray seem to be less deleterious than the normal detergents which you can buy from various manufacturers. For thirty years I have worked with bearings, and I never dreamed that the problem of wetability could be one of the big factors in the breakdown of a bearing. For instance, in a gyro, I am sure that is one of our greater problems today.

J. E. Stern: What is the frequency of types of failures that you have encountered in the bearings that have failed in the Air Force program? Do most failures involve retainers, races, or materials? Are most failures by fatigue?

K. M. Glaesner: Forty percent fatigue of the raceways in rolling elements, forty percent due to mishandling of the bearings, and twenty percent having to do with corrosion.

J. E. Stern: How do you evaluate the remaining life of recycled bearings?

K. M. Glaesner: Only through experience. We keep a log on each bearing, recording the hours of operation each time it is removed. Examination of these bearings over a 30 year period has given us the knowledge of how many defects, how many pits, what size of indentation, etc. we can stand in order to provide enough operating hours in the bearing. It is by experience only, and to my knowledge we have never lost an aircraft engine because of one of these bearings. I hope some day that we can predict the service life with more accuracy. I think if we could measure the subsurface structure of the bearing and if we could calculate the amount of load, heat, RPM's and some of the environmental conditions of the bearing, we could design a bearing and predict a life expectancy of X number of hours and I think when that X number of hours came, the bearing would all collapse.

J. E. Stern: Do you use acoustic signature analysis?

K. M. Glaesner: We do on small bearings in generators and in similar auxiliary equipment.

J. V. Deller, Naval Ship Engineering Center: You indicated that 75% of the bearings are recycled. Is that the first recycle? How is the number of hours before the first recycle related to the L_{10} life of the bearing?

K. M. Glaesner: As I indicated before, we keep a log on bearing life so we can tell how many hours a bearing has been in service. With the exception of a few engine bearings, in no place do we approach the actual life design expectancy of a bearing. When the Barkhausen analysis and measurements of subsurface changes due to operational conditions are obtained, it may be necessary to decrease the life expectancy of the bearing.

J. V. Deller: Would you try to revise your design criteria to a higher L_{10} life?

K. M. Glaesner: No, most aircraft engines are well over our design requirements.

H. P. Jost, K. S. Paul Products, Ltd: To what extent do you use vibration analysis to determine the life that is still in the bearing? In some applications of larger bearings we now know through vibration analysis whether the bearing has 5 hours, 50 hours or 500 hours life still in it, so we can use it right to the very end of its commercial life.

K. M. Glaesner: Ten years ago we designed what we call an andarometer, and since it turned out to be a lousy design, I might as well tell you I designed it. Its purpose was to measure the vibratory loads and the frequencies of the vibrations in a bearing. We found it valuable for research and development, but not for actual bearing monitoring in the shops. We didn't know how to pick up the vibrations 10 years ago as we know today. But in the Air Force we do not use any mechanism to measure vibratory forces on bearings.

H. P. Jost: I don't know of any application in Britain in aircraft, but on rolling mills it's being used to lengthen the life of bearings. It cuts down on mechanical failures to a tremendous extent and it is used actually on the rolling mills themselves.

K. M. Glaesner: Battelle Memorial Institute has done some marvelous work in automating what happens within a bearing during its normal cycle of operation at a given speed. Because of automation, they can tell, for instance, the kind of vibration in the retainer, the thickness of the lubricant film, the drag of the retainer on the ball, and many other things simultaneously. I don't think that a bearing has really been examined until it has been subjected to a dynamic analysis.

T. L. Daugherty, Naval Ship Research & Development Center: We find increasing importance in the use of the andarometer in low noise bearings in submarine applications. We use it not only in failure analysis but also in inspection testing.

K. M. Glaesner: I can go to heaven in peace now.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Holographic NDI of P–3 Wing Plank Splices*

A. J. Koury

Naval Air Systems Command
Washington, D.C. 20361

M. J. Devine and J. F. Erthal

Naval Air Development Center
Warminster, Pennsylvania 18974

P. D. Bhuta

TRW Systems Group
Redondo Beach, California 90277

Continuous wave and pulsed laser holographic techniques were applied to inspect stress corrosion cracking in P–3 wing plank splices. Holographic results were compared with those of currently used ultrasonic inspection techniques. Holographic techniques successfully detected all the stress corrosion cracking areas that were identified by the ultrasonic inspection technique and, in addition, several other crack areas were located whose presence was later verified by dye penetrant inspection after careful disassembly of the panel joints. It was also shown that the holographic technique can locate areas of stress corrosion cracking without removal of paint and sealant materials. The currently used ultrasonic technique requires a paint and sealant stripping operation prior to inspection. Considerable savings in cost will be realized when inspection using holographic techniques is implemented. The holographic technique was demonstrated to Navy personnel at the Naval Air Rework Facility, Alameda, California, in an air rework environment. The concept of "structural signature" to verify structural integrity of complex aircraft structures is briefly described.

Key words: Aircraft component failure; continuous wave holograms; crack detection; holographic techniques; pulsed laser holograms; stress corrosion cracking; ultrasonic inspection.

1. Introduction

Naval aircraft are subject to stress corrosion cracking because of constant exposure to the ocean environment. A particular problem under investigation dealt with the Navy's P–3 patrol aircraft. The commercial version of the P–3 aircraft has also shown the presence of stress corrosion cracking. Approximately 900 wing plank splices have shown stress corrosion cracking as reported in references [1] and [2].[1] Stress corrosion cracking results in the necessity to repair wing fuel leaks. The current method of inspection requires ultrasonic inspection by a point probe. Prior to performing the ultrasonic inspection, it is necessary to strip the area completely of paint and render the surface clean, This technique provides satisfactory results; however, it is time consuming and expensive because it is necessary to inspect large areas of the wing, and also the stripping process necessitates repainting.

In the interest of cost savings, it is desired to perform the inspection without requiring the removal of the paint and the sealant material which is time consuming. To satisfy this requirement, holographic interferometry techniques were considered for this application. Holographic techniques have the capability of recording very small displacements. These techniques take advantage of the fact that a structural member with a crack or flaw will respond differently to a load than will a member which is not defective. These differences in response are evidenced by marked changes in the interferometric fringe patterns as they traverse a defective area. Based on the nature of the materials and the details of the structure, it was considered that holographic techniques offered a potential to inspect P–3 wing plank splices in a cost effective manner.

Using continuous wave holographic techniques in the laboratory, its ability to detect stress corrosion cracking was verified by comparing the holographic results with those of ultrasonic inspection. The presence of additional flaws or cracks indicated by the holographic techniques was verified by comparison with dye penetrant results after disassembly of the structural joint.

*Work performed under contract N62269–72–C–0400 for the Naval Air Development Center in support of Navy's Analytical Rework Program.

[1] Figures in brackets indicate the literature references at the end of this presentation.

Subsequent to the continuous wave holographic inspection of small P–3 wing plank panels, a larger wing plank panel was inspected using the pulsed ruby laser in a demonstration at the Naval Air Rework Facility at Alameda on December 5 and 6, 1972. The purpose of the demonstration at Alameda was to verify that inspection can be performed using pulsed laser techniques in a Naval Air Rework Facility environment and to brief Navy personnel on the potential use of holographic techniques. Details of these results and their comparison with ultrasonic and dye penetrant results, after disassembly of the structure, are described.

Use of the holographic techniques to perform inspection without requiring removal of the paint or sealant was later verified by performing the inspection on a panel on which four different thicknesses of paint were applied and a bare strip was left at the top of the panel. Cracks discovered in the earlier inspection showed up again even after the coats were applied.

A brief description of the pulsed laser holographic system capable of performing in situ stress corrosion cracking inspection on an aircraft in a Naval Rework Facility is given.

2. Description of the Problem

Figure 1 illustrates the geometry of the wing plank splice specimens taken from a P–3 Naval aircraft. Particular attention is drawn to the fillet and to the crack originating there despite the installation of sealant. The method currently used to locate these cracks is a point-by-point search with an ultrasonic probe.

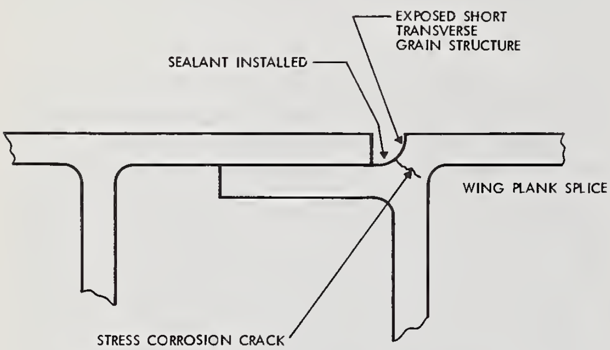


FIGURE 1. *Wing plank splice configuration.*

Figure 2 shows the original locations of the three wing panels. Because of apparent mislabeling these panels will be identified in this paper by their physical dimensions to eliminate ambiguity.

At the outset, it was postulated that a properly strained panel would exhibit localized structural weaknesses in the vicinity of cracked areas. Anomalies in deformations would be recorded as irregularities in the holographic fringe patterns.

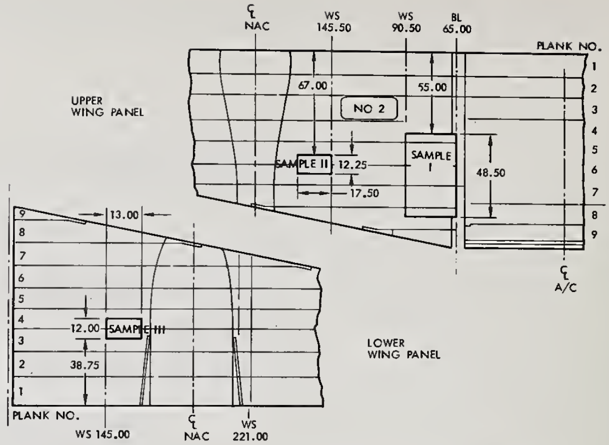


FIGURE 2. *Station sketch showing location of samples.*

3. Procedure for Continuous Wave Holographic Inspection

One of the smaller panels, 12″ x 13″, was clamped in cantilever fashion so as to be able to strain it by applying a simple bending load. It was expected that the application of a small preload before the first exposure would enhance the detection process.

The apparatus used for testing the small panels in the laboratory was as follows: one 15 milliwatt continuous wave Helium-Neon laser, one beam splitter, four mirrors, two spatial filters, one neutral density filter, one film plate holder, two right-angle blocks, assorted clamps, rised blocks, optics holders and a stable shock isolated mechanical platform. Figure 3 is a photograph of the actual setup. The various elements are arranged so as to provide a scene illuminating beam and a reference beam. The various elements and the optical paths followed by the reference and scene beams

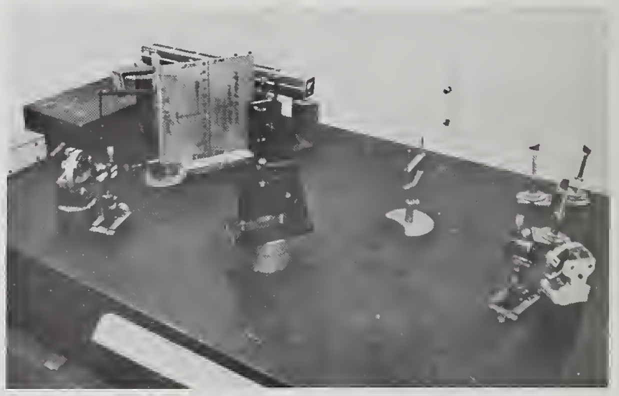


FIGURE 3. *CW holographic setup for testing P–3 wing plank splice.*

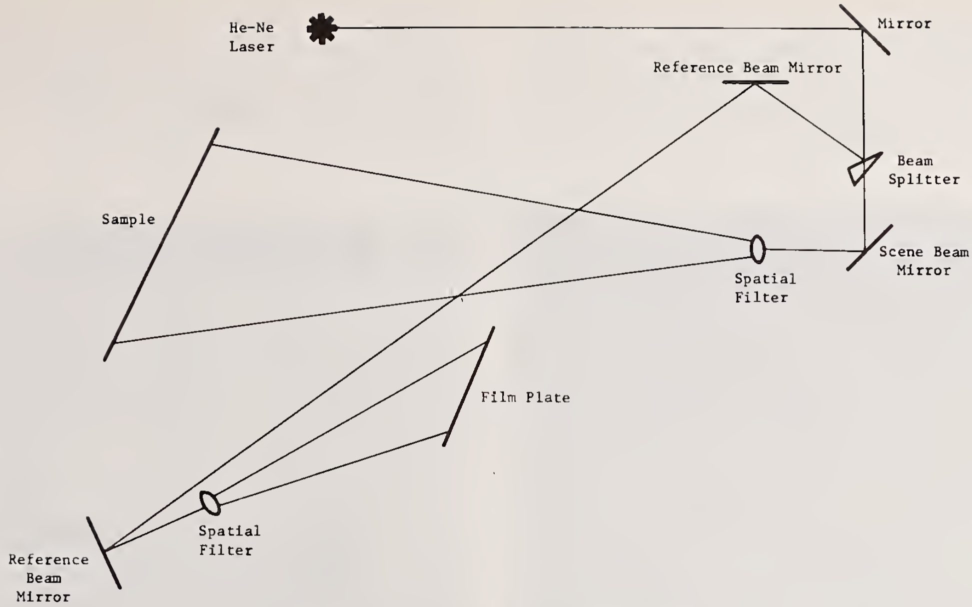


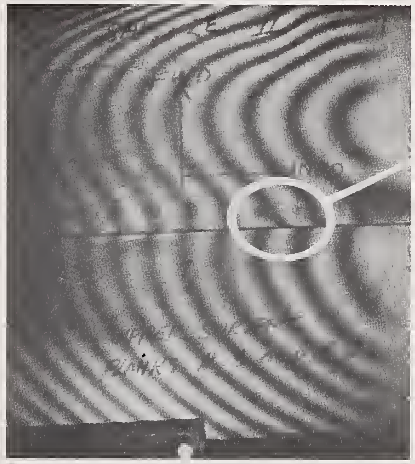
Figure 4. *Diagram of CW holographic setup.*

are shown in figure 4. The specimen was clamped to an angle block and was loaded in bending by means of a knurled machine screw at the right-hand edge of the specimen. Initial testing in the laboratory was performed using a Helium-Neon laser because it permits use of stored beam interferometry techniques which provide real-time inspection. A continuous wave Helium-Neon laser is not suitable for use in a Naval Air Rework Facility environment because it requires a greater degree of mechanical stability and isolation from the environment.

A holographic interferogram of the 12″ x 13″ panel was made by performing the following steps.

1. Arrange optical setup so as to match path lengths of the illuminating and reference beams.
2. Apply preload to specimen and make first exposure.
3. Apply small additional load, minute rotation of knurled screw, and make second exposure.
4. Develop, fix and wash film plate. Allow to dry.

A hologram made in this manner can be "reconstructed" by illuminating it with the reference beam only, impinging upon the hologram at the same angle employed in the original exposures. When the eye (or a camera) looks through the hologram, as if it were a window, it sees a virtual image of the panel sitting there in space, even though the panel is removed from the scene. Presence of a flaw is detected by examining the fringe pattern obtained in the holographic interferogram. Any anomaly in the fringe pattern, such as discontinuities in the fringes themselves and/or their slopes, indicates the presence of stress corrosion cracks.



CRACK INDICATED BY FRINGE DISCONTINUITY

Figure 5. *Small 12″ x 13″ panel subjected to continuous-wave holographic inspection.*

Figure 5 is a picture of an interferogram of the 12″ x 13″ panel. It was strained and holographed and the consequent fringe pattern contains a marked discontinuity, shown in the encircled area. The discontinuity in fringe pattern is identified by reconstructing the hologram and viewing the fringes in three dimensions.

The second of the two smaller panels measured $12\frac{1}{4}''$ x $17\frac{1}{2}''$. It was inserted into the same setup as the first panel and inspected in the same manner. Figure 6 depicts the holographic results. The fringe orders have been assigned numbers to aid the reader in tracing them across the joint. Four cracks were indicated by discontinuities in the fringe pattern and are shown in rectangular blocks marked 1, 2, 3, and 5.

The pulsed ruby laser system, which was used for this program, is shown in figure 7. This laser has the capability of producing up to four, long-coherence, optical pulses (wavelength, $\lambda = 6943Å$) within one millisecond. The separation between pulses can be varied from one to 250 microseconds.

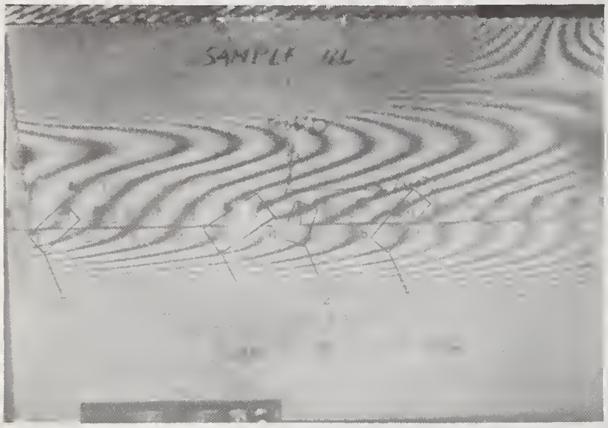


FIGURE 6. *Crack locations in 12.25" x 17.50" panel as indicated by holographic inspection.*
Cracks are indicated by fringe discontinuities shown in rectangular blocks marked 1 through 3 and 5. Numbers assigned next to fringes indicate fringe orders and serve as a means for improved data interpretation.

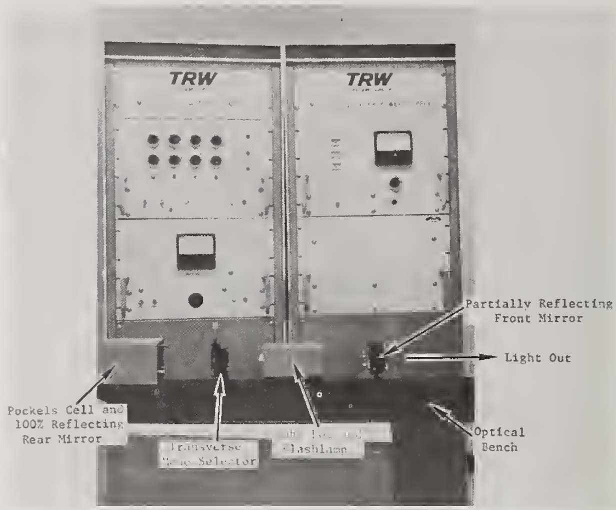


FIGURE 7. *Multipulse ruby laser system.*

4. Pulsed Laser Holographic Inspection of Larger P–3 Wing Plank Panel

In this section pulsed laser holographic inspection of the larger P–3 wing plank panel is described. To permit inspection in a Naval Air Rework Facility environment, pulsed laser holographic techniques were utilized. The section provides a description of double-pulse ruby laser holography, test setup, timing sequence and making of the holographic interferogram.

The double-pulse approach consists of exposing the film plate while the panel assumes an arbitrary "background" pose and then exposing it a second time after having impacted the panel at a low level by a ballistic pendulum. Both sets of wave fronts reflected off the panel are thereby captured on the same emulsion and interfere constructively and destructively, depending upon their relative phases at various stations on the hologram. This is, in turn, a function of the relative positions of the panel at the times of the first and second exposures. Upon reconstructing the hologram, interference fringes will be formed wherever the displacements equal odd numbers of quarter wavelengths.

The double-pulse system permits firing of two pulses in rapid succession, several microseconds apart. This time interval is so short that the panel cannot experience any appreciable displacements other than the ones induced by the impact. The pulse duration is approximately 50 nanoseconds, which permits "freezing" of the motion due to the induced stress waves.

The basic components of the ruby laser are two mirrors, a ruby rod, a flashlamp, and a Pockels cell Q-switch. These are shown schematically in figure 8. The two mirrors form opposite ends of the optical cavity. The ruby rod provides the optical energy (light) in a coherent, monochromatic form. The flashlamp "pumps" the ruby rod, making it ready to lase. The Pockels cells is situated between the mirrors and acts as a shutter for the cavity. When the proper voltage is applied to the Pockels cell, light is allowed to pass through it forming a complete optical cavity allowing the laser to lase. When there is no voltage applied, the Pockels cell isolates one mirror from the other and the optical cavity is no longer complete.

The actual operation of the pulsed ruby laser consists of a series of steps which must be performed in the proper order. The flashlamp requires considerably more power than can be supplied on a continuous basis so it is necessary to store energy for use during the pump cycle. This is accomplished by charging a large capacitor. The amount of stored energy is adjusted by controlling the voltage to which the capacitor is charged. Once the capacitor has been energized, the laser is ready to operate.

In order for the laser to lase, the ruby rod must be "pumped" by a Xenon flash tube. The pumping period approximates one millisecond, after which the ruby rod is capable of lasing. It is necessary, then, to initiate the pumping cycle before the ball impacts the specimen. The signal for this is provided by breaking the CW

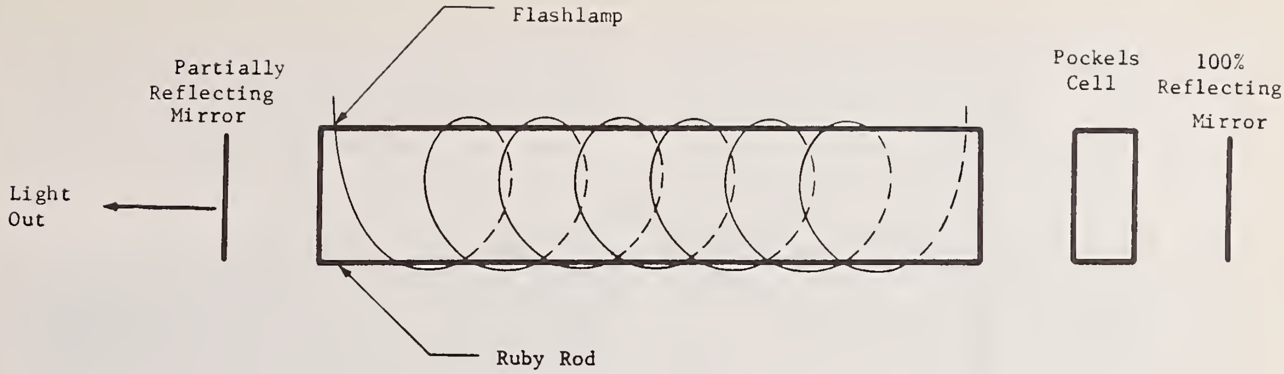


Figure 8. *Schematic representation of a pulsed ruby laser.*

laser timing beam with the ball as it nears the specimen. This signal is fed into a delay circuit where it is held in abeyance as the ball continues its swing. At the proper moment before impact, this signal is fed into the flashlamp circuit and the pumping cycle begins.

As the ball contacts the panel it acts as a mechanical switch, closing the path through the conducting support wire, the ball, and the panel. This circuit activates the Pockels cell control which in turn causes the laser to fire the first pulse. The stress wave is launched upon ball impact and the first pulse may be fired simultaneously or may be delayed to give the stress wave time to progress. At an arbitrary time later, approximately 10 to 300 μs, the second pulse may be fired capturing the interferogram on the film plate.

The large wing panel measured 25½″ x 48″ and contained four splices. It was clamped, cantilever fashion, to the left side of a support structure which was secured to the table on which the test was conducted. See figure 9 for a photograph of the setup. A small preload was applied to the panel by means of the small knurled screw shown at the lower right side of the panel.

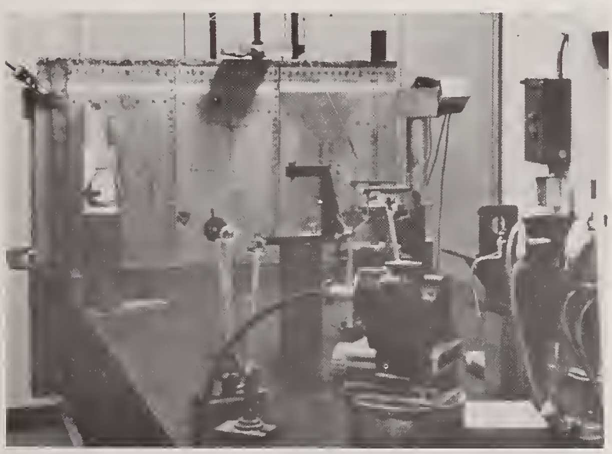


Figure 9. *Inspection setup.*

The ballistic pendulum was a 1″ diameter brass ball suspended by a wire which was 16¾″ long. It was caged and dropped by a special release mechanism seen at the top center of the picture in figure 9. A camera shutter cable can be seen dangling from this device, and is used to trip the release lever. The drop height of the ball was 6½″.

On top of the taller aluminum pedestal near the left side of the panel was a small rectangular mirror which was used to direct a one milliwatt CW laser beam across the face of the panel and into a photocell secured to the right end of the support structure. This served as a timing beam, the function of which is described previously in this section. The path of the timing beam and the positioning of the other optics are shown in the diagram of figure 10.

The optical elements for the illumination beam are, in order, one beam splitter and two negative lenses. The reference beam path incorporates the same beam splitter, a neutral density filter, a small mirror, a negative lens, an iris diaphragm and a final mirror. The holographic setup is completed by the pulse laser, the film plate holder, and the CW timing beam elements.

The hologram is made in the following manner. The pendulum ball is placed in the retaining nest, automatically cocking the release lever. The room lights are turned off (during an on-site inspection, the area to be inspected can be covered by light proof curtains) and a fresh film plate is placed in its holder. Following an oral warning, the capacitor bank of the pulse laser is charged. This takes only a few seconds. With a countdown of "3, 2, 1, fire," the operator then presses the release cable and the ball commences its swing. The event proceeds from this point according to the programmed timing sequence.

After the firing of the laser, the film plate is removed and placed in a carry box (before turning on the room lights). The film plate is then taken to the darkroom for developing, fixing, washing, and drying. The finished product is referred to as a hologram.

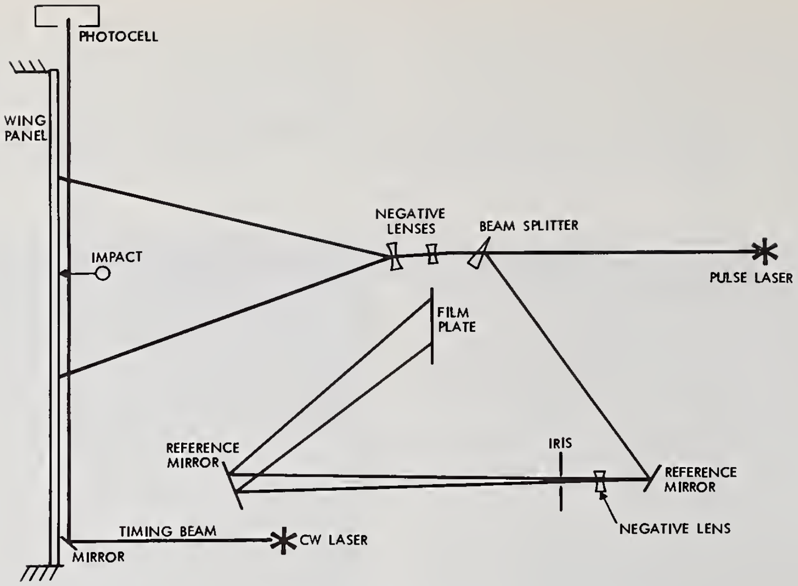


Figure 10. *Diagram of pulsed holographic inspection setup.*

5. Holographic Results of the Pulsed Laser Inspection

Several interferograms of the larger panel were produced which revealed a total of 10 marked discontinuities in the fringe patterns. One particular hologram was chosen for illustration in this paper in that it shows all 10 of these suspicious areas. Refer to figure 11 for a picture of this interferogram. Again, anomalies in the fringe pattern are enclosed in rectangular boxes and are numbered for ease in identification. Inspection was confined to the two center splices. The flaw location chart of figure 12 indicates the positions of the cracks along these splices as detected by the pulsed laser holographic system.

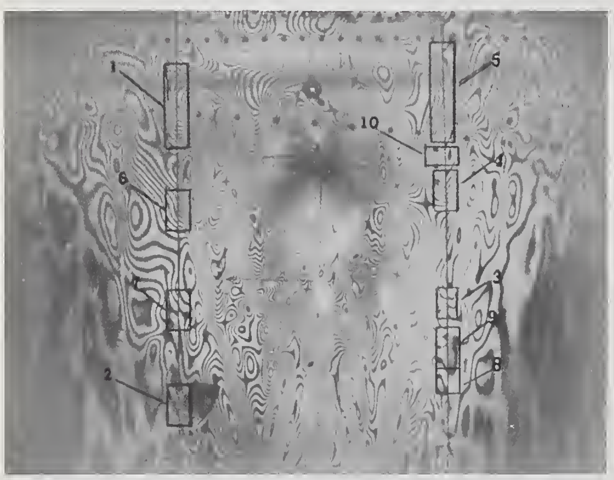


Figure 11. *Fringe pattern on 48.50″ x 25.50″ panel obtained using the pulsed holographic inspection technique.*
(The areas of stress corrosion cracking are indicated.)

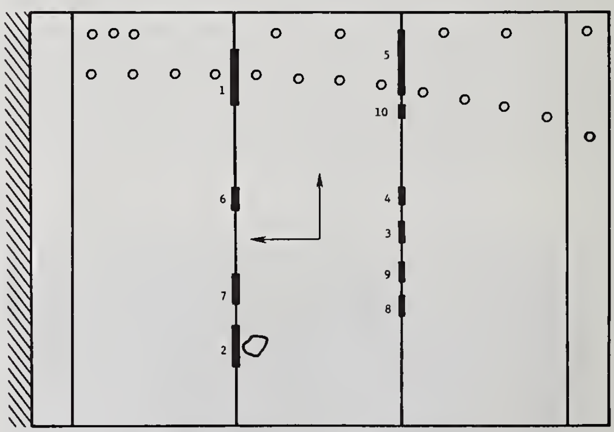


Figure 12. *Flaw location chart.*

6. Comparison of Holographic and Ultrasonic Results

Figure 5 shows the holographic results of inspection of the 12″ x 13″ panel where the discontinuity in the fringe pattern indicates presence of one crack. The same panel was tested by the Naval Air Rework Facility/Alameda personnel using the current ultrasonic technique. The results of ultrasonic testing are shown in figure 13. A comparison of figures 5 and 13 shows that both holographic and ultrasonic techniques located the same crack.

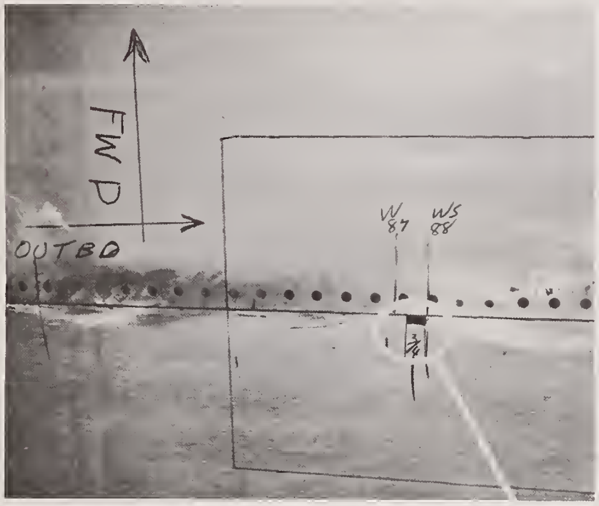


FIGURE 13. *Ultrasonic crack location in 12.00" x 13.00" sample.*

Results of holographic inspection of the 12.25" x 17.5" panel are shown in figure 6 where 4 crack locations are indicated by numbers 1, 2, 3, and 5 identifying rectangular blocks. The same panel was tested by the ultrasonic technique and the results provided by Naval Air Rework Facility/Alameda are shown in figure 14. The ultrasonic technique detected two cracks which were also detected by the holographic technique. However, the holographic technique indicated the presence of four cracks. To further investigate these suspected crack indications, the panel was carefully disassembled and subjected to dye penetrant inspection. The results of this comparison are provided in a later section.

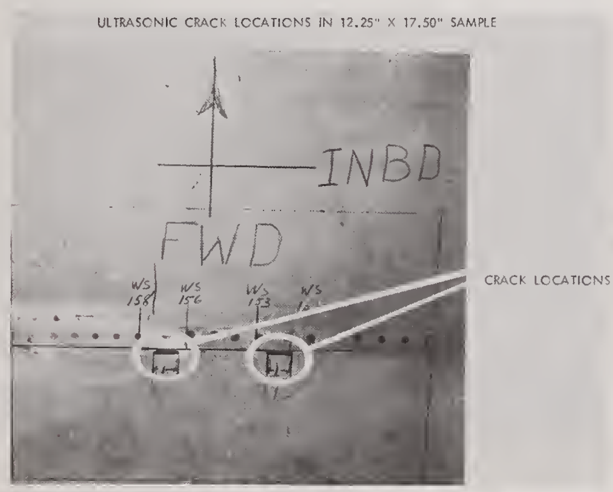


FIGURE 14. *Ultrasonic crack locations in 12.25" x 17.50" sample.*

Holographic results of inspection of the 48.5" x 25.5" panel are shown in figure 11. For ease in identification and location, these flaws are shown in figure 12. Figure 15 shows the locations of 4 cracks detected by the ultrasonic-system used by the Naval Air Rework Facility/Alameda. The flaws located by the ultrasonic system correspond to crack Nos. 1 through 4 shown in figure 12. Holographic results indicated presence of additional cracks and/or anomalous joint behavior. To further investigate this aspect the panel joints were disassembled and subjected to dye penetrant inspection.

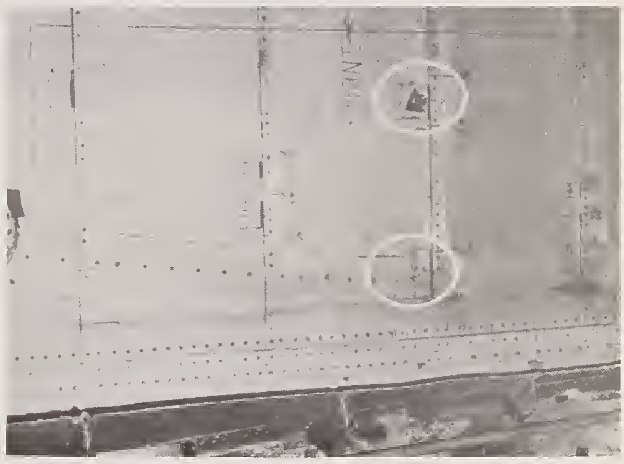


FIGURE 15. *Locations of the 4 cracks detected using ultrasonic inspection.*

7. Comparison of Holographic and Dye Penetrant Inspection Results

Figure 16 shows the results of dye penetrant inspection after disassembly of the 12.25" x 17.5" panel. In the figure the crack locations are identified by circled areas 1 through 5. Further magnified (5 times) views of each of the crack locations are shown in the same figure. Figure 16 shows that the cracks do not lie in the fillet but rather on the flat portion of the tank, a short distance away from the fillet. During the demonstration at the Naval Air Rework Facility/Alameda, a comment was made that the ultrasonic system currently used does not provide crack locations accurately. The current repair practice of removing the suspected area as indicated by the ultrasonic technique might miss the actual defective site and hence the defective area would not be repaired. A comparison of figure 16 with figure 6 shows that the holographic technique

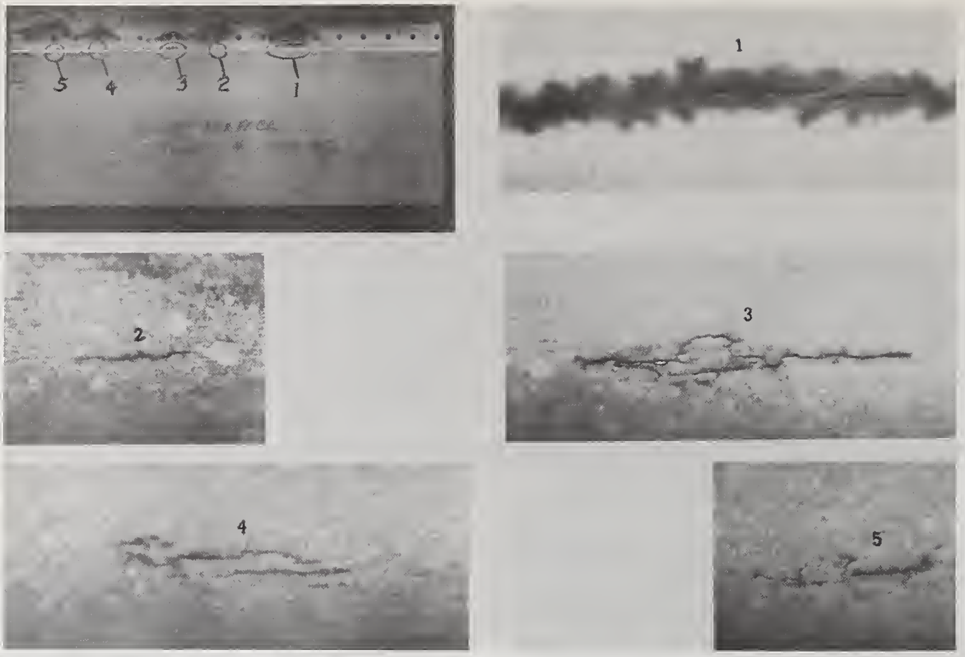


FIGURE 16. *Disassembly of the 12.25" x 17.50" panel and subsequent dye penetrant test revealed 5 cracks (shown individually in frames 1 to 5 at a magnification of 3×).*

detected only 4 of the 5 cracks. Crack No. 4 in figure 16 was not detected by the holographic technique. By developing an improved straining technique, it should be possible to obtain a more uniform and finer grid spacing of the interferometric fringes on the joint. Finer and more uniform grid spacing on the joint should provide improved sensitivity to detect the crack whose presence was not visible in the results obtained to date.

Figure 17 shows the photographs of the two center splice joints showing locations of stress corrosion cracks, upon dye penetrant inspection, after disassembly. For a comparison of the holographic results with dye penetrant results, it is more convenient to work with one splice at a time. The left photograph in figure 17 shows crack locations by circled areas 1, 2, and 6. Holographic results in the flaw location chart (fig. 12) indicate the presence of the same cracks and, in addition, an anomaly in the holographic fringe pattern at location 7. Since the dye penetrant results do not confirm the relationship of the fringe pattern irregularity at location 7 to the presence of a crack, it was suspected that the anomaly in the fringe pattern may be attributed to a loose rivet which is indicated by the irregular deformation pattern. This was later verified.

The right photograph in figure 17 shows the location of cracks determined by dye penetrant inspection and are indicated by Nos. 11, 9, 3, 4, 4a, and 10. The holographic results are shown in figure 12. Cracks 3 and 4 are located in both holographic and dye penetrant results. Cracks 8 and 9 detected to be separate cracks by the holographic results are both confirmed

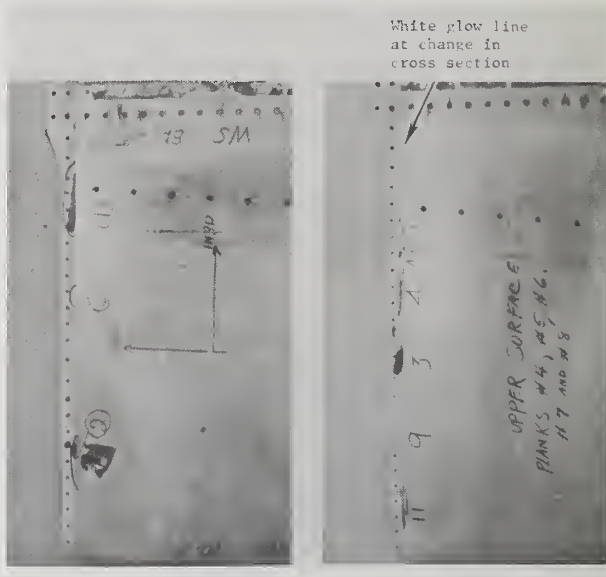


FIGURE 17. *Photographs of the two center splice joints showing stress corrosion cracks located by a dye penetrant inspection.*

to be present by the dye penetrant results as a single crack shown in location 9 in figure 17. Crack No. 10 is located in both holographic and dye penetrant results.

An unlabeled crack between locations 3 and 4 in figure 17 giving the dye penetrant results was not detected by holographic results. At the area of location

5 in figure 12 giving the holographic results, no crack was found by dye penetrant inspection. However, the presence of an anomaly in the fringe pattern, it is felt, results from the fact that there is an abrupt change in the cross-sectional thickness of the panel. This abrupt change in the cross-sectional thickness is indicated by a white glow line in figure 17. Since we would indeed have in our possession the drawings of structures we wish to inspect, we would have a prior knowledge if such behavior is to be encountered from sudden changes in thickness.

The dye penetrant result at location 11 in figure 17 does not indicate a crack but rather a "scab" of encrusted material.

In summary, the pulsed laser inspection technique of the 48.5″ x 25.5″ panel located 3 additional stress corrosion cracks which went undetected by the ultrasonic system and also indicated anomalous behavior of the joint at locations 7 and 5. The reasons for the anomalous behavior were postulated and verified by subsequent testing.

8. Effects of Surface Paint

The question arose as to whether a painted wing panel would respond as well to holographic inspection as a bare panel. To resolve the question, the one panel remaining intact, the 12″ x 13″ panel (the other two had been disassembled), was taken to Naval Air Rework Facility/North Island to be painted. Four different thicknesses of paint were applied and a bare strip was left at the top of the panel. Starting at the top of the picture of figure 18, the respective thicknesses, in horizontal strips, are: bare, 6, 12, 18, and 24 mils.

FIGURE 18. *Demonstration that surface paint does not degrade the quality of the interferometric fringe pattern.*
Original flaw still evidenced by discontinuity in fringe pattern in boxed area.

The panel was set up as before and strained in bending after an initial preload. The resulting fringes were sharp and of good contrast, showing that in no way does the paint degrade or hamper the holographic process. As pointed out earlier, the holographic interferometry inspection technique eliminates the necessity of stripping the wings as a part of the process of searching for stress corrosion cracking.

It will be noted also that the crack discovered in the earlier test shows up again in this picture. It is outlined by the upper rectangular box. Another short discontinuity appeared immediately below this one (lower box) which may be a separate crack or an extension of the effect of the original crack.

9. Description of a Portable Holographic System

A conceptual design of a portable holographic system for in-situ inspection of Naval aircraft wings is shown in figure 19. The design requires mounting the optical elements and laser head on an optical mounting plate as shown in figure 19. The power supply and the cooling system would be mounted on a transporter or on the floor in a Naval Air Rework Facility area.

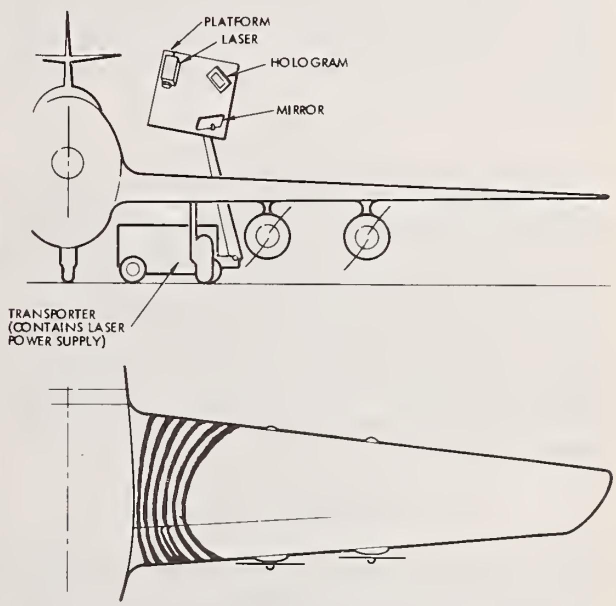


FIGURE 19. *Inspection setup in an air rework facility.*

The concept of "structural signature" to verify structural integrity of complex aircraft structures is described. In brief the concept is based upon comparing the *structure against itself* periodically as the service life of the structure progresses. As the structure experiences progressive duty cycles, changes in the "structural signature" of a highly redundant structure should result from fatigue or reduction in a cross-sectional area of a structural component. It is proposed to use the "structural signature" to verify integrity of a complex structure without its disassembly or any special surface preparation.

10. Conclusions

The following was accomplished:

(1) All the cracks previously detected by the currently used ultrasonic technique by Naval Air Rework Facility/Alameda in P–3 wing plank panel splices were detected by the holographic technique.

(2) Additional cracks were also detected whose presence was later verified by dye penetrant inspection after careful disassembly of panel joints.

(3) Stress corrosion cracks in P–3 wing plank splices were detected without requiring stripping of the paint prior to inspection. The confirmation was made by performing the inspection holographically with and without the panel painted with up to four coats of polyurethane paint. This feature of the inspection technique offers substantial savings because stripping and resealing operations, which are currently necessary when the ultrasonic technique is used, will be eliminated. Reduction in inspection time by a factor of 3 for location of stress corrosion cracks in P–3 aircraft wing plank splices should result when holographic techniques are used.[2] If automated film processing and data reduction techniques are developed and incorporated into the holographic inspection system, a reduction in inspection time by a factor of 5 may be achievable.

The authors wish to thank J. E. Wright, J. L. Jacoby, W. S. Tierney and E. K. Burchman of the Advanced Technology Staff, Space Vehicles Division, TRW Systems Group, for their providing the experimental results reported in the paper. The work described in the paper was performed under the sponsorship of the Analytical Rework Program, Naval Air Systems Command (AIR–4117) to verify structural integrity of Naval aircraft. The authors also wish to thank Naval Air Rework Facilities personnel at Alameda and North Island, California, for their cooperation.

11. References

[1] Case histories—stress corrosion and corrosion enclosure (4) to letter NARF–313–RAW, from A. W. Tucker, Naval Air Rework Facility/Alameda, to P. G. Bhuta, TRW Systems Group (12 June 1972).

[2] P–3A BUNO 151387, wing panel splices; excessive cracking of, Naval Air Rework Facility/Alameda from S. G. Kurz, Enclosure (5) to above letter NARF–313–RAW.

12. Discussion

A. J. McEvily, University of Connecticut: I was interested in the fact that you identified the cracks as stress corrosion cracks in the Electra and wondered how you made that determination in the sense that here is a structure that is subjected to cyclic loading? How do you know whether you have corrosion fatigue or stress corrosion cracking as your primary problem?

J. Erthal: This work was done by the engineering and materials group and a metallurgist. Experience in examining planes over the last 15 years has led to the belief that it's stress corrosion rather than fatigue. Of course it could be a combination; pitting corrosion can develop that leads to nuclei for the stress raisers that allow fatigue to enter the picture. But, they are usually identified as stress corrosion cracks.

A. J. McEvily: What led them to believe that they were stress corrosion cracks?

J. Erthal: The decision is based on a failure analysis by the metallurgical department.

D. C. Drucker, University of Illinois: I would think, to look at the effect holographically, that the loading would be equivalent to that which the aircraft had in service. I wonder why impact type loading was used for the hologram instead of the much simpler, and perhaps more direct, method of putting a vibrator at the end of the wing. I believe this would open the cracks in the proper direction while the impulse might or might not do that.

J. Erthal: It is planned to determine the best way of loading the wing.

C. E. Vest, NASA, Goddard Space Flight Center: What do you do after the crack is found?

J. Erthal: That depends on where it is and what kind of data we have for crack propagation for the particular aircraft. It is up to the stress analyst, the designer, and the materials people to decide from the fracture toughness data, or fracture mechanics data, to say fix it, take it out, repair it, and so forth.

C. E. Vest: Are the cracks large or very small?

J. Erthal: The smallest one was roughly a little less than a quarter of an inch. We found them up to one and one-half inches long.

C. E. Vest: How big does a crack have to be in order to ground an aircraft?

J. Erthal: That depends on the aircraft, where it is, and who says ground it, or what kind or operation it's used in. No doubt, I have seen aircraft fly with 3 or 4 inch cracks.

A. J. Babecki, NASA, Goddard Space Flight Center: Inasmuch as you are using lasers in the inspection system, can you discuss the safety aspects?

J. Erthal: The laser beam is similar to a welding arc. The light should be kept from the eyes. We follow

[2] Verbal communication between Naval Air Rework Facility/Alameda and TRW personnel. Sixteen h to prepare splice gaps, 40 h to ultrasonic inspect and 75 h to reseal gaps.

OSHA requirements as well as requirements of our own safety people.

J. K. Miska, Naval Ship Engineering Center: How often do you perform a holographic inspection? What is your confidence that if the crack does exist, you really find it? In other words, what is the reliability of that kind of inspection?

J. Erthal: The inspection is so far about 95 percent reliable, but as I said, this is a feasibility study and limited in extent. It does look feasible.

J. K. Miska: And how often do you perform these? Is it routine?

J. Erthal: If it proves satisfactory, it would be used as a routine tool when an aircraft comes in for rework.

CONSEQUENCES OF MECHANICAL FAILURE

Session III (cont)

Chairman: J. R. Strang
Air Force Materials Laboratory

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Economic Considerations in Failure Prevention

Howard E. Morgan

**Office of the Director
National Bureau of Standards
Washington, D.C. 20234**

Economic considerations normally determine if product failure prevention is justified by a producing firm. Some failures are not worth preventing, even if prevention is technically feasible. Since the risk of failure is inherent in all products, the expected benefits of failure prevention should be compared with expected costs, both discounted to present value in a life-cycle context and compared with alternatives.

Producers and users of a product may evaluate the risk of failure and its economic cost differently. This may lead to inefficient and inequitable production decisions affecting the public interest. Consequently, externalities (third-party effects) may require that regulatory or other agencies of Government intervene to protect the public and to insure socially efficient and equitable allocation of resources to failure prevention.

Failures represent a gap between expectations and performance. Both involve measurement and standards technology which the NBS can provide. However, economic analysis should be used to determine the conditions under which failure prevention is desirable. Examples of economically desirable conditions for failure prevention are provided.

Key words: Economic cost of failure prevention; frequency of failure; government action; life-cycle costing; marginal analysis; quality control; safety standards.

1. Failure Prevention Management Requires Economic Considerations

Some mechanical failures are not worth preventing. This statement may be heresy to an audience composed of professional people dedicated to failure prevention. Yet, upon reflection, most probably would agree with it. This paper provides arguments in defense of that statement. The central question addressed here is: "When is it economically desirable to prevent mechanical failures?" If this question can be answered, we have a basis for setting priorities and allocating resources to failure prevention. We can then sort the trivial from the important problems and perform needed research and engineering work more efficiently. Since zero defects (nonfailure) usually is an economically unrealistic objective, economic considerations should prevail in the management of a failure prevention program.

Mechanical failure prevention is a problem of importance to manufacturers of durable consumer goods and industrial goods which increasingly are becoming more complex. Servicing of these products involves the correction of defects often not recognized at the time of manufacture. These failures cost a firm money, both in warranty costs and the loss of repeat business. In addition, mechanical failures are receiving increasing public attention with numerous Government programs designed to improve public safety endangered by these failures. Automobiles and other consumer products are receiving particular attention. Economic losses in property and personal injuries—often involving passive third parties—have made failure prevention important to both private businesses and Government.

This discussion proceeds by indicating some of the possible economic consequences of failures. It is then argued that failure prevention can be managed by quality control techniques. This leads to a discussion of the definition of quality, the frequency distribution of failures, and the economic implications. Techniques for economic analysis in quality control are then presented and illustrated. The paper concludes with a brief discussion of the appropriate role of Government in failure prevention.

Specifically excluded from the paper are techniques of statistical quality control, theoretical issues of interest to economists, and a discussion of risk-benefit analysis to decide "how safe is safe enough." Rather, the focus is on those analytical techniques of value to business management in deciding which failures are worth preventing.

2. Economic Cost of Failures

Comprehensive statistical data on the economic cost of mechanical failures do not exist. Sellers do not wish to publicize their mistakes; this may account for the fact that we have only anecdotes rather than aggregate data on failures. Consequently, one is forced to use proxies to indicate the possible magnitude of the economic cost of mechanical failures. Some of these data appear in tables 1 through 6.

Some fraction of all accidental deaths undoubtedly are caused by mechanical failures. Table 1 indicates the number of deaths and death rates from accidents occurring in 1967. Of a total of 113 169 deaths, nearly half were attributed to automobile accidents. Another one-fifth was caused by accidental falls. If only one percent of these nearly 75 000 deaths were attributable to mechanical failures, 750 deaths would have been caused by mechanical failures in 1967. Table 2 indicates the total cost of accidental injuries and deaths in 1963. This amounted to nearly $10 billion dollars of which a large share is the discounted value of lost earnings of those killed or injured in accidents. If only 1 percent of these injuries were caused by mechanical failures, the economic cost would have been around $100 million dollars in personal injuries and deaths, exclusive of property damage.

TABLE 1. *Deaths and death rates from accidents, 1967.*

Cause all accidents:	Number of deaths 113,169	Death rate per 100,000 population 57.2
Motor vehicle	52,924	26.7
Railway	997	0.5
Aircraft	1,799	.9
Accidental falls	20,120	10.2
Miscellaneous		
Machinery	2,055	1.0
Fire and explosion	7,423	3.8
Firearms	2,896	1.5
Drowning	5,724	2.9
Poisoning	4,080	2.1
All causes [a]	1,851,323	935.7

[a] Mostly due to diseases.
Source: Statistical Abstract of the U.S., 1970, p. 57

TABLE 2. *Total economic cost off accidental personal injuries, 1963.*

Cost of illness	Amount (mil. dollars)
Direct expenditures	1,703
Indirect costs [a]	1,811
Mortality cost [a]	6,427
Total cost	9,941

[a] Lifetime earnings discounted to present value at 6 percent
Source: HEW, Estimating the Cost of Illness, Health Economics Series, No. 6, 1966, Table 32

Where do accidents occur? Most injuries occur to persons at home or at places other than at work. Of nearly 50 million persons injured in accidents in 1968, 20 million were injured at home and another 20 million in motor vehicles, or at other places. This suggests that mechanical failures in equipment used in the home or at locations other than work may be worthy targets for investigation (table 3).

TABLE 3. *Accidental injuries by circumstances, 1968.*

	Persons injured (mil.)
While at work	9.3
While at home	20.5
Other (includes motor vehicles)	20.8
	49.0

Source: Statistical Abstract of the U.S., 1970, p. 82

The incidence of injuries in industry is indicated in table 4. The metal producing and forming industries, of particular interest to this audience, are not especially high in injury frequency or severity rate. On the other hand, the extractive industries, especially coal mining, is hazardous. The construction and transportation industries also have a high injury frequency rate, some portion of which may be due to mechanical failures.

Injuries occurring at home are now receiving the attention of the new Consumer Product Safety Commission. Among the more hazardous sources of injuries, thus far identified, glass bottles and glass doors rank conspicuously (table 5). Undoubtedly, some of these injuries are attributable to mechanical failures in glass. Perhaps this audience should give attention to this problem, along with mechanical failures in other materials.

A somewhat different way of viewing mechanical failures appears in table 6. Between 1960 and 1970, the cost of automobile repairs as percent of depreciation declined from about two-thirds to one-half. To the extent these data reliably indicate the magnitude of the economic cost of mechanical failures in automobiles, the auto industry appears to have made progress in reducing the economic burden of these failures on consumers, if not in the inconvenience of required repairs [1].[1] Similar analyses might be performed for other types of consumer durables and industrial equipment where mechanical failures may have considerable economic cost. (Life-cycle costing discussed below provides a technique for comparing the lifetime cost of alternatives.)

To identify industries using metals and other materials subject to mechanical failure, the Input-Output table for the U.S. economy is useful [2]. This permits one to measure the dollar value of purchases of materials by using industries; if the failure rate of that material were constant in all applications, one would have an indication of the magnitude of failure-relative economic losses in all industries using that material. For example, if the impact of copper corrosion were in proportion to the copper purchases by the industries using primary copper in 1963 (industrial chemicals $11.1 million, blast furnace products $12.5 million, iron and steel foundries $6.9 million, primary copper manufacturing $659.2 million, copper rolling and

[1] Figures in brackets indicate the literature references at the end of this presentation.

drawing $684.1 million, nonferrous wire $257.8 million, pipe valves and fittings $45.7 million) this would indicate where copper corrosion is apt to be found a problem. Of course, this offers only a first approximation to the incidence and magnitude of copper corrosion because a corrosive environment (e.g. chemicals) may cause higher corrosion losses in some industries than others.

TABLE 4. *Injury frequency rates and severity measures: selected industries, 1968.*

Industry	Frequency rate injuries per mil. man-hours	Severity rate days per mil. man-hours
Manufacturing:		
Food products	26.8	1,023
Lumber and wood	36.1	2,973
Furniture	22.3	945
Stone, clay, glass	21.6	1,212
Primary metals	15.2	917
Fabricated metals	21.1	924
Mining:		
Coal mining	40.8	9,859
Metal mining and milling	21.1	3,227
Nonmetal mining	22.5	3,010
Sand and gravel	20.2	2,440
Stone quarrying	16.6	2,505
Construction	26.9	1,992
Transportation:		
Passenger	23.6	1,220
Motor freight	31.7	1,821
Services:		
Auto repair	14.5	568
Misc. repair	20.1	710
State hospitals	20.0	535

Source: Statistical Abstract of the U.S., 1970, p. 237

TABLE 5. *Personal injuries by class of consumer product, 1969.*

(Hospital emergency room reporting)

	Frequency/severity index	Estimated extra cost to reduce hazard
Glass bottles and containers	127.7	2/10¢ per bottle
Floors and flooring materials	119.8	N.A.
Outside walls and fences	110.7	N.A.
Bicycles	96.0	N.A.
Stairs	83.2	N.A.
Gas furnaces	70.1	N.A.
Structural glass and glass doors	40.9	$10 per door

Source: National Commission on Product Safety: Final Report, 1970, pp. 40 and 68
N.A.: Not available.

TABLE 6. *Estimated 10-year cost of automobile depreciation and repair: 1960, 1968, 1970.*

	1960	1968	1970
Depreciation	$2,542	$2,806	$3,185
Repairs and maintenance.	1,722	1,788	1,521
Repairs as percent of depreciation.	68	64	48

Source: Statistical Abstract of the U.S., 1970, p. 547

3. Failure Prevention = Quality Control

The title of the group sponsoring this conference reflects a pessimistic viewpoint. Failure prevention implies a pessimism which can be translated into the more optimistic term, "quality control." Technical know-how and effective quality control management can prevent those failures which are worth preventing. Quality control represents the optimist's point of view [3].

Statistical quality control relies heavily on the use of probability sampling to estimate the characteristics of a group from which the sample is drawn. We shall not discuss this technique here; it is important only to note that some variability is inherent in the characteristics of any product. Quality control for failure prevention should measure and recognize the implications of that variability.

There are two concepts of quality which must be distinguished in quality control for failure prevention; table 7 identifies and illustrates these concepts. Design quality is a specification a product must meet in a given use; it may be expressed either as a set of product attributes or as a set of performance requirements for a particular use. Design quality is focussed on user needs. In contrast, conformance quality indicates the extent to which a product meets the design requirements; i.e., to what extent do the physical attributes of the product comply with design specifications? The term "standard" may refer to either the quality of design, considering use, or the extent to which a product conforms to design.

Shewhart, in his classic *Economic Control of Quality of Manufactured Product*, neatly illustrates this distinction graphically as illustrated in modified form in figure 1 [4]. In figure 1a., design quality is represented as a three-dimensional box having specified limits or dimensions. A product whose attributes lie within the box meets the specified design quality or standard. In figure 1b., conformance quality is illustrated by a particular point lying entirely within the box. A mechanical failure could occur because of either a design failure or the failure of a product to conform to design.

Table 7. *Quality of design versus quality of conformance.*[a]

Term	Defined	Example	Source of information for quality control
Design quality	A specification for a given functional use.	"Grades" of steel differ with use. Value to user may depend on grade of steel.	Users of product and other sources external to producer.
Conformance quality	The fidelity with which a product conforms to the design specification.	Steel of a given grade which meets the design specification for that grade of steel.	Internal to producing firm—laboratory or production department.

[a] A "standard" may refer to design quality (described by physical characteristics or by performance in use) *or* to conformance quality.

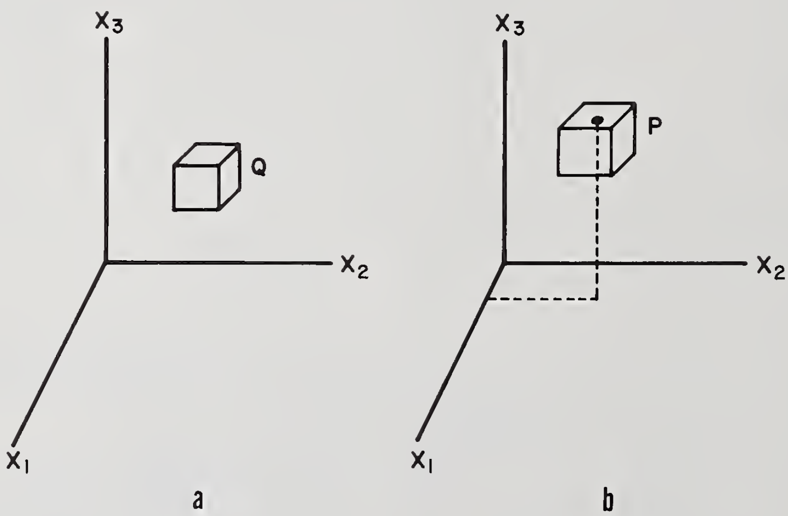


Figure 1. *Quality of design versus quality of conformance.* (*Source: Shewhart, page 40.*)
(a) Quality of design is represented by a multi-dimensional space with specified limits.
(b) Quality conforms if a point falls within design space.
$Q = X_{1\,(1\ldots n)}, X_{2\,(1\ldots n)}, X_{3\,(1\ldots n)}$ $P = X_{11}, X_{21}, X_{31}$

4. Distribution of Failures

Experience indicates that there are a few important failures and many trivial failures [5]. A frequency distribution of economic losses or severity of failures typically is skewed and described by one or more of the familiar mathematical functions. Figure 2 illustrates one such distribution described by the Lorenz curve which indicates the cumulative percent of failure losses plotted against the cumulative percent of quality defects. For example, 80 percent of warranty costs may be due to 20 percent of all of the defects in an automobile.

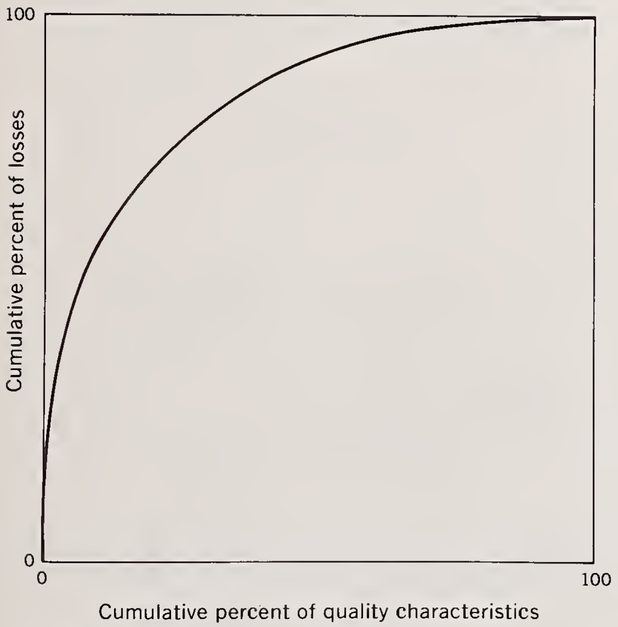


Figure 2. *Pareto- (Lorenz-) type curve NOTE: The Lorenz curve has incorrectly been called a "Pareto" curve by quality control specialists.*
(Source: Joseph M. Jurin and Frank M. Gryna, Jr., Quality Planning and Analysis, copyright 1970 by McGraw-Hill Company. Used with permission of McGraw-Hill Book Company.)

The typical frequency of failures over time is illustrated in figure 3 with the exponential distribution. A high proportion of defects will appear early in the life of a product; the frequency of defects then declines as the product ages [6]. This suggests that efforts to correct defects early in the life of a product may be economically more promising. However, without information about the value of prevention and the cost of correcting defects, one cannot assert that this is true. These and similar statistical distributions suggest, however, that not all failures are equal in magnitude and therefore, the prevention of some may be economically more worthwhile than others.

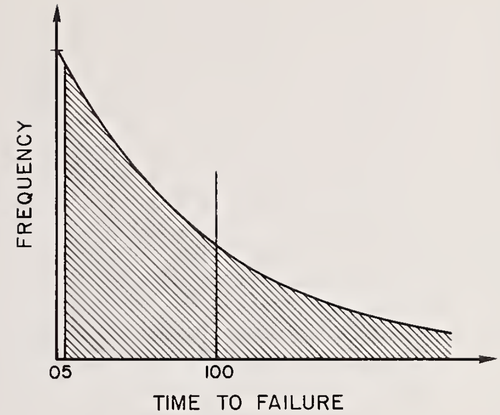


Figure 3. *Distribution of time of failure.*
(Source: Joseph M. Jurin and Frank M. Gryna, Jr., Quality Planning and Analysis, copyright 1970 by McGraw-Hill Company. Used with permission of McGraw-Hill Book Company.)

5. Economic Analysis for Quality Control

In deciding how to allocate resources for failure prevention and quality control, familiar techniques of economic analysis are applicable. Specifically, marginal analysis and life-cycle costing are useful techniques illustrated in figures 4, 5, and 6, and tables 9 and 10.

First, however, table 8 indicates the elements of value and cost which should be considered in quality control. It is necessary to obtain information on some of these items before the techniques of marginal analysis and life-cycle costing can be used to make choices.

Table 8. *Components of the value and cost of quality.*

Value of quality considers user's	Cost of quality considers producer's
—Knowledge of product and market	Quality design costs: —Market research costs —R&D costs for product —Advertising quality features
—Service needs over time	
—Aesthetic preferences	Conformance costs: —Planning and maintenance of quality control processes and machines
—Safety and health	—Quality control inspection and action to correct defects
—Life cycle cost	—Warranty and other costs due to defects or failures

Note: Assume quantity sold is given, unit price may vary with quality.
Source: Juran, pp. 42–46.

Table 8 suggests that the value of quality (design quality) must consider the users needs and wishes. This includes both economic and noneconomic considerations facing the user of a product. The cost of quality includes those costs associated with design and the cost of insuring conformance with the specified design. To the extent that these value and cost elements can be identified and appropriate data obtained, the analytical techniques which follow can be used.

An economically optimum level of design quality is illustrated in figure 4. As the quality of design is improved as one moves on the horizontal axis from 1 to 3, the value of quality first increases more rapidly than the cost of quality, then declines. Area B is greater than Area A, indicating that the marginal value of quality improvement exceeds the marginal cost of quality control as one moves from design quality 1 to 2. At 2 a maximum difference (i.e., profit) exists between the value and cost of quality. Between design quality 2 and 3, the marginal value of quality improvements represented by Area D is smaller than the marginal cost of quality control in Area C.

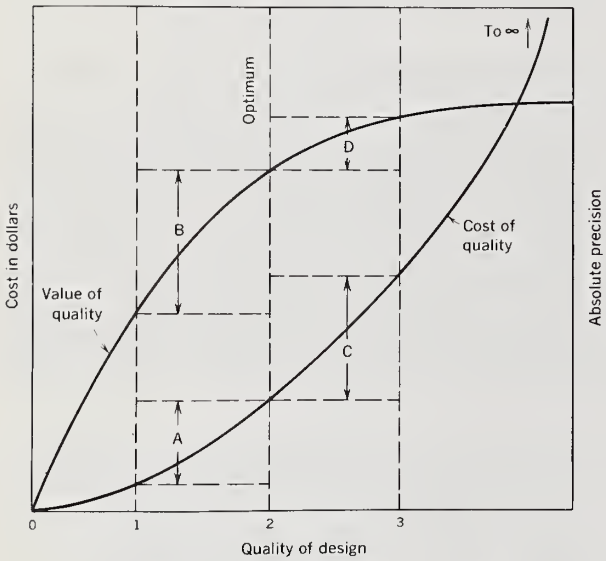


FIGURE 4. *Economics of quality of design.*
(Source: Joseph M. Jurin and Frank M. Gryna, Jr., Quality Planning and Analysis, copyright 1970 by McGraw-Hill Company. Used with permission of McGraw-Hill Book Company.)

Marginal analysis similarly can be used to decide upon the best level of quality conformance as illustrated in figure 5. As the percentage of defects declines along the horizontal axis, the cost of quality control rises first slowly, then rapidly to infinity as one tries to obtain zero defects. At the same time, however, the cost of defects is high at a level of 50 percent or more defects, but declines rapidly to zero with zero defects. Where the cost of quality control and the losses due to defects intersect, the total cost of quality conformance is at a minimum. This point lies somewhat to the left of zero defects.

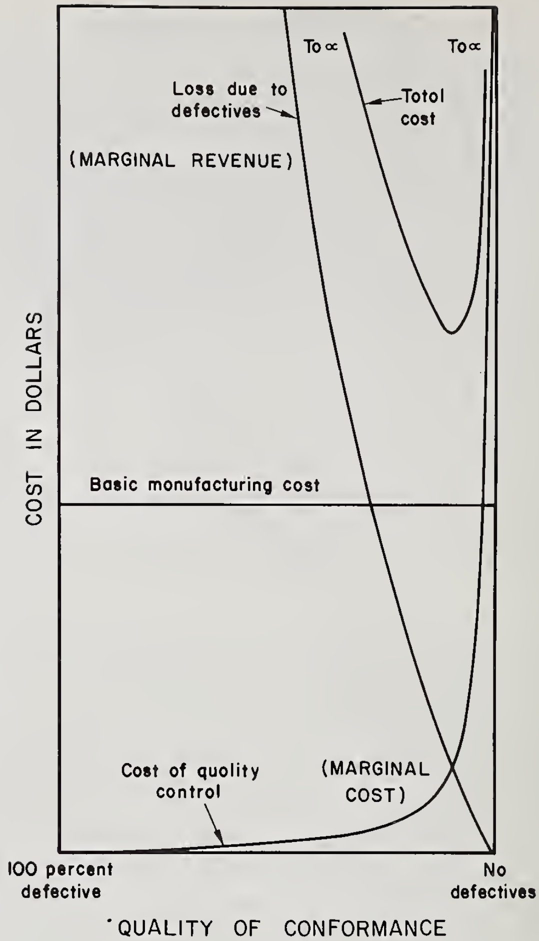


FIGURE 5. *Economics of quality performance.*
(Source: Joseph M. Jurin and Frank M. Gryna, Jr., Quality Planning and Analysis, copyright 1970 by McGraw-Hill Company. Used witb permission of McGraw-Hill Book Company.)

Frequently, a manufacturer is faced with defects in each of several products which may be components of a finished product. Table 9 illustrates a benefit-cost analysis for quality control of defects in five products (A through E). The second column indicates that products A and B have the major share of all defects, but the present annual cost of these defects (column 3) is not as great as for products C and E, which account for only 14 and 9 percent of all defects, respectively. The probability of success in quality control is indicated in the fourth column for each of the five products. Multiplying this likelihood of success by the present annual cost of defects, one obtains the data in the fifth column. Products C and E still offer the greatest potential savings in quality control. But, the five products differ in their expected life as indicated in the sixth column. Recognizing this fact by calculating the expected annual value of quality control over

Table 9. *Benefit-cost analysis for quality control of a set of products.*

Product	Percent of all defects	Present annual cost of defects	Probability of success in control	Expected annual value of control	Expected product life (yrs)	Discounted present value of control 10 percent rate	Discounted present cost of control 10 percent rate	Net benefit	Priority
A	32	$ 35K	0.70	$ 24.5K	10	$ 150.4K	$200K	$ −49.6K	5
B	25	45	.85	38.3	8	204.1	70	134.1	2
C	14	300	.65	195.0	3	485.6	400	85.6	4
D	15	45	.50	28.5	7	138.7	10	128.7	3
E	9	240	.80	192.0	5	727.7	300	427.7	1
Total	100	$665K		$472.3K		$1,706.5K	$980K	$726.5K	

the expected product life of the five products, one obtains the discounted present values shown in the seventh column. Products C and E still account for the major potential value of quality control, but products B and A have increased in relative importance because of their longer expected life. The eighth column indicates what it will cost in present dollars to control the quality of each of the five products. Net benefits (column 9) indicate that quality control of E may yield large dollar benefits, while quality control of A may incur losses. The last column indicates the priority which should be assigned to quality control of each of the five products. Note that product E, which has the lowest percent of all defects, may yield the largest dollar benefit from quality control.

One of the important user considerations noted in table 8 is the life-cycle cost of operating alternative pieces of equipment. Table 10 illustrates the application of life-cycle costing to two machines which have the same design quality. Machines A and B each meet the same performance standard or design quality in use.

Table 10. *Life cycle cost to user of two machines having the same design quality but differing in conformance quality.*

	Machine A	Machine B
Initial cost	$5 mil.	$6 mil.
Annual operating cost (includes down time)	$0.3 mil.	$0.2 mil.
Service life	40 years	50 years
Salvage value	$0.1 mil.	$0.5 mil.
Discount rate	10 percent	10 percent
Life cycle cost [a]	$7.87 mil.	$7.74 mil.

[a] Based on formula: $LC_o = FC + \sum_{t-1}^{l} \left[\frac{OC_t}{(1+r)^t} \right] - \frac{S_l}{(1+r)^l}$

Where:
- LC_o = Present value
- FC = First or initial cost
- OC = Annual operating cost
- l = Service life of machine
- S_l = Salvage value at end of life
- r = Discount rate
- t = Time in Years of Operation

However, Machine B has higher initial cost, but lower annual operating cost than Machine A. The life-cycle cost of Machine B is less than that of Machine A, considering the differences in initial cost, operating cost, service life, and salvage value as calculated with the formula shown in table 10. Buyers of industrial equipment and consumers buying durables frequently consider only the initial cost of the product. A failure to recognize the other elements of cost (including the cost of defects) may yield an incomplete picture of the total cost of alternative products [7]. Figure 6 illustrates graphically the tradeoff existing between the acquisition or initial cost of equipment and the net operating cost. The intersection of these cost curves indicates the optimum tradeoff and minimum life-cycle cost. In realistic applications of life-cycle costing in a system involving many elements, the tradeoff analysis becomes much more complex. However, the principles of marginal analysis and life-cycle costing remain valid.

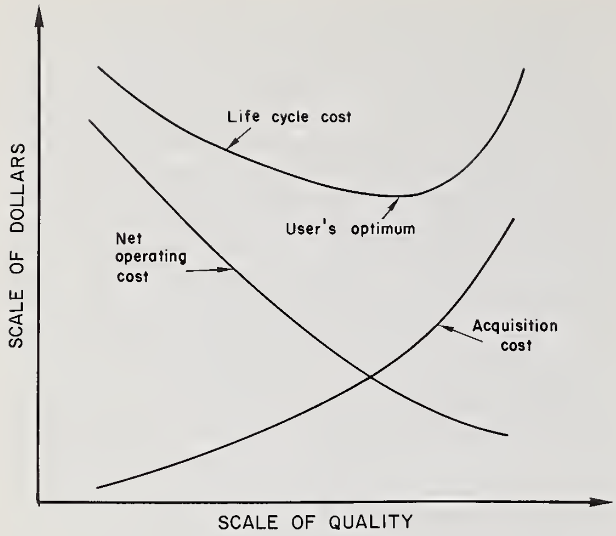


FIGURE 6. *Life cycle cost of quality.*
(Source: Joseph M. Jurin and Frank M. Gryna, Jr., Quality Planning and Analysis, copyright 1970 by McGraw-Hill Company. Used with permission of McGraw-Hill Book Company.)

6. Role of Government in Failure Prevention

One may ask why a paper dealing with economic considerations in failure prevention should discuss the role of Government in preventing failures. The answer is that business firms operating in the marketplace frequently cannot or do not allocate resources efficiently in failure prevention. Apart from humanitarian considerations, failures which threaten public safety represent an economic loss [8]. It then becomes necessary and desirable for Government to intervene in some manner. In short, a more efficient allocation of resources and greater public satisfaction may result from Government action to correct imperfections in the marketplace.

Some rules (criteria) for deciding when Government action is appropriate appear in table 11 [9]. There are three basic conditions or criteria for Government involvement, either in the form of Government programs which produce services related to failure prevention or Government policy governing private business.

TABLE 11. *Criteria for Government action in failure prevention.*

1. When private firms cannot capture in the market for their product all the benefits (disbenefits) which accrue to third parties.
2. When production by decreasing cost industries leads to monopoly and output less than socially optimal.
3. When imperfect knowledge about production or use discourages socially desirable high risk investments by private firms.

First, Government action is appropriate when private firms cannot capture in the marketplace all of the economic benefits which may accrue to third parties other than the initial users of their product. For example, efforts by the Ford Motor Company to sell safety several years ago did not appear successful because the public was not aroused to the need for safety and because buyers of safety equipped automobiles may not have been willing to pay a premium for safety benefits which may accrue to other persons as well as the users of the vehicle. Similarly, firms frequently do not charge buyers of their product for adverse economic harm which accrues to third parties. For example, a railroad may not charge its customers for the cost of injuries inflicted on third parties occasioned by a train wreck brought about by mechanical failure in wheels or track. If third parties suffer economic losses without being compensated by users of the railroad, Government intervention may be necessary to insure equity.

Secondly, Government action is appropriate when an industry has decreasing unit costs at an output less than that which is socially optimal. The best example of this is the public utility selling electric or gas service. These firms obtain economies of scale and lower unit cost as their systems expand and they are granted a monopoly position to service a given area. If left to their own decisions, their prices probably would be higher and their level of service lower than the public might wish. In such cases, Government regulatory action is appropriate to insure reasonable pricing and service levels leading to more efficient resource use. For example, the Federal Power Commission regulates gas pipeline pricing in the public interest. Recently, pipelines have become a matter of public concern because of mechanical failures which have resulted in damage to persons and property. Government regulation is then appropriate (under the first but not second rule) to insure public safety.

Thirdly, Government action is appropriate when private firms are discouraged from making high risk investments in socially desirable projects because they lack the resources or knowledge about the costs of production and payoff from such investments. This is illustrated by NASA research to reduce defects in space vehicles when the space program was considered to be important to the U.S. Similarly, the Department of Housing and Urban Development and National Bureau of Standards research in buildings is intended to reduce failures which may affect public safety or waste building materials. Private firms in the construction industries, left to themselves, may be discouraged from making the high risk investments necessary to provide technology needed to reduce failures. In such cases, Government acts as a risk-taker in providing technical service to protect public safety and improve the use of materials.

Many of the technical activities at the National Bureau of Standards assist industry and Government in quality control to prevent mechanical failures. For example. Standard Reference Materials which provide a basis for accurately measuring the metal content of engine oil may help prevent mechanical failures by monitoring engine wear; should wear become excessive, engines may be overhauled sooner than normally and failures avoided. Similarly, Standard Reference Materials are available for determining the composition of steel and other products. These assist in insuring conformance quality; i.e., the extent to which the product complies with a given standard. The Standard Reference Data system provides evaluated data on the characteristics of materials which helps design engineers determine design quality. Some of these reference materials and data are sold to users at a fee when their benefits are capturable by users in the marketplace. When some users or beneficiaries cannot be excluded from the benefits in the marketplace, it is appropriate for the Government to supply these at the expense of the general taxpayer, provided the economic benefits exceed the economic cost relative to alternative investments in public funds. In reality, many of these technical activities are carried out cooperatively by industry and Government.

7. Summary

Mechanical failures may impose substantial economic losses on business firms and the public. No aggregate statistical data exist to measure these losses; however, to the extent that accidental injuries and deaths are due to mechanical failures, the economic losses may represent a substantial sum. This is an economic loss which may justify much of the technical effort to prevent mechanical failures.

But, not all failures are worth preventing. Typically, there are a few large and many trivial failures. This suggests that the prevention of some failures may be worth more than others.

The techniques of quality control are useful in managing technical work in failure prevention. These techniques are well established and rely upon the disciplines of statistics and economics. Quality control relies heavily on the concept of probability sampling to estimate the characteristics of a group from which the sample is drawn. It also requires that a distinction be made between design quality and conformance quality. Design quality refers to the characteristics of a product in use and may be stated in terms of standards. Conformance quality refers to the extent to which a product complies or conforms with the given standard or design quality.

Economic analysis for quality control relies on the techniques of marginal analysis and life-cycle costing. Typically, there is an optimal level of quality at which the value of design quality is greatest relative to its cost. Similarly, the cost of conformance quality usually reaches a minimum at some point less than zero defects; i.e., it is usually not economically feasible to obtain zero defects. Life-cycle costing permits a comparison of two or more alternatives available to users, considering all costs over the life of the product. This method is preferred over a consideration of initial cost only. These techniques permit one to decide which failures are worth preventing so that technical and other efforts to reduce mechanical failures can be more effective in reducing the economic losses incurred.

Government action to help reduce failures is appropriate when public safety is endangered by the inability of firms to capture in the marketplace the benefits or disbenefits accruing to persons other than users of their product or service.

8. References and Notes

[1] One estimate of the economic cost of failure of five common automobile parts is four billion dollars per year or $400 over the average vehicle life. This source also notes the inadequacy of data on vehicle parts failures. See: Booz Allen Applied Research, Maintainability and Repairability of Vehicles in Use: Volume I Summary (National Technical Information Service, U.S. Department of Commerce, June 1971, Report # PB–202 531), pp. III–17 and IV–2.

[2] Bureau of Economic Analysis, U.S. Department of Commerce, The Input-Output Structure of the U.S. Economy: 1967, Survey of Current Business (February 1974), p. 24.

[3] In 1748, Voltaire defined an optimist as a madman who maintains that everything is right when it is wrong.

[4] Shewhart, W. A., Economic Control of Quality of Manufactured Product, (New York, D. Van Nostrand Co., 1931), p. 40.

[5] Juran, Joseph M. and Gryna, Jr., Frank M., Quality Planning and Analysis (New York, McGraw Hill Co., 1970), p. 63.

[6] Nelson, Wayne, Charts for confidence limits and teests for failure rates, Journal of Quality Technology (October 1972), p. 190.

[7] For a discussion of product durability and service as affected by the competitive structure of producing industries see: Richard W. Parks, The demand and supply of durable goods and durability, American Economic Review (March 1974), p. 37.

[8] There is considerable economic literature on the economic cost of safety and legal aspects of liability. See for example: Simon Rottenberg, Liability in law and economics, American Economic Review, March 1965, p. 107 and note by T. E. Borchering on that article in the December 1970 issue of that Journal, p. 946. See also, R. A. Holmes, On the economic welfare of victims of automobile accidents, American Economic Review (March 1970), p. 143.

[9] See: Richard A. Musgrave, The Theory of Public Finance (New York: McGraw Hill Co., 1959).

9. Discussion

W. D. Compton, Ford Motor Company: You made the comment, and I think it's quite correct, that there are examples where the private sector has not had proper market incentives to move as a total industry in a particular direction. I was curious, however, that you did not suggest how government controls might be used to develop market incentives. How, in fact, does the government develop the incentives that will allow the free market to operate?

H. E. Morgan: We are looking at that question with our Experimental Technology Incentives Program (ETIP) here at the National Bureau of Standards. Experiments are conducted with private firms to identify the barriers to technological innovation, some of which may relate to mechanical failures, and then suggest to the Congress changes in policies which will provide incentives to firms. Of course, many of us think about tax incentives or a reduction of anti-trust restrictions. These are the classic arguments often used by business for relaxation of controls so that they will have more incentive. Our program is looking particularly at procurement practices of government. What sort of innovations, for example, specifying performance in use and letting manufacturers come in with various designs to meet these performance requirements, are necessary to achieve objectives which are publically desirable. We're looking at regulations that inhibit technological innovation in order to suggest changes which will still protect the public and yet which will provide incentives to firms. Although the effort is underway, it is too early to say how it will pay off. Many other countries have gone way beyond this in that government acts as a partner with business in attempting to solve these problems.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Economic Impact of Tribology

H. Peter Jost*

Angel Lodge Laboratories
K. S. Paul Products, Ltd.
Nobel Road
London N18 3DB, England

The emergence of tribology as a subject in its own right arose out of industrial need. In the early 1960's there was a large increase in reported failures of plant and machinery due to wear and associated causes. At the same time, technology was increasing the capital intensity of plant and the use of more continuous processes; thus breakdowns of such plant and machinery were becoming more costly.

In 1966 a U.K. Government investigation came to the conclusion that by application of tribological principles, very large savings were obtainable, in most cases without appreciable capital investment.

Tribology was defined as "the science and technology of interacting surfaces in relative motion and of subjects and practices related thereto." In other words, it deals with all aspects of rubbing, sliding and rolling surfaces and included the subjects of wear, friction and lubrication. Application of tribological principles would lead not only to greater operational reliability, efficiency and productivity, but also to conservation of materials and energy.

To gain the advantages outlined in the above Report, estimated at the time to be in the region of £515 million per annum (p.a.) ($1285 million approximately), the British Government set up the Committee on Tribology, to implement the recommendations in the education, research and industrial fields.

Actual savings already resulting to British industry are conservatively estimated to be in excess of £100 million p.a. ($230 million), and potential savings (allowing for inflation since the 1966 estimate) to over £1,000 million p.a. ($2,300 million).

In less than seven years, the new concept of tribology has been accepted by the majority of the industrial countries as part of their scientific and technological background in the field of obtaining greater plant efficiency, better performance, fewer breakdowns and significant savings in other directions.

Key words: Economic savings of failure prevention; friction; lubrication; mechanical failure; tribology; wear.

1. Foreword—The Consequences of Preventing Mechanical Failure

Friction and wear-caused mechanical failures and maintenance normally have their root in phenomena based on tribology, the science and technology of interacting surfaces in relative motion.

Like other new industrial technologies, the problems of tribology are seldom solved by the application of the knowledge or techniques of a single subject area, such as physics or chemistry, but often require teams of people, bringing their expertise and techniques together for successful operation. New concepts and techniques develop from the linking of ideas from two or more different disciplines in new ways, but there are often difficulties in persuading specialists in established fields of knowledge to co-operate in such multi-discipline activities. Developments in tribology since the publication of the "Jost" Report have utilised existing and new knowledge from physics, chemistry, mathematics, statistics, engineering, etc. and advances in diverse fields from metalworking to medicine or space technology have been produced as a result.

The economic benefits that may accrue to industry from the application of such new technologies can be substantial; in the case of tribology in the United States, they could amount to as much as $16 billion per annum.

Some of the means which were utilized in the United Kingdom in order to reap the benefits attainable by application of the principles and practices of the multidisciplinary subject of tribology and some of the results obtained are described in the following paper.

2. Introduction

It is only eight years ago that the word "tribology" was used for the first time. It saw the light of day in a Government Report [1],[1] published by the United

*Chairman, Committee on Tribology, Department of Industry, U.K., President International Tribology Council. The views expressed by the Author in this paper are his own and not necessarily those of Her Britannic Majesty's Government nor those of H.M. Committee on Tribology or of the International Tribology Council, of which the author is Chairman and President respectively.

[1] Figures in brackets refer to reference and bibliography listings at the end of this presentation.

Kingdom Department of Education & Science, as a result of an investigation of a Working Group under my Chairmanship.

Whilst none of the constituent members of this new science and technology were new, some had indeed been known for many centuries [2], tribology brought together—for the first time—all the constituents under the overall description of "interacting surfaces in relative motion." This approach was new, inasmuch as nobody seemed to have regarded such subjects as "wear," "friction," and "lubrication" as wholly interdisciplinary—embracing Physics, Metallurgy, Chemistry, Mechanical Engineering, Mathematics, etc., as well as being parts of the same concept, namely that of forces being transferred from one surface onto another, at a time when the surfaces were in relative motion. Tribology is therefore a realignment and a polarisation of thoughts in many spheres, all basically dealing with "force transference between surfaces moving relative to each other," and as such, it was new.

Now it is not unknown that a new concept, particularly one that consists of a new orientation of some very old ones, may not always be willingly accepted and may indeed meet with fierce resistance, particularly from the experts in a particular constituent of the new concept.

This is exactly what happened in the case of tribology. Indeed it is true to state that some of the greatest reluctance to accept the new concept came from a few experts in lubrication—some of great eminence and distinction—and from those countries, including the United States of America, in which lubrication research was particularly advanced. Yet in spite of this, there are now tribology societies and tribology groups—as distinct from lubrication societies and groups—in the United Kingdom, Western Germany, Holland, France, Norway, Sweden, Poland, India, and in formation, in Italy. Tribology, or tribotechnology, has also become a recognized subject in the German Democratic Republic and others.

Why did this happen in such a short span of time and in spite of the lack of support from some academic quarters? The reason is very simple: the new approach which tribology provided was needed by industry.

The findings of the original Working Group expressed the view that with increasing mechanisation and automation, with greater use of capital intensive equipment and higher production rates, breakdowns and mechanical failures were becoming increasingly expensive, both directly and consequentially. Amongst the principal reasons for the normal and general occurrence of maintenance, breakdowns and mechanical failures in industry, breakage, corrosion and wear feature prominently, the latter being the predominant group of causes of maintenance and not insignificantly those of failures.

Efforts to prevent such wear-caused maintenance, breakdowns and mechanical failures have led to the fundamentals of tribology, i.e., the fundamentals of "interacting surfaces in relative motion," becoming recognised at a principal consideration in engineering design, and therefore of considerable importance to industry.

There is no doubt that, at least in the United Kingdom, the tribology movement was carried on the shoulders of industry. When, in five years' time, our present instructional campaign is completed, it is hoped that there will be at least 150 000 people in the United Kingdom, most of them at the technician level, but also quite a number of professional and qualified engineers and other scientists, capable of utilising knowledge of the practical facets of tribology for design purposes, in much the same way as they now use "strength of materials" and other "materials science" subjects.

A great deal of this training will have to be performed by industry and in industry; plans to achieve this are now well advanced. There is little doubt or argument that if industry is to profit from tribology, industry itself will have to learn how to apply tribology in all its aspects.

3. The Macro-Economics of Tribology

In 1965, the Department of Education and Science Working Group estimated that tribology could reasonably save British industry about £515 million per annum ($1236 million). Details of the investigation and the basis of the estimated savings were presented in 1966 in a U.K. Government Report which has since been referred to and may be known to you as the Jost Report [1].

Figure 1 gives a breakdown, taken from this report, which shows that the vast majority of the savings, namely those falling under headings d–g, have a direct relevance to the theme of this conference. One can rea-

	£ Million	$ Million
a). Reduction in Energy Consumption through lower friction	20	48
b). Reduction in Monpower	10	24
c). Sovings in lubricont costs	10	24
d). Savings in maintenace and replacement costs	230	552
e). Savings of losses consequentiol upon breokdowns	115	276
f). Sovings in investment duc to higher utilizotion ond greoter mechonical efficiency of machinery	22	52
g). Sovings in investment through increased life of mochinery	100	290
1965 Totals (as published in 1966) (1)	515	1,236

Conversion rate of $2.40 per £1.00 is used throughout this paper.

FIGURE 1. *Estimated figures of savings obtainable by British industry.*

sonably assume that after taking into account the inflation of the last 9 years, this figure is likely to be in the region of £850 million per annum ($2040 million). It should however be noted that these and the following calculations do not take account of the increased oil prices due to the recent Middle East crisis.

There is some evidence that, in the United Kingdom, through application of tribological principles and practices, industry is effecting considerable savings which may exceed £200 million per annum (480 million) [3]. The direct cost of government expenditure, including the salaries of the Civil Servants involved, to achieve this end was in the region of £1.25 million ($3 million). However, included in this figure are about £750 000 ($1.8 million) for the financing and setting up of three Centres of Tribology. These are principally consulting centres, all of which sell their services to industry, all of which are now self-supporting and which together have earned foreign currency to the value of nearly $1.5 million.

What would the equivalent figures of potential savings be for the United States?

Only very rough approximations can be made, based on limited comparative parameters, and even these may not be wholly accurate. Taking the employment in manufacturing in 1971 to be: U.K. 8.1 million, United States 20.9 million; crude steel production in U.K. 24 million metric tons, U.S.A. 109 million metric tons; gross domestic products (purchasers' values in U.S. million dollars); U.K. 134 838; U.S.A. 1 045 753, and estimated energy consumption in million metric tons coal equivalent: U.K. 307, U.S.A. 2327, [4, 5] the estimated range of potential savings obtainable at present in American industries, through tribology, could be between $12 and 16.3 billion (fig. 2).

Whilst these amounts are of course only a small portion of the gross domestic products of either country, they are both substantial and significant and represent sums which the industries of both our countries would be very pleased to save, particularly as such savings can be obtained without the deployment of large capital investment. For the macro-economies of tribology are of a nature which enable large returns to be obtained from minimum capital outlay.

It is not often that an intermediate technology can claim to procure a measurable economic effect within a period of a few years. However, at least as far as the United Kingdom is concerned, tribology—in the economic sense—has become one of the few applied sciences and technologies which is believed to have succeeded in making an impact, albeit a small one, on both national productivity and the economy.

4. Micro-Economics of Tribology

The spectrum of micro-economics of tribology extends from the summation of a multitude of small savings, obtained by the application of correct tribological principles, designs and practices on the one hand, to single or group savings of large individual amounts, the significance of which has been outlined in many papers in Europe and which is of particular interest to this conference. For the purpose of this paper, these two types of savings will be referred to as indirect and direct savings, respectively.

Direct savings are those which the direct application of tribological knowledge can procure, e.g., by direct expert advice leading to a design change.

Indirect savings are those savings which are applied only when the tribological knowledge, say that acquired at school or university, has penetrated to its ultimate points of application, e.g., design offices.

	1971 Employment in manufacturing Industries (A)	1971 Crude Steel Production (B)	1971 Gross Domestic Product (purchasers value) (C)	1971 Estimated Energy Consumption (D)	1974 Estimated Savings through Tribology (E)
	Millions	Millions Metric Tons	Million Dollars	Million Metric Tons Coal Equivalent	Million Dollars
UK	8.1	24	134,838	307	2040
USA	20.9	109	1,045,753	2,327	12,000 to 16,000

SOURCES:
(A) ILO Year book of Labour Statistics (1972) Table 2 Adjusted for U.K.
(B) U.N. Statistical Yearbook (1972) Table 124
(C) U.N. Statistical Yearbook (1972) Table 188
(D) U.N. Statistical Yearbook (1972) Table 140
(E) Author's calculated estimates, based on D.E.S. Report (1) and price Indices Numbers, U.N. Statistical Yearbook (1972), and above tables.

FIGURE 2. *Potential savings comparison, UK—USA.*

Although the figures for both U.S.A. and U.K. are based on the present systems of National Accounts of our countries, comparisons of this nature must of necessity be somewhat inaccurate. However, even if, for a number of reasons, such inaccuracies were ± 25 percent (or even ± 50%), the order of magnitude of savings is so great that they should not be lightly ignored.

The procurement of savings by the indirect route must be part of an integrated policy of developing and transferring tribological knowledge into industry. This includes knowledge transfer of the application of multitudes of minor elements of tribology into the everyday life of the industrial bloodstream. This normally necessitates a policy involving education, training, and appropriate promotion on a nationwide scale. If this is done, a permanent improvement of industrial processes, of significant magnitude, and of industrial designs of all machinery and equipment can be obtained. Indirect savings may, however, be difficult to quantify accurately; to do so requires the establishment of a basis connected to more definite and to more quantitatively measurable savings, after which by extrapolation and other means, the more difficult measurable indirect savings through tribology can be estimated.

In any case, as in any other subject, the industrial benefits of education and training are difficult to quantify with any degree of accuracy; however, there is evidence to show that—as far as tribology is concerned—in terms of productivity and profitability, their value is very considerable.

The other end of the spectrum of the micro-economics of tribology are the direct savings, i.e., those of individual or group savings of significance; these are more easily measurable. Within this group fall the large tribological savings obtainable by those industries which are based on capital intensive plants with costly production rates, particularly the process industries.

An interesting example of such industry, in which the application of tribological principles and practices has led to considerable and measurable savings in both the United Kingdom and in the United States, is the iron and steel industry [3].

When, in 1965, the original British Working Party estimated savings obtainable through better application of tribology and better education and research in tribology, out of an estimated saving of £515 million per annum ($1236 million), as being applicable to the whole British Industry [1], a saving of about £20 million per annum ($48 million) was calculated in respect of the Iron & Steel Industry. It was found that such saving could be obtained by the application of modern tribology to iron and steelmaking practices and machinery.

About 2½ years ago, the British Steel Corporation set up a Tribology Section. This was an entirely new venture, being distinct and different from the Lubrication Engineering activities which the British Steel Corporation engages in in its many steelworks.

To run this section, which is headed by Dr. Fred Westlake, a former student of Imperial College, London and of MIT, costs the British Steel Corporation approximately £50 000 ($120 000) per annum. (It would probably cost twice as much in the United States.) One of its main purposes is to look at tribological problems, which cause significant losses, and to solve them.

The criteria employed in the selection of its work were outlined by Dr. Westlake in May, 1973 [3]. They are as follows:

Firstly: The solution of a problem must be capable of leading to savings of at least £1 million ($2.4 million) per annum.

Secondly: A problem must be capable of being solved within a period of 12 months.

Thirdly: There must be a 90 percent plus chance that the problem can be successfully solved.

It was estimated that through tribology, by the end of 1973, the British Steel Corporation would have already saved £6 million per annum ($14.4 million), and a figure of savings of £20 million per annum ($48 million) was confidently expected by the end of this year.

Thus the original 1965 estimate of a total of £20 million per annum ($48 million) capable of being saved by the British Steel industry, through the application of tribological principles and practices, was an under-estimate; at least that is the view of the British Steel Corporation. It may therefore be equally true that this degree of under-estimation may also apply to the potential savings of £515 million per annum ($1236 million p.a.) in respect of British industry as a whole.

As a result of similar actions in the field of tribology, the equivalent savings that would accrue to the U.S. Steel industry, calculated on the basis of the difference in crude steel production between our two countries, could exceed the $200 million per annum mark. You may feel that even if through greater application of technology in U.S. steelworks this figure were only seventy-five percent achieveable, it is still one of considerable substance.

Savings at the above level are, however, not confined to the Iron and Steel industry. Evidence from another great process industry, viz: the chemical industry, showed that by the correct application of tribological principles, considerable savings were made:

(a) by the reduction of outage time of plant
(b) to the production efficiency of the new plant, largely through increased reliability, and
(c) by direct savings in purchasing costs.

On the latter point it is interesting to note that correct tribological design need not be more expensive, but can in fact save money.

Two outstanding examples relate firstly to the start of the low pressure ammonia plant, on which a rolling bearing experienced three catastrophic failures on the vetrocoke pumps. Because of the inherent hazards involved, the process operators refused to run the plant—not unreasonably—until reliable bearing arrangements could be installed. The Engineering Department's traditional proposals were to convert the pumps to plain bearings using an oil circulation system; this would have taken 3 months. Tribological investigation succeeded in identifying the cause of failure; and this was so well explained that the Trade Union representatives accepted the modified rolling bearing arrangement which reduced the outage time from 3 months to less than 3 weeks, representing an ammonia saving of 126 000 tons or (at 1966 prices) £1.26 million ($3 million).

Another example from the same company dealt with reciprocating compressors, which handled a variety of different gases. Before the tribological exercise was conducted, a conservative estimate of cost (both maintenance and production losses) over the 2-year period and arising mainly from tribological causes connected with piston rings, cylinder liners and glands, was about £500 000 ($1.5 million). Since the modifications resulting from the tribological exercise were carried out, the production losses in this area were negligible and the maintenance costs small.

Summing up, we have on one end the micro-economics of indirect application, largely introduced by a change of practices brought about by education, training and awareness promotion, and on the other end of the spectrum, significant direct savings arising out of individual applications or groups of applications.

To meet the requirements of indirect application, facilities for education, training, and appropriate promotion are required; to meet the requirements of direct application, expertise must be provided.

Since in industry the majority of tribological applications occur between these two extremes, both education, training and promotion on the one hand, and direct expertise on the other hand, are necessary in order to enable industry to gain the benefits of tribology.

4.1. Some Indirect Means of Tribological Savings

In the United Kingdom, in order to effect industrial savings through the application of tribological principles and practices, the Government Committee on Tribology set up a programme to encourage tribology to penetrate the mainstream of industrial life, in particular in education and training.

Universities, Technical Colleges, Industrial Training Boards and others were encouraged to cover the whole spectrum of tribology, and industry was encouraged to make use of the facilities thus made available. In addition, advice on research was offered and some research priorities established. Advisory services to government Departments and public bodies and others were provided by the Committee on Tribology.

The work of the Committee on Tribology was fully reported in "The Introduction of a New Technology." [6]

Let it suffice to report that there are now in the United Kingdom three Chairs in Tribology, over 70 Masters' Degrees in Tribology have been awarded, a basic tribology module—representing the minimum coverage of tribology which should be present in the education of all mechanical engineers—was established; and teaching material in tribology is now available from school to post-graduate levels.

4.2. The Tribology Handbook

The original Working Group found that there was a need for a reference book for use by engineering draughtsmen, designers and works engineers, its contents to be expressed in simple terms, primarily nonmathematical, fully illustrated, comprehensive and free from bias.

The work—called "The Tribology Handbook" [8]—was produced by an independent publisher, under the auspices of the Committee on Tribology, and is regarded by many as a breakthrough in the presentation of a new technology to the designer, draughtsman, technician and works engineer for the purpose of enabling him—without the use of much mathematics—to select tribologically correct designs, to use the tribologically suitable materials, and in general to ensure that tribological considerations are not omitted in design. In addition the effects of many tribological failures are shown, the causes explained and remedies listed.

This 500 page book—believed to be the first of its kind in any technology—is already helping U.K. industry to prevent numerous breakdowns, save costs and improve the quality of its products. A list of subject headings as well as examples of some of the information contained in it, are given in appendix 1.

It is intended to use the Tribology Handbook as a training means for technician and crafts courses both at technical colleges and within industry and to train at least 45,000 engineers and technicians in its use.

4.3. Industrial Centres of Tribology

The Tribology Handbook already veered in the direction of direct savings through tribology by providing the user with some "expertise" in condensed and readily useable form. In the main, however, to deal with the direct application of tribology, some of the largest companies in the United Kingdom now employ their own trained tribologists or have tribology sections.

To provide expert advice for the majority of companies, however, three Centres of Tribology were established with the help of governmental funding [6, 9]. They are the Industrial Centre at the University of Leeds, the National Centre of Tribology at Risley, and the Tribology Centre at the University College of Swansea. (In addition the National Engineering Laboratory at East Kilbride, Scotland, also provides tribological services of high quality.)

These Centres, although started off with government deficiency payments [6], are now viable concerns, selling their services to industry, for which they have procured savings of considerable magnitude, and—as will be seen—generally at comparatively very low cost.

It is important to point out that the work of these Centres does not conflict with the work of the many excellent universities offering their services in the field of tribology. The principal difference is due to the Centres having to earn enough money to keep going and growing without the need of further government aid, e.g., the amount of research work that they can afford to engage in on their own is very limited; indeed such work would normally have a direct selling relevance, as in the case of the National Centre of Tribology, which spent some money on developing a window hinge, which was then patented and is now licensed to commercial manufacturers.

Thus, in general, in the United Kingdom the universities on the one hand, and the Industrial Centres on the other, do different jobs and have different functions.

Since the Industrial Centres are obliged to be economically viable, companies seeking their advice, or considering seeking advice, may ask themselves "How much will tribology cost me?" (This is, of course, the wrong question. The right question should be "How much is bad tribology costing me at present?")

That a company need not spend much money to get profitable advice from an Industrial Centre is shown in figure 3. The cost of the contracts placed with the Leeds Industrial Centre, varied in value from £10 ($24) to over £20,000 ($48,000); the graph shows the distribution of contracts by value, and demonstrates that the majority of problems were solved for less than £500 ($1200), and a high percentage for less than £100 ($240).

The National Centre of Tribology at Risley, which is a larger unit than either the Leeds Industrial Unit of Tribology or the Swansea Tribology Centre, shows a distbution curve not dissimilar, but the average value of individual jobs is somewhat higher (fig. 4). This

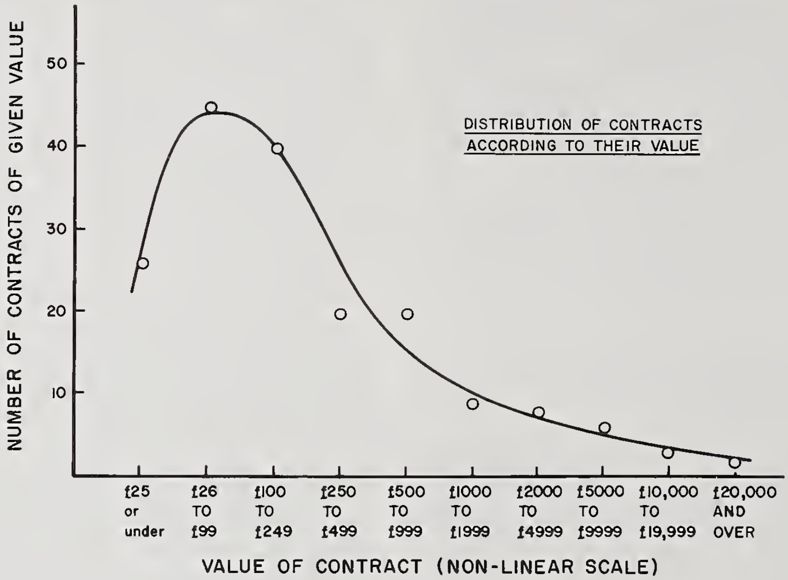


FIGURE 3. *Distribution of contracts according to value.*
Industrial Unit of Tribology—Leeds University.

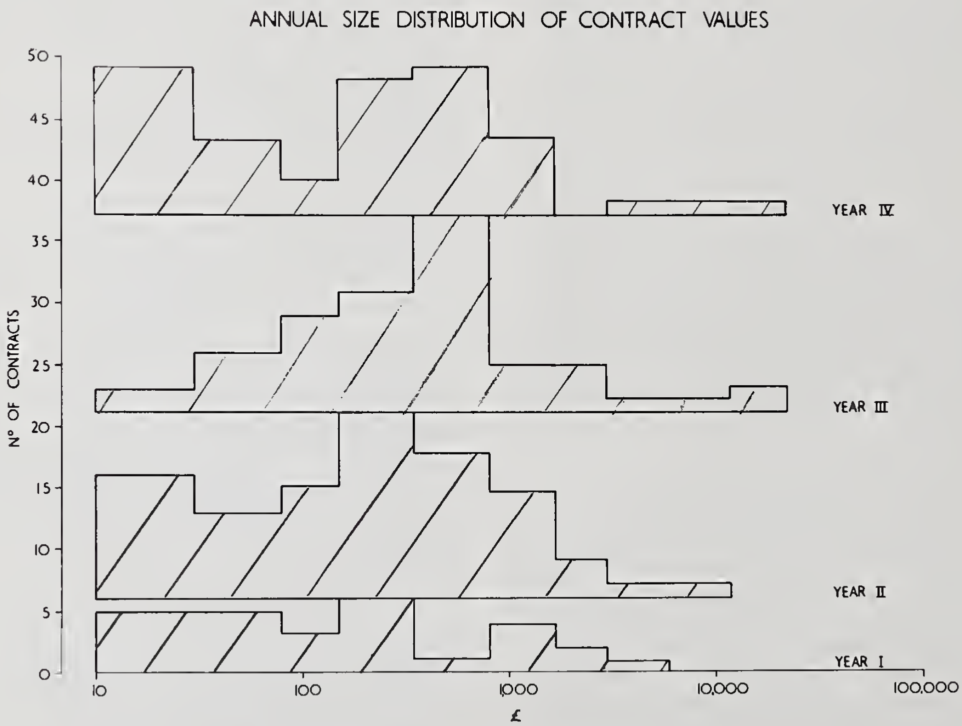


FIGURE 4. *Distribution of contracts according to value.*
National Centre of Tribology.

table excludes the $1 million contract which was awarded to the National Tribology Centre by E.S.R.O. for the setting up of European Space Tribology Laboratory.

A detailed government survey (in 1971) of the benefits of the Swansea Centre's work to its clients showed a net benefit considerably more than £125 000 ($300-000) to 50 clients on projects costing them only £16 000 ($38 000). This attractive financial picture is supported by a significant proportion of contracts being repeat orders, which would seem to show the satisfaction of clients with the work of the Centre.

Some work was carried out by the Swansea Tribology Centre, in conjunction with the Department of Economics of the College, to find potential savings through tribology in a very ordinary and not very modern engineering works. The survey identified 4 areas as the most obvious in need of improvement. The estimated annual savings were £8040 (nearly $20 000) for an investment of £5245 (nearly $13,000), a financial return of the order of 150 percent per annum. Even allowing for some inaccuracy in the calculations, this still leaves a very big margin over the usual financial return on industrial investment.

A cost benefit analysis conducted in respect of the University based Centres revealed the following interesting features [7]:

(1) More than 70 percent of the Centres' customers had approached them for assistance in solving particular problems. The remainder were about equally divided between those making a general enquiry and those who wanted organisational assistance.

(2) Seventy percent of the Centres' customers who co-operated stated that as a result of their use of the Centres' services, they had either saved money or expected to do so.

(3) Out of these, 30 percent had saved, or expected to save, a reduction in maintenance and down-time, 20 percent had saved the time of an information search, 16 percent had benefitted from the building up of general background knowledge, and the remaining 34 percent reported savings of a miscellaneous nature.

(4) Even the most conservative extrapolations of the assessed benefits suggested that the net savings, discounted to the net present values, were well in excess of the total grants given to the Centres by government.

(5) At one Centre, the cost/benefit ratio was as high as 1:6.

4.4. Direct Tribological Savings and Costs

The work of the three centres has been of a wide variety. A few examples to illustrate the range of subjects in the sphere of tribology with which the Industrial Centres of Tribology try to meet industry's needs of services in tribology are given in appendix 2. They generally show the savings obtained and the cost of the work by the Industrial Tribology Centres in the United Kingdom. They include some of the applications which the widening of the spectrum of the new concept of tribology entailed, and show how these Centres are slowly becoming an integral part of the industrial design process.

The seven cases referred to in appendix 2 cover: (1) industrial problems of reduction of friction in hostile environments (wire in acid bath); (2) the effect of difficult materials on seals (molasses); (3) co-operation in the development of probably the most important breakthrough in car cafety since the last war (safety tyre); (4) the more traditional redesign of pump bearings; (5) the detection of the tribological cause of computer amnesia; (6) the saving of energy (float glass process); and (7) medical applications.

It must be added at this juncture, that since complete commercial security is maintained in the work at these Centres, only those examples can be quoted which have specifically been cleared by the customers of the Centres, and only in such form as has been permitted. Confidentiality between thc industrial centre and the customer is of vital importance and the heart of the customer relationship of the centres. It is for these reasons that some of the interesting technical details of the cases quoted in appendix 2 are omitted.

The significance of these examples is therefore not their relative or actual values nor the tribological technicalities involved, but the exposition of a trend towards an integration process of tribological design services in industry which they reveal [9].

However, the Industrial Centres do not merely deal with tribological problems of a mechanical engineering nature. Since tribology is based on natural laws, and since the laws of nature apply equally in other spheres, the Tribology Centres helped in the interface between tribology and medical engineering, in which area there are many unsolved problems encompassing the concept of interacting surfaces in relative motion. Some examples are given in Case Group 7, Medical Applications.

It is a source of great satisfaction that, as a by-product of our work for industry, the application of tribological knowledge has in some areas contributed considerably to the alleviation of human suffering and to the enjoyment of life.

5. Conclusion

It has been shown that the acceptance and application of the concept of tribology can be of benefit to industry, in the spheres of macro-economics as well as micro-economics. It may well be that the savings, reasonably obtainable in the United States, could exceed a figure of $16 billion. Whether this figure is wholly accurate or whether it is 25 percent or even 50 percent inaccurate is immaterial. A very high value is involved.

The question now is whether such savings can be effected in the United States. There are many examples in the sphere of science and technology in which the United States has led the world. There are numerous areas in the application of modern technologies in which the United States may reasonably be regarded as being without equal. However—and this may come as a surprise to you—in the acceptance and application of the admittedly diffuse concept of tribology, and in the realization of the economic impact of this new technology, amongst all major industrial nations, the United

States could rightly be called—even by its best friends—perhaps not backward, but certainly underdeveloped.

It is therefore a question of some considerable interest whether, in spite of the large savings obtainable without correspondingly appreciable capital investment, the American system permits full advantage to be taken of the benefits of a diffuse and intermediate technology, such as tribology, and other similar intermediate technologies, particularly as such benefits can be gained only by an almost complete integration of the subject into the educational and industrial fabric of the nation.

To achieve such integration is not an easy task in a free and democratic society which particularly values, and rightly safeguards, its academic, educational and commercial freedom. However, in this matter, there is little difference between our two countries and, as has been stated, we have advanced some way towards succeeding in our task. In other words, it can be achieved by the free and voluntary cooperation and collaboration of all concerned.

Therefore, irrespective of whether in the case of the United States the estimate of potential savings of $12–16 billion per annum is accurate or not, I suggest to you that it is worth every effort that needs or should be taken to promote the science and technology of tribology.

In so doing, you will not only help your own industries and provide support for the research workers in the field, but you will also help to increase the wealth of the world without having to find corresponding investment capital. If only for the last reason and the need for the industrial nations to create the additional wealth necessary to assist in meeting the requirements of the rest of the world, you may feel as I do, that introduction of the concept of tribology is a task well worth undertaking.

The author wishes to thank the Directors of the University of Leeds Industrial Unit of Tribology and the Swansea Tribology Centre, and the Manager of the National Centre of Tribology, Risley, for their help in providing the Case Studies contained in this paper.

6. References and Bibliography

[1] Lubrication (Tribology)—A report on the present position and industry's needs, Department of Education and Science, (1966), H.M. Stationery Office, London.
[2] Dawson, D., Tribology, (Inaugural Lecture 1969), (Leeds University Press).
[3] Owen, Kenneth, Tribology speeds the wheels of British industry, The Times, London (1 June 1973).
Darnell, H., Applying tribology to British steel, iron and steel internation (1973).
Carvel, John, Cutting down on wear and tear, The Guardian (1 June 1973).
Tribology savings £200 million a year, The Birmingham Post (1 June, 1973).
[4] I.L.O. Year Book of Labour Statistics (1972).
[5] U.N. Statistical Year Book (1972).
[6] The Introduction of a New Technology. Department of Trade and Industry, 1973. H.M. Stationery Office, London.
[7] Tribology Notes on Science and Technology in Britain. Prepared for the British Information Services by the Central Office of Information in London. Ref. NST 56, Classification 6 (October, 1973).
[8] The Tribology Handbook, Edited by M.J. Neale, (Butterworths, 1973, London).
[9] Jost, H. Peter, Centres to help industry, The Financial Times (Nov. 25, 1969).
'Eight years after Jost,'
'Tribology in action'—Leeds industrial unit of tribology. (University of Leeds, London, England, 1974).
Jost, H. Peter, Tribology in engineering design, New Scientist (June 8, 1968).

6.1. Books

Summer-Smith, D. Introduction to Tribology in Industry (Machinery Publishing Company, Brighton, England. 1969).
Quinn, J. J. F., Application of Modern Physical Techniques to Tribology (Butterworth & Co. Publishers Limited, London, England, 1971).
Hondros, E. D., Tribology (Mills & Boon Limited, London, England. 1971).

6.2. Friction

Bowden, F. P. (deceased) and Tabor, D., An Introduction to Tribology, Anchor Books (Anchor Press, Doubleday Garden City, New York. 1973).
The Principles of Tribology, Edited by J. Halling. (MacMillan's London, to be published 1974).
War on Wear, Tribology in Iron and Steel Works. Proceedings of the Iron and Steel Institute, London, S.W.1. M69 Price £12.50.
Tribology Projects for Schools, Department of Trade & Industry, London (1972).

7. Appendix 1. Subject Matters of Tribology Handbook

A. Component Selection, Design and Performance

Bearings

A1 Selection of bearing type
A2 Selection of journal bearings
A3 Selection of thrust bearings
A4 Dry rubbing bearings
A5 Steady load, pressure-fed journal bearings
A6 Ring and disc-fed journal bearings
A7 Grease, wick and drip-fed journal bearings
A8 Pourous metal bearings
A9 Hydrostatic bearings
A10 Gas bearings
A11 Crankshaft bearings
A12 Oscillatory journal bearings
A13 Spherical bearings and universal couplings
A14 Instrument pivots
A15 Palin thrust bearings
A16 Profiled pad thrust bearings
A17 Tilting pads thrust bearings
A18 Selection of rolling bearings
A19 Rolling bearing installation
A20 Plain bearing form and installation
A21 Radial stiffness of bearings
A22 Bearing vibration

B. Lubrication

C. Properties of Materials for Tribological Components and Surfaces

D. Environmental Factors

E. Failures and Repairs

Failure of common components

E1 Failure patterns
E2 Plain bearing failures
E3 Rolling bearing failures
E4 Gear failures
E5 Failures of friction surfaces
E6 Seal failures
E7 Fretting problems

Operating and failure limits of components and machines

E8 Failure detection methods
E9 Failure limits of noise and vibration
E10 Failure limits of operating temperatures
E11 Allowable wear limits

Repair and maintenance methods

E12 Repair of plain bearings
E13 Repair of worn surfaces
E14 Repair of friction surfaces
E15 Lubrication maintenance planning

F. Basic Information

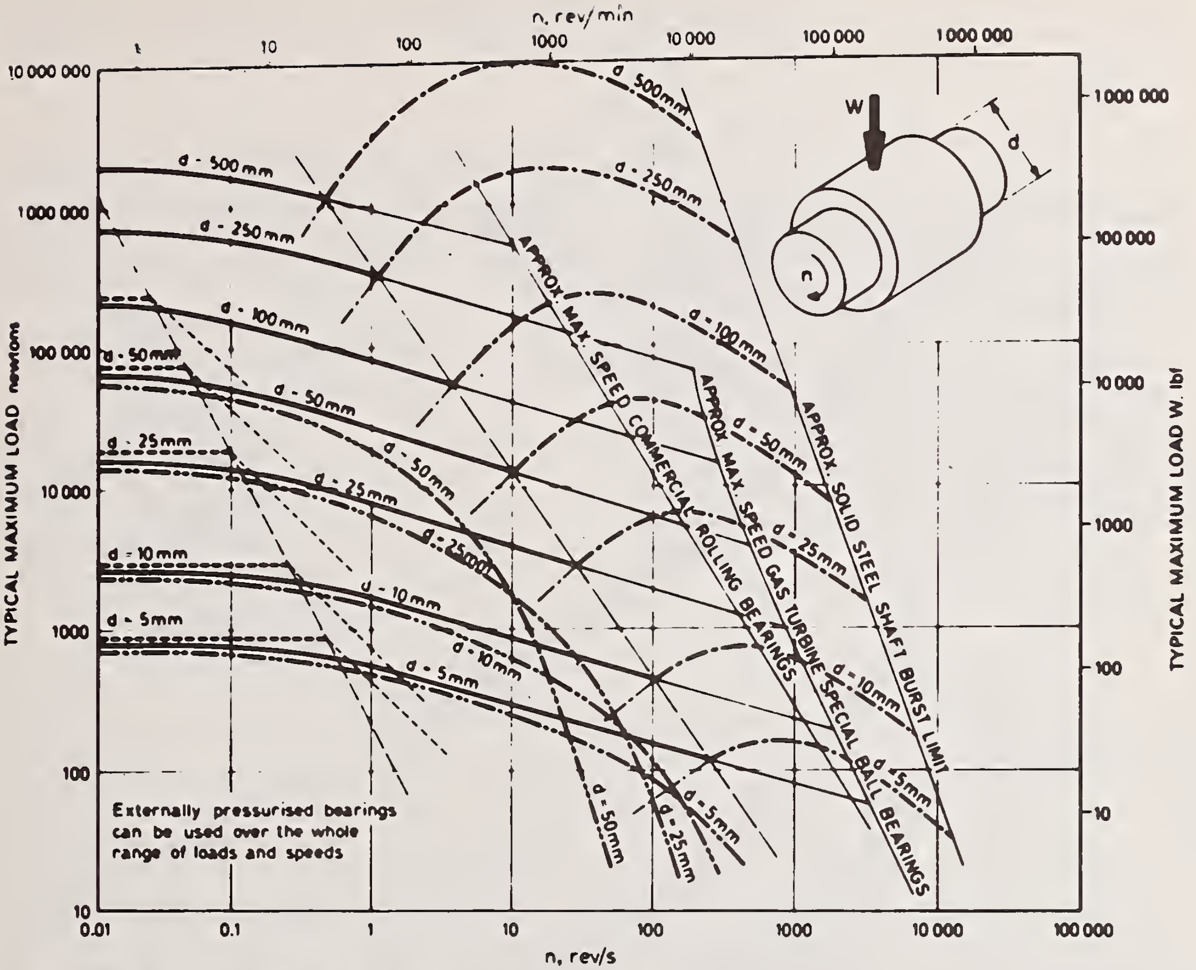
Basic tribology

F1 Nature of surfaces and contact
F2 Surface topography
F3 Friction mechanisms, effect lubricants
F4 Viscosity and rheology
F5 Methods of fluid film formation
F6 Mechanisms of wear

General design information

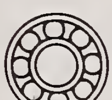
F7 Heat dissipation from bearing assemblies
F8 Shaft deflections and slopes
F9 Shape tolerances of typical components
F10 Relevant standards ISO and BSS
F11 SI units and conversion factors

A selection of examples are shown in figs 5–11.



Rubbing plain bearings in which the surfaces rub together. The bearing is usually non-metallic.

Plain bearings of porous metal impregnated with a lubricant.



Rolling bearings. The materials are hard, and rolling elements separate the two moving components.

Fluid film plain bearings. A hydrodynamic pressure is generated by the relative movement dragging a viscous fluid into a taper film.

FIGURE 5. *Example from Tribology Handbook.*
Selection by load capacity of bearings with continuous rotation.

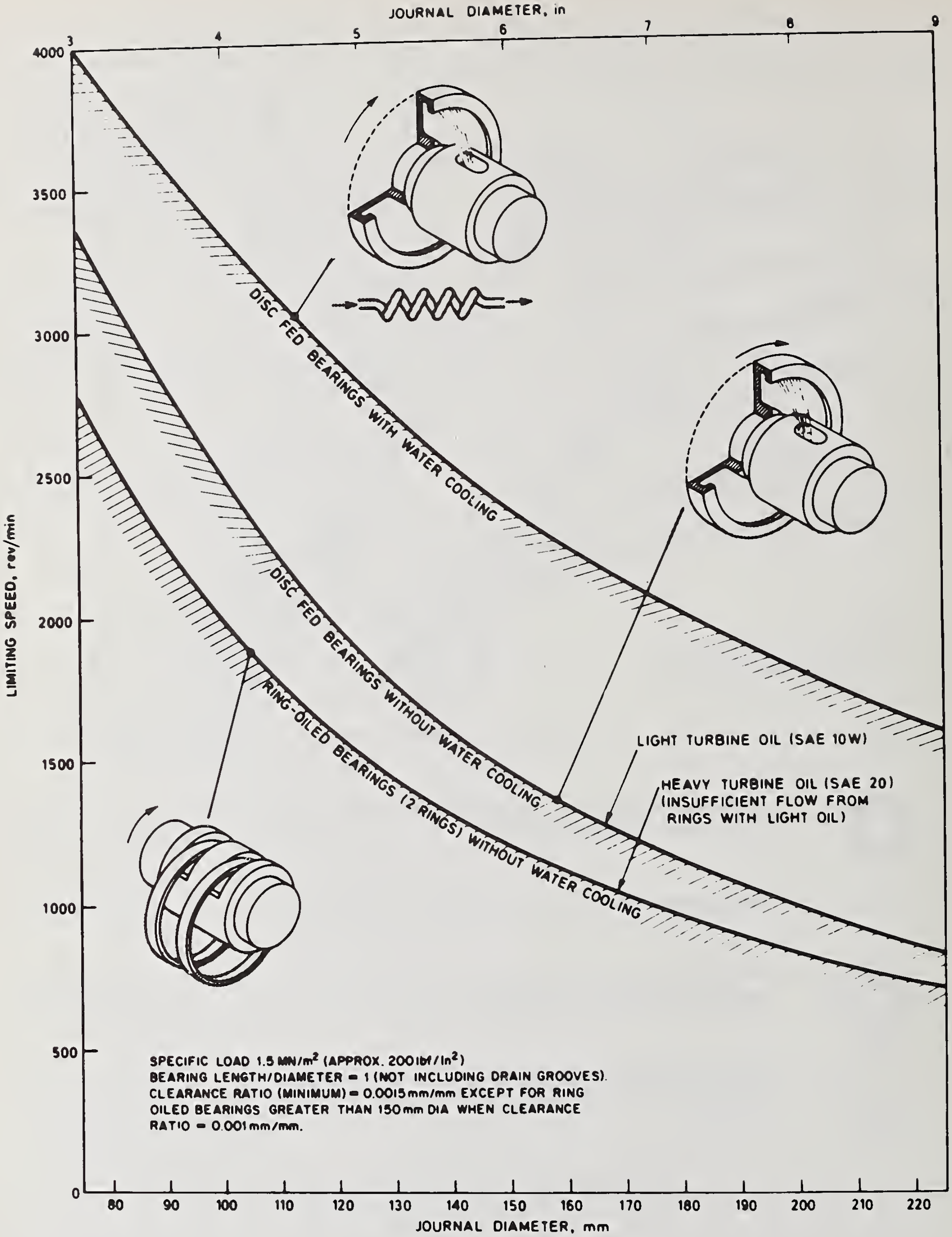


FIGURE 6. *Example from Tribology Handbook.*
General guide to limiting speed for ring and disk lubricated bearings.

CHECK-LIST FOR SEAL SELECTION

Temperature (see Fig. 12): seals containing rubber, natural fibres or plastic (which includes many face seals) may have severe temperature limitations, depending on the material; for example:

Natural rubber	−50 to + 80°C
Nitrile rubber	−40 to +130
Viton rubber	−40 to +200
PTFE plastic	−100 to +280

Speed (see Fig. 13).
Pressure (see Fig. 13).
Size (see Fig. 14).

Leakage (see Fig. 15).

Fluid compatibility: check all materials which may be exposed to the fluid, especially rubbers.

Abrasion resistance: harder sliding contact materials are usuallv better but it is preferable to keep abrasives away from the seal if at all possible, for example by flushing with a clean fluid.

Vibration: should be minimised, but rubber seals are likely to function better than hard seals.

Having narrowed down the choice of seals the reader should refer to the detailed treatment of the seals in question but is also recommended to discuss his requirements with a seal manufacturer or an independent information centre such as the British Hydromechanics Research Association.

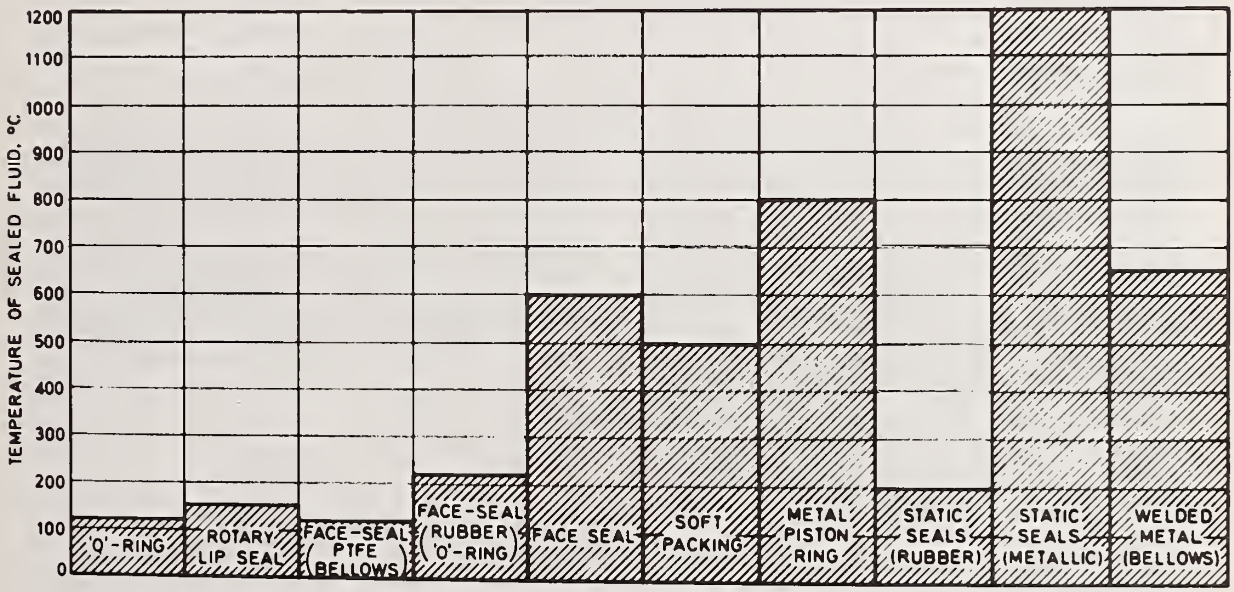


FIGURE 7. *Example from Tribology Handbook.*
Selection of seals.

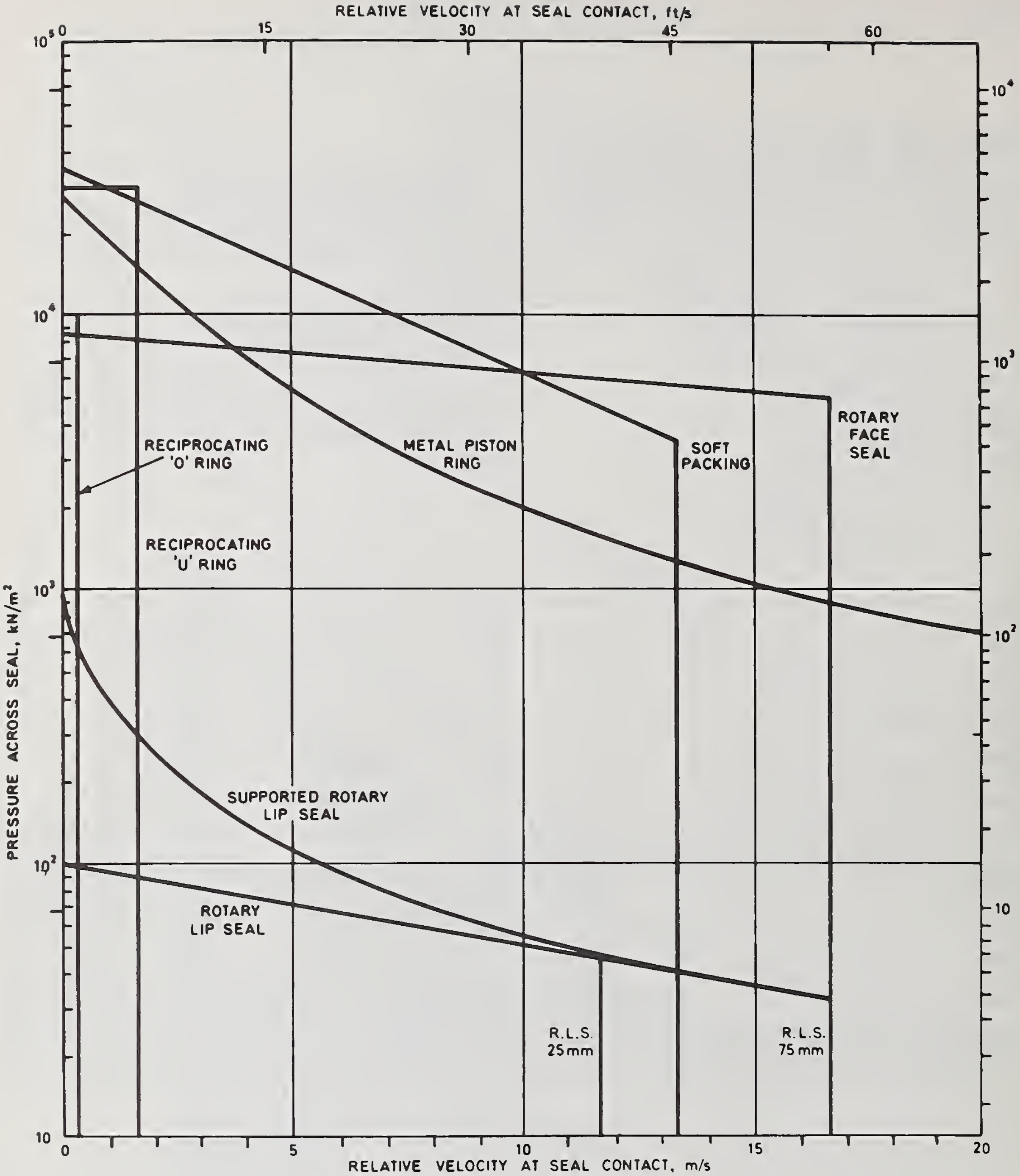
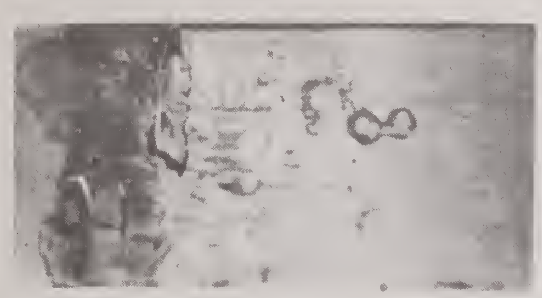


FIGURE 8. *Example from Tribology Handbook. Limits of pressure and rubbing speed for various types of seals*

Plain bearing failures E2



Foreign matter

Characteristics

[illegible]

Causes

[illegible]

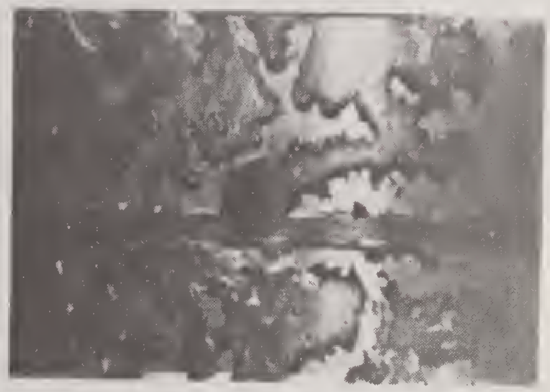
Wiping

Characteristics

[illegible]

Causes

[illegible]



Foreign matter

Characteristics

[illegible]

Causes

[illegible]

Fatigue

Characteristics

[illegible]

Causes

[illegible]

FIGURE 9. *Example from Tribology Handbook.*
Plain bearing failures.

Plain bearing failures E2

'Wire wool' damage

Characteristics

[illegible]

Causes

[illegible]

'Wire wool' damage

Characteristics

[illegible]

Causes

[illegible]

FIGURE 10. *Example from Tribology Handbook.*
Plain bearing failure—"wire wool" damage.

Fig. 1. Mechanical seal faces after use. *(a) normal appearance, some circumferential scoring (b) parallel radial cracks (c) radial cracks with blisters (d) surface crazing (b) – (d) are due to overheating particularly characteristic of ceramic seal faces*

Fig. 2 Tungsten carbide mechanical seal face showing symmetrical surface polishing characteristic of mild hydraulic or thermal distortion

Fig 3 Tungsten carbide mechanical seal face showing localised polishing due to lack of flatness, this seal leaked badly *The inset illustrates a typical non flat seal face viewed in sodium light using an optical flat to give contour lines at 11 micro inch increments of height*

FIGURE 11. *Example from Tribology Handbook.*
Seal failures.

8. Appendix 2. Case Studies from the Work of the Tribology Centres Including Savings and Costs

8.1. Case 1 (National Centre of Tribology) Tribology in Hostile Conditions (fig. 12)

Wire, in an acid bath, is processed round two sets of immersed pulleys in order to achieve the requisite residence time. The life of the set of 2 shafts was 13 weeks and that of the associate pulleys, only 3 weeks. The replacement costs for these components amounted to over £1,500 ($3,600) per annum. In addition, production losses and maintenance costs were extremely heavy and believed to be multiples of that figure. This wear problem was analysed by tribological consideration in metallurgy, in engineering design and in materials science. Thereafter the assembly was redesigned according to proper tribological principles.

FIGURE 12. *Aciding plating bath pulleys—National Centre of Tribology.*

The National Centre of Tribology manufactured a new assembly. This has now operated for more than 12 months without significant wear.

The charges which covered the redesign and manufacture of the first set of shafts and pulleys and subsequent supply of two further sets was under £1,000 ($2,400). Savings must have been extremely high because the company is now operating two machines on a 2-shift system with a much increased rate of production and virtually no down-time.

8.2. Case 2 (National Centre of Tribology) Molasses on Seals (fig. 13)

One of the leading British food cake manufacturers uses positive displacement pumps to meter quantities of molasses into animal feedstuffs. Leakage from the shaft seals had been continually causing trouble. Molasses is a difficult substance to handle, since besides being very viscous, it corrodes cast iron and mild steel. Moreover, it is abrasive, containing crystals of sugar at low temperatures, whilst at temperatures about 50 °C it turns irreversibly into a kind of caramel toffee. Existing seals needed replacing every 3 months with tightening at much more frequent intervals; wear of the shaft aggravated the situation. Ultimately the stainless steel shafts themselves required replacing.

The requirement from the National Centre of Tribology was to look into this matter tribologically and design some specific seals that (a) could be incorporated into existing pumps and, (b) would be leaktight for at least 6 months without requiring any maintenance.

After a study of the problem on the basis of tribological principles, a complete test rig was set up so that the typical pump conditions were simulated and an ingenious arrangement of seals packed with special lubricants was tested. The pump was operated under start/stop

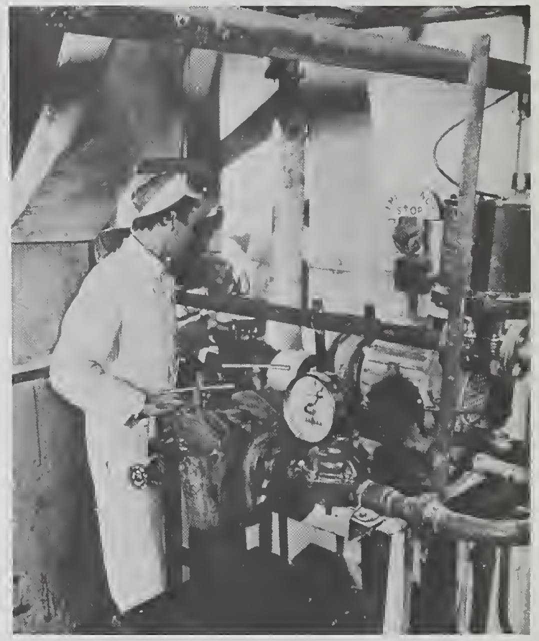


FIGURE 13. *Mollasses pump—National Centre of Tribology.*

and continuous running conditions and no significant leak had occurred during 600 h of running and 60,000 starts and stops.

The pumps were therefore modified and have now operated for periods well in excess of the target period without the slightest leakage. Similar pumps used in the works are in the process of modification. Several other factories of the same Group will be adopting the same design.

The cost to the customer, including laboratory proving tests, was less than £1000 ($2400). Total savings are now well over £5000 ($12 000) per annum.

8.3. Case 3 (Leeds Industrial Unit of Tribology) The Dunlop Safety Tyre (figs. 14–17)

Perhaps one of the most interesting and unusual involvements in an industrial centre with industry took place during the development of the Dunlop Safety Tyre. This is a tyre which, when punctured during use, is capable of running on at speeds of 50 m.p.h. for up to 100 miles, and with little or no effect on the handling of the car.

The special features of the wheel are that it contains no well. The rim is such that the tyre stays firmly against it even under conditions of tight cornering.

The special design of the side walls enables them to squash down onto the inside of the tread section and continue to support the weight of the car.

When the tyre deflates a special type of lubricant, designed by the Leeds Unit of Tribology, is spread between the side walls and the tread, performing as follows:

Firstly, it lubricates the surfaces between side wall and tread, thus preventing the degradation of the rubber by frictional heat. Full or elasto-hydrodynamic lubrication conditions apply in this condition which is referred to as "internal aquaplaning".

Secondly, it seals any small holes.

Thirdly, it partly vapourises and reinflates the tyre to a small pressure.

Fourthly, it has a high specific heat and acts as a transfer medium, removing heat from the rubbing surfaces and flexing parts.

Figure 14 shows the difference between a standard tyre and the new type of safety tyre, which contains pressurised lubricant containers. When the tyre is punctured, it is deflated, figure 15, until the canister releases the lubricant which coats the internal surfaces and seals them, figure 16. A partial reinflation takes place which allows the vehicle to continue running (figure 17).

It has been said that the development of this tyre is the most fundamental and most important breakthrough in road safety since the war. It is a source of great satisfaction that one of our Industrial Tribology Centres should have been involved in its development.

The actual cost/benefit of this particular exercise is impossible to establish; the benefits are so vast that they can be measured in £ millions; on the other hand, the work described in tribology was a team effort between the designers of the company and the Industrial Centre of Tribology at the University of Leeds.

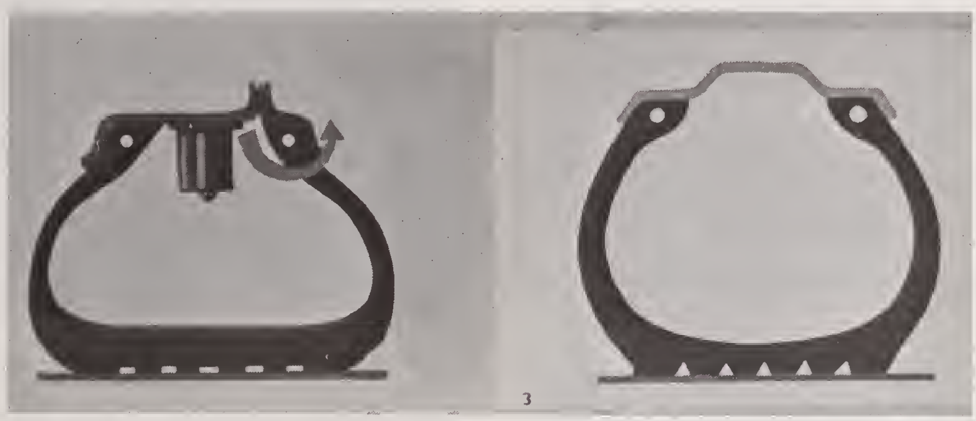


FIGURE 14. *Comparison between standard tyre and Dunlop Safety Tyre—Leeds Industrial Unit of Tribology.*

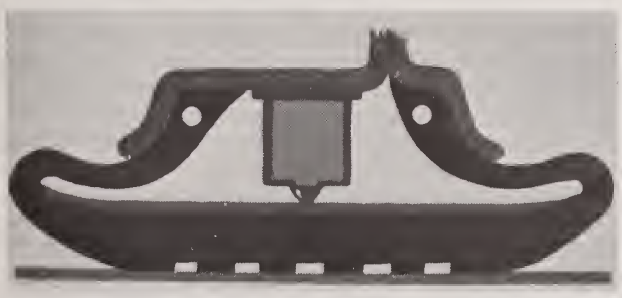


FIGURE 15. *Safety tyre deflating.*

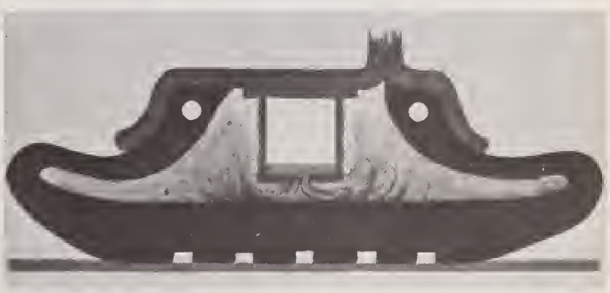
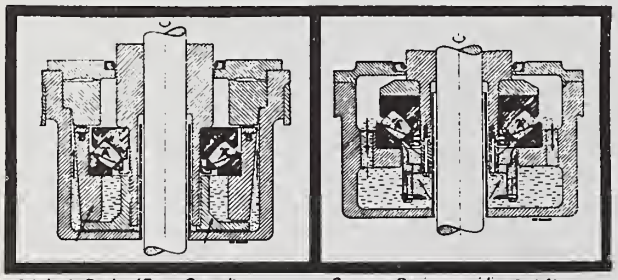


FIGURE 16. *Cannister releases fluid and tread and side wall are lubricated.*

FIGURE 17. *Partial reinflation.*

8.4. Case 4 (Swansea Tribology Centre) Redesign of Pump Bearings (fig. 18)

This involved the redesign of water cooling plant bearings for a large international chemical company where pump bearings required replacing every 750 h. Modification by the pump manufacturers produced very little improvement and the Swansea Tribology Centre was called in to solve the problems.



Original Design (Pump Capacity 10000 gallons water per minute at 150 ft head)

Swansea Design providing a vortex discharging oil into the bearing and carrying it through by the bearing's natural pumping action.

Since modification the pump has run over 2000 hours without bearing failure. This represents a saving to the operators of at least £2400 per annum for their once only investment of £340.

FIGURE 18. *Pump bearing redesigned by Swansea Tribology Centre.*

This was a classic example requiring both redesign and manufacture of the modified components by the Centre, and in respect of which the improvement in performance and the resultant savings could be estimated with some accuracy.

The key to the Swansea tribology analysis was the tribological principle that bearings of this type exhibit a natural pumping tendency in the reverse direction to the thrust, i.e., the direction (a) figure 18. The problem in the original design was a pumping ring which was attempting to force oil through narrow passageways path (b) in the opposite direction to the natural pumping action.

The Swansea design, including construction, installation, strain gauging, bearing load data and testing on the Swansea bearing test rig, was completed in a total "down-time" of the pump of 72 h and a total charge of £340 ($816). Since modification the pump has run for over 2000 h without bearing failure. This represents a saving to the operators of at least £2400 ($5760) per annum for their once only investment of £340 + £25 own costs.

8.5. Case 5 (National Centre of Tribology) Computer Amnesia (figure 19)

This case was known as "computer amnesia" or "loss of memory". It concerned a large memory store for digital computers capable of holding up to 1200 million 'bits' at one time (fig. 19). The store consists of a number of magnetic discs which rotate about a horizontal axis at a speed of 1200 rpm. Small magnets, imbedded in pads which float above the disc surface on a film of air, can record or read off information on the disc.

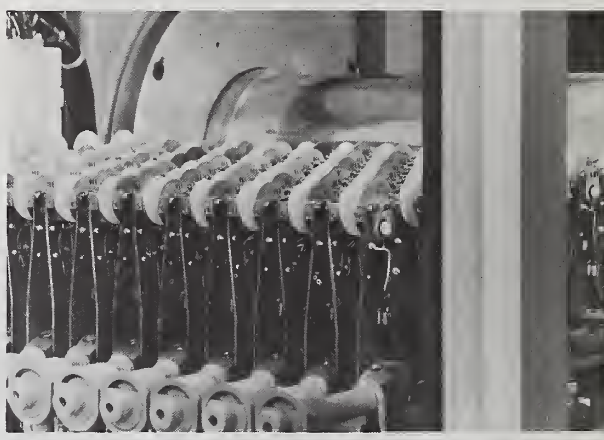


FIGURE 19. *Computer flying heads—National Centre of Tribology.*

The pads, or flying heads, have to swivel about a pivot in order to fly at the correct angle and to adjust the inevitable but small amount of wobble of the disc as it rotates (the one metre diameter discs have a permitted wobble of ± 0.07 mm at their outside edge). It was found that under certain conditions, a head would stiffen unduly on its pivot and result in "touchdown" onto the disc. Then, in an avalanche process caused by dust, many heads would "crash" and the whole disc, (or even several discs) would then be spoiled, with all the recorded data being lost.

Tribological examination found that the cause of this stiffening was due to a "fretting" action between the pivot spindle and the bore of the pad. A build-up of fretting debris was identified by the very sensitive technique of electron micro analysis and the Centre recommended a change in the manufacture and setting up procedures to reduce the risk of seizure, including a number of chemical treatments designed to eliminate fretting.

The diagnosis of the problem as being a tribological one of fretting damage has led to not only the savings of several thousand pounds per annum, but to the prevention of loss of possibly invaluable information. The charge made by the Centre was £500 ($1200).

8.6.Case 6 (National Centre of Tribology) Energy Preservation (fig. 20)

This was the first case in which the prime principal objective was in the sphere of energy saving. Since it is acknowledged that a third of all energy generated is used in friction, the relationship between the subjects of tribology and energy is not difficult to establish.

The case concerned dealt with the "float" process of glass making. Higher output at reduced power was the requirement (fig. 20).

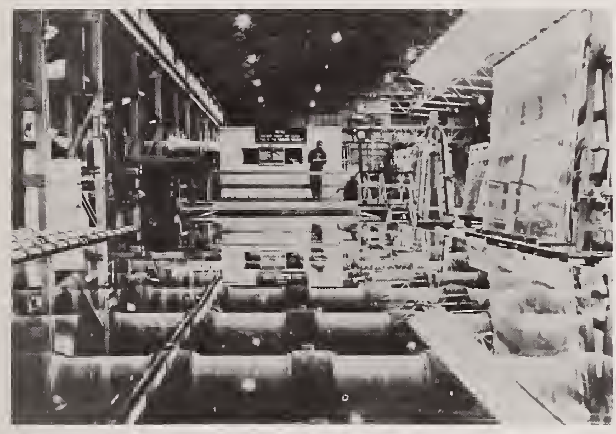


FIGURE 20. *Float glass process—National Centre of Tribology.*

In the manufacture of "float" glass, a ribbon of glass up to 11 feet wide leaves the float tank and enters the annealing lehr at temperatures in the region of 600 °C. In the design of a new lehr employing nearly 200 rollers (18 ft long and weighing 1300 lbs) it was necessary to cater for an increase of the speed of the glass by a factor of about 3. The plant is designed to run for two years without stopping.

By careful analysis of the whole design of all interacting surfaces in relative motion, including materials, lubrication, etc., it was possible to give the increased output without an increase in energy.

8.7. Case Group 7 Medical Applications (fig. 21)

Tribology has helped in the interface between engineering and medicine, in which area there are many unsolved problems encompassing the concept of interacting surfaces in relative motion.

The following are but a few examples of this work:

(a) The studies of synovial joints have led to a better understanding of the performance of these remarkable "bearings" and contributions have been made towards the alleviation of osteo-arthrosis. The introduction of synthetic lubricants into joints showing early signs of osteo-arthrosis and the development of total endoprostheses are instances.

Experiments have been undertaken to develop lubricating fluids which are compatible with the body and which possess rheological characteristics similar to those of synovial fluid. Clinical trials designed to assess the potential of this approach to some forms of failure of the human tribological system have begun.

(b) Design and development of total endo-prostheses replacement "bearings" are now available in a variety of forms for the hip, knee, shoulder, elbow and finger joints. Work on 'total replacement' artificial knee joints is progressing.

(c) But perhaps one of the most unusual applications of tribology was that in respect of the design of artificial heart valves (fig. 21). Heart valves are required to operate continuously for 40 million cycles per annum for many years. Plastic replacements for diseased heart valves, therefore, need to be extremely reliable and free from significant wear.

FIGURE 21. *Heart valve—National Centre of Tribology.*

At the request of the manufacturer of the newly designed heart valve, the National Centre of Tribology subjected it to tests for placidity, load and impact resistance, and sterility under various environmental conditions and temperature.

The substance pumped, which also forms the lubricant, is like salt water with very badly suspended semicolloidal particles in it. From a colloidal dispersion point of view it would not pass the tests of any reputable manufacturer. Wear took place on the hinges. In some cases this would become severe after 18 months to 3 years only. Tribological study, a knowledge of the operation of the valve, the effect of the environment on its materials, and the employment of tribological principles, led to the redesign of the valve. As a result the hinges will not now become troublesome for many years.

9. Appendix 3. Particulars of Industrial Centres of Tribology in the United Kingdom

National Centre for Tribology,

UK Atomic Energy Authority,
Risley, Lancashire.
Warrington 31244
Manager:
Dr. W. Roberts.

Industrial Unit of Tribology,

University of Leeds,
Leeds, Yorkshire, LS2 9JT
Leeds 31751.
Director:
Dr. R. Wakelin.

Swansea Tribology Centre,

University College of Swansea,
Singleton Park,
Swansea, Glamorgan, SA2 8PX
Swansea 25678.
Director:
Dr. A. R. Lansdowne.

National Engineering Laboratory,

East Kilbride, Glasgow.
East Kilbride 20222
Director:
Mr. R. Weir, CB.

10. Discussion

H. Corten, University of Illinois: It wasn't quite clear what you were suggesting to be introduced into the curriculum of a university mechanical engineering department.

H. P. Jost: I didn't suggest you should do anything. I just told you what we have done. You cannot be a mechanical engineer designer without a minimum of tribological knowledge. Therefore we have laid down, and when I say we, I mean not the Government but the Tribology Group of the Institute of Mechanical Engineers together with the universities and colleges, a module which is a minimum amount of tribology which every mechanical engineer ought to have in his curriculum, even if it means leaving something else out. Because without it, he cannot design properly and half of the failures which you are talking about will still be in, whereas by proper design, they could be left out. The polarization of the subject has been of great help to our universities, but basically the minimum knowledge in the curriculum of a mechanical engineer is the vital point.

A. J. Babecki, NASA Goddard Space Flight Center: We do, in the United States, teach our mechanical engineering students subjects which are all encompassing in the word tribology. We teach them metallurgy, chemistry, mathematics, etc., so the fact that we don't have a course named tribology does not mean that mechanical engineering students do not get the basic information. How did you arrive at the numbers that you were quoting as the estimated savings the United States could effect by using tribological applications? Did you assume that there's no technology being used currently in solving these problems? Certainly in the United States we have literally hundreds of organizations that are not necessarily called tribological centers that have research facilities that are in the lubrication area, the friction and wear area, and they are solving these problems. Are you discounting all of that work when you quote those numbers?

H. P. Jost: No sir. In the first place I didn't suggest any degree courses in tribology; in fact, I am very much against a first degree in tribology. Tribology is a part and parcel of mechanical engineering. I am very much for a post graduate course to become specialized in the subject. In the second place, my figures are based on a very thorough investigation of United Kingdom figures transposed using the United Nations statistics. Now, in the United Kingdom, we too, have all these research laboratories and schools. But we've found that the transfer of knowledge, in general, throughout industry to the designer is insufficient. Our savings in Great Britain are based taking into consideration all of the things you mentioned that can be obtained by a better transfer of that kind of intermediate type of technology in a way which is completely multidisciplinary.

A. J. Babecki: We have a lot of technology centers in this country to which industry can resort to solve its problems. I think it would be a duplication of effort to set up additional organizations in the United States as you have in England which would be called tribology centers, because we have those facilities already available.

H. P. Jost: I very much question that from personal knowledge. You have the same type, and perhaps better, organizations available as we have; you have excellent universities, you have excellent research institutes. But

I very much doubt whether you have any means of transferring this knowledge, whether from research or knowledge that exists, and most of this knowledge is in existence, into the everyday curriculum at work and the industrial bloodstream of your industry. Mind you, the argument you put forward is extremely good. It was put forward by our own people when they tried to oppose the original setting up of the Committee on Tribology.

J. E. Ryan, General Electric Company: I think philosophically this deserves comment because here we've had numerous component technologies including lubrication, the metallurgy of wear, etc., in existence often at the graduate level. It takes people at Battelle or in similar laboratories to "put it all together" themselves, but they do it only in limited areas. I'm not waving the flag for your tribology, but I do think that it gives us some pause for the value of putting things together in what we might call intermediate packages so we can see everything at once. I am a mechanical engineer by training and I've practiced mechanical engineering. I had to learn to "put it together" myself, and many a seal manufacturer I wrestled with over what you now call tribology. We didn't have the name but we had the problem. I think that there is some value in putting together for consideration new alignments of basic knowledge so that the practicing man who has to, I use the term again, "put it all together", may see not only the kinematics and the strength of materials or material science as such, but may see it integrated with lubrication and all the other relevant component sciences.

H. P. Jost: Thank you very much. This bears out the point which I mentioned: The whole tribology movement in England was borne on the shoulders of industry who felt they needed it, and against the reluctance of some of the finest and best experts in the field of research who just couldn't see that something wider was involved. Tribology in Britain serves industry, it's been borne on the shoulders of industry, and carried forward by industry.

G. M. Ugiansky, National Bureau of Standards: Since a lot of the failures in industry are caused by corrosion, a similar argument could be used for introducing studies of corrosion into the universities for mechanical engineers.

H. P. Jost: I quite agree. You're absolutely right. We have made a similar investigation into corrosion. Last year a committee on corrosion was set up by the Department of Trade and Industry with very similar terms of reference to a very similar work. There are one or two other intermediate technology subjects being considered for similar treatment. If you want to go to the moon, the government is prepared to appropriate millions of dollars so you can do it, but they are unwilling to fund areas of study in which basically all of the knowledge is already known, knowledge you need for everyday work. Corrosion and tribology are parallel sister subjects that are being dealt with in the same way. Talking as an industrial dealing with industrial problems, if there were no more fundamental research in these two subjects for the next 10 years I believe that all the answers are already there. It's just for us to use them.

J. R. Strang, Air Force Materials Laboratory: I don't think the study of tribology should be limited to mechanical engineers. There are tribo-chemical and tribophysical aspects to consider and I think tribological education should be provided for chemical engineers, surface chemists, and surface physicists also.

CONSEQUENCES OF MECHANICAL FAILURES

Session IV

Chairman: H. P. Jost
K. S. Paul Products, Ltd.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Nuclear Power Plant Safety

Peter A. Morris

Office of Operations Evaluation
U. S. Nuclear Regulatory Commission
Washington, D.C. 20545

Protection against the potential radiological consequences of malfunction or mechanical failure of reactor systems and components in nuclear power plants is achieved by a combination of procedural and technical constraints. For routine operation, plants are to be designed and operated so that radioactive effluents are maintained as low as practicable. To assure that the probability of serious accidents is acceptably low requires conservative design, construction and operation using redundant and diverse systems and components and a rigorous program for quality assurance at all times.

Key words: Failure prevention; mechanical failure; mechanical malfunction; quality assurance; reactor systems and components; safety standards.

1. Introduction

The potential consequences of mechanical failures in nuclear power plants conceivably could be very severe. The character of the consequences, where radiation and radioactivity are involved, are different from the consequences of mechanical failures in nonnuclear activities, and generally are not well understood by the layman. This paper discusses the general subjects of nuclear power plant design, nuclear safety philosophy and policy, Atomic Energy Commission nuclear plant regulatory procedures and nuclear power plant operating experience. Through this discussion, it is hoped that the importance of prevention of mechanical failures in nuclear power plants is put in proper perspective.

2. Discussion

2.1. Nuclear Power Plant Description

Nuclear steam supply systems for electric power generating plants are made up of extensive, complicated, advanced technology equipment and systems. Because of the nature of the nuclear chain reaction taking place in the reactor cores, there are very high radiation fields and very high levels of radioactivity involved in the operation of such plants. Because of the goal of achieving economic production of electricity, the steam-generating aspects of the plants require high power, high temperature and high pressure of water and steam, which leads to the storage of an immense quantity of energy in the reactor system. Further, because of the delayed release of heat from the decay of radioactive fission products in the reactor core, large amounts of heat are released even after the reactor itself has been shut down for a long time. The combination of a very large amount of stored energy in conjunction with a very large amount of radioactivity presents a significant potential hazard to the public and to the environment in the vicinity of a nuclear power plant.

Many different kinds of nuclear reactors for use in electric power plants have been designed and operated. All of them depend upon the fissioning of uranium, thorium or plutonium for operation. The major differences in kind are in the design of the fuel elements, the way in which heat is removed from the fuel, in the choice of materials used to cool the fuel elements and in the method used to control the average energy of the neutrons produced from fission before further interaction. Most fuel element designs have used natural uranium, enriched uranium dioxide or uranium-metal alloys. Moderators, the materials used to absorb the energy of fission neutrons, principally have been graphite, ordinary water (called light water) or heavy water (which has an extra neutron in the nucleus of the hydrogen atoms). The primary coolants have been light water, heavy water, liquid sodium, carbon dioxide, and helium gas. Within the United States, the great majority of the power reactors in operation today and to be built in the foreseeable future are of two basic designs and the rest of this discussion will be confined to those two designs. Both designs use uranium dioxide pellets, slightly enriched in the isotope uranium-235 for fuel, Zircaloy tubing for the fuel cladding and light water as both moderator and primary coolant. In the boiling water reactor (BWR) type, the pressure in the reactor primary coolant system is designed to be approximately 1000 pounds per square inch (psi). The primary coolant water is heated sufficiently while flowing through the reactor core to form steam in the reactor vessel itself and this steam is piped directly to the turbine. In the pressurized water reactor (PWR) type, the pressure in the reactor primary coolant system is designed to be approximately 2200 psi. In this design, the primary coolant water is piped to heat exchangers, i.e., steam generators, where the heat is transferred to a secondary coolant water system at lower pressure, in which steam is generated to drive the turbine. The BWR design is called a direct cycle system and the PWR design is called an indirect cycle system.

The light water reactor systems are made up basically of the fuel elements in the reactor core, a reactor pressure vessel and associated pipes, pumps and valves, control and auxiliary systems, engineered safety features and a containment building. Of course, both designs have a large turbine and generator. While this sounds relatively straightforward, consider for a moment or two what really is involved in the principal systems of a nuclear power plant. (The following listing is taken from a draft of a user's manual being developed by a nuclear subcommittee of the American National Standards Institute, N18–20, Nuclear Plant Reliability Data):

Standard System Coding of Nuclear Plants
System Description

Reactor

- Reactor vessel internals
- Reactivity control systems
- Reactor core

Reactor Coolant System and Connected Systems

- Reactor vessels and appurtenances
- Coolant recirculation system and controls
- Main steam systems and controls
- Main steam isolation systems and controls
- Reactor core isolation cooling systems and controls
- Residual heat removal systems and controls
- Reactor coolant removal systems and controls
- Feedwater systems and controls
- Reactor coolant pressure boundary leakage detection systems
- Other coolant subsystems and their controls

Engineered Safety Features

- Reactor containment systems
- Containment heat removal systems and controls
- Containment air purification and cleanup systems and controls
- Containment isolation systems and controls
- Containment combustible gas control systems and controls
- Emergency core cooling systems and controls
- Control room habitability systems and controls
- Other engineered safety feature systems and their controls

Instrumentation and Controls

- Reactor trip systems
- Engineered safety feature instrument systems
- Systems required for safe shutdown
- Safety related display instrumentation
- Other instrument systems required for safety
- Other instrument system not required for safety

Electric power systems

- Offsite Power Systems and Controls
- AC Onsite Power Systems and Controls
- DC Onsite Power Systems and Controls
- Onsite Power Systems and Controls (Composite AC and DC)
- Emergency Generator Systems and Controls
- Emergency Lighting Systems and Controls
- Other Electric Power Systems and Controls

Fuel Storage and Handling Systems

- New fuel storage facilities
- Spent fuel storage facilities
- Spent fuel pool cooling and cleanup systems and controls
- Fuel handling systems

Auxiliary Water Systems

- Station service water systems and controls
- Cooling systems for reactor auxiliaries and controls
- Demineralized water make-up systems and controls
- Portable and sanitary water systems and controls
- Ultimate heat sink facilities
- Condensate storage facilities
- Other auxiliary water systems and their controls

Auxiliary Process Systems

- Compressed air systems and controls
- Process sampling systems
- Chemical, volume control and liquid poison systems and controls
- Failed fuel detection systems
- Other auxiliary process systems and their controls

Other Auxiliary Systems

- Air conditioning, heating, cooling and ventiliation systems and controls
- Fire protection systems and controls
- Communication systems
- Other auxiliary systems and their controls

Steam and Power Conversion Systems

- Turbine-generators and controls
- Main steam supply system and controls (other than CCX)

Main condenser systems and controls
Turbine gland sealing systems and controls
Turbine bypass systems and controls
Circulating water systems and controls
Condensate clean-up systems and controls
Condensate and feedwater systems and controls (other than CHX)
Steam generator blowdown systems and controls
Other features of steam and power conversion systems (not included elsewhere)

Radioactive Waste Management Systems

Liquid radioactive waste management systems
Gaseous radioactive waste management systems
Process and effluent radiological monitoring systems
Solid radioactive waste management systems

Radiation Protection Systems

Area monitoring systems
Airborne radioactivity monitoring systems

Over 60 systems are identified in this table, almost every one of which has direct relevance to nuclear safety. Although the primary characteristics of these systems may be thought of as thermal-hydraulic, electrical, mechanical, or what-have-you, every one of these systems has important mechanical features whose failure would jeopardize the availability, reliability or safety of the plant. Needless to say, the possible failure modes cover just about the entire spectrum known to man.

Aside from the problems of high pressure and high temperature steam, which are common to any steam-turbine system whether fueled with uranium, gas, oil or coal, the different potential hazard from a nuclear plant arises from the possible release of radioactivity either as a routine effluent or as a result of an operational transient or accident. The fuel elements themselves, both fuel and cladding, become intensely radioactive. The uranium dioxide pellets retain some of the products of fission and transmutation, but during normal operation some of the products, particularly the gaseous isotopes, escape from the fuel material into the spaces between the fuel pellets and between the pellets and the tube-like containers in which they are located. Some of these radioactive materials diffuse through the fuel cladding, i.e., the walls of the tubes, and some leak out through pin-holes or other defects in the cladding, into the primary coolant water. In addition to this source of radioactivity during normal, routine operation, radioactive materials are also formed by the irradiation of the coolant water itself and any contained impurities, such as corrosion products and abrasion particles from pump seals, that might be in the water.

As just described, it is, therefore, expected and normal that the primary coolant water, after the reactor has been operated for any appreciable time, will contain a certain amount of radioactivity. It is also expected and normal that a certain amount of water, gas or steam containing radioactivity will escape from the reactor primary coolant system. Some radioactivity will escape during normal operation as a result of the irreducible low leakage of water and steam that occurs from high pressure systems with many valves, pumps, flanges and seals. Some radioactivity will escape when the reactor vessel is opened up—by removing the vessel head to permit replacement of fuel elements (about once a year) and some radioactivity will escape during the time that newly discharged fuel elements are stored underwater in a storage pool before shipment off-site to a fuel reprocessing plant. There are other minor sources of radioactivity release, such as from water chemistry sampling procedures, that are not of concern in this discussion.

Because radioactive fluids and gases are constantly being generated and escaping from the primary coolant system during operation of the reactor, there need to be extensive systems and equipment provided for processing these effluents. Such equipment needs to operate reliably for long periods of time, just as do the reactor itself and the electricity generating equipment.

2.2. Consequences of Reactor Accidents

As mentioned earlier, it is very easy to conceive of accidents in which large amounts of radioactivity might be released. The character of the consequences and the magnitude of the *conceivable* consequences of a very large release of radioactivity are clearly more severe than for the consequences of accidents in non-nuclear activities that normally are considered. Severe non-nuclear accidents that one thinks of are explosions, fires, collisions, ship sinkings, aircraft crashes and the like, as a result of which a hundred or a few hundred people might be killed and from which damage might cost in the tens or hundreds of millions of dollars to repair. From natural catastrophes, such as floods, tornadoes and earthquakes one might think of consequences resulting in thousands of people killed and property damage costs exceeding hundreds of millions of dollars. While such events would cause severe social consequences, they would not generally be expected to produce severe generic consequences or to render uninhabitable large areas of land, as might be the case following a nuclear accident. Nevertheless, they are serious accidents of a kind that have occurred and will occur well within the span of human history.

The consequences of several hypothetical large releases of radioactivity from a nuclear power plant were studied and reported by the Brookhaven National Laboratory in 1957 in a report known as WASH–740 [1].[1] The results of this study are frequently misrepresented to the public by those who are critical of the nuclear community. The largest release—for which no mechanism was postulated—assumed that most of the radioactivity contained in the fuel that could be released from the fuel if all of it were in a molten state was in fact not only released from the fuel, but also completely escaped the primary system and completely escaped from the containment building and was transported by unfavorable weather conditions to regions of high population. These conservative assumptions; assumed

[1] Figures in brackets indicate the literature references at the end of this presentation.

to apply simultaneously, constitute an almost impossible situation. No estimate of probability of such a release was made in the Brookhaven report, except to characterize it as very small. While this study did quantify the upper limit of consequences that could be conceived for such a hypothetical accidental release of radioactivity,[2] it did not and was not intended to, characterize the risk of nuclear power plant operation. The upper limit of theoretical consequences calculated in the Brookhaven report was used as a basis for selection of the maximum level of third party liability insurance to be provided by the government under the Price-Anderson Act, but was never intended to imply that there was any meaningful probability that such a release would in fact occur. Any discussion of risk needs to discuss not only conceivable consequences, but also probability or frequency of occurrence of such consequences.

A more recent AEC discussion of the probabilities and consequences of postulated accidents is provided in the report, WASH–1250 [2], "The Safety of Nuclear Power Reactors (Light Water-Cooled) and Related Facilities", that was made available in July 1973. Among others, two very important results were presented in this report. First, it was shown that when realistic models for postulated accidents are used, rather than extremely conservative models, the calculated off-site consequences for a given event are several orders of magnitude less. Second, quantitative estimates were reported of the probability of accidents leading to release of radioactivity. Assuming only independent mechanisms for failure, that is, not common mode mechanisms, it was estimated that the chance of an accident leading to the release of the equivalent of about five million curies of radioiodine is approximately one in 100 trillion each year (i.e., approximately 10^{-14} per reactor year). This, of course, would be a really very serious accident, and it is comforting that the estimate of the probability is as small as it is, even though the chance might be increased by a factor of 10 or so if common mode failure possibilities were taken into account. For a relatively innocuous release of from one to ten curies, the probability is still only approximately one in a thousand per year (i.e., approximately 10^{-3} per reactor year).

For about the last year and a half, a special study of reactor safety has been in progress for the AEC under the direction of Professor Norman C. Rasmussen of the Massachusetts Institute of Technology. This study was undertaken to provide a realistic assessment of the risk of nuclear power plant operation using the reliability analysis technology developed principally by the aerospace and missile industry for the Defense Department and NASA to calculate the probabilities of accidents where actual experience data were not available. Seven major tasks were defined for carrying out this study, as follows:

(1) Identification of Accident Sequences—A systematic study was made of possible accident initiators, such as equipment failures or natural catastrophes, that could lead to melting of fuel and subsequent release of radioactivity to the environment.
(2) Assignment of Probabilities—Existing experience data for components and systems similar to those in a nuclear plant were adapted to synthesize overall probabilities for the various accident sequences, using fault tree methodologies.
(3) Definition of Source Terms—For each of the accident sequences, estimates were made of the amounts of radioactive isotopes that would be released to the environment, taking into account filtering, plate-out or other mechanisms for retention within the containment structure.
(4) Dispersion of Radioactivity—Models were developed for calculating the dispersion of the various isotopes as a function of meteorology.
(5) Health Effects and Property Damage—Consequences of accidents were calculated in terms of fatalities, injuries, long term health effects such as latent cancer and genetic effects, and property damage.
(6) Other Risks—To provide a basis for placing nuclear risk in perspective, some risks in non-nuclear activities were investigated. These included high probability risks resulting from events that happen often enough so that their frequency can be measured and low probability risks resulting from events that have not happened, but whose frequency can be estimated.
(7) Overall Risk Assessment—This last task was an effort to convey an accurate impression of the risk of low probability events to a non-technical audience.

Although the report of the Rasmussen study is not yet available, some general conclusions have been released by Dr. Ray, Chairman of the AEC. The chance of a reactor core meltdown is caluclated to be in the range of one in a million to one in 10 million per reactor year (10^{-6}–10^{-7} per reactor year). The probability of such an event leading to severe consequences affecting a large number of people is calculated to be in the range of one in a billion to one in 10 billion (10^{-9}–10^{-10}) per reactor year. The consequences of this latter event are characterized as being comparable to the consequences of a large dam failure or to a large aircraft crash in a large, fully occupied stadium. In other words, even for the very serious nuclear incident, whose probability is extremely low, the consequences would be the order of several hundred people killed immediately. Depending on specific weather conditions, population distribution and emergency measures, several hundred additional persons might be killed or injured and land areas might be inaccessible for a number of years but not several tens of thousands of people killed as one hears from those who quote or extrapolate from the early Brookhaven report.

2.3. Prevention of Failures

The AEC regulatory program for assuring that operation of nuclear reactors does not present an undue risk

[2] A maximum of 43,000 deaths and the equivalent seven billion dollars damage.

to the health and safety of the public has been evolving for some 20 years, i.e., since the passage of the Atomic Energy Act of 1954. Originally, this program reflected to a large extent the safety concepts that had grown up since 1942 during the design, construction and operation of the Commission's production, test and research reactors, which generally were located at remote sites. Power reactors are preferably located near large population centers or industrial complexes, however, so that isolation alone cannot be relied upon for protection of the public from potential accidents. Current designs of nuclear power plants incorporate engineered safety features and containment structures in addition to the basic plant safety features to compensate for smaller isolation distances.

In addition to conservative siting practices, the three general concepts used in assuring that nuclear safety objectives are achieved are:

(1) Conservative design, construction and operation.
(2) A defense-in-depth philosophy.
(3) Quality assurance.

Clearly, the best way to avoid unforeseen and unacceptable consequences resulting from nuclear plant operation is to be sure that unforeseen performance and accidents don't happen in the first place. This truism leads to the requirement that designs be conservative, even though extra cost is incurred. Conservative designs are achieved by careful selection of materials and design principles, by use of redundant systems in parallel, and by use of systems incorporating diverse principles of operation and providing backup systems and physical separation of all systems.

Similarly, the use of conservative practices during construction helps to provide reliable systems and the use of conservative practices during operation reduces the chance for mistakes in operation, or deterioration of operating conditions.

Even though conservative design, construction and operation are mandatory and do produce a relatively small chance for unacceptable performance, the degree of assurance needed that public health and safety are protected has led to the requirement of a policy of defense-in-depth. What this means is that even though accidents should not happen, equipment and systems are required to be built into the plant first, to detect reliably any abnormal or impending abnormal condition and second, to provide reliable and positive control measures. The most important example of this policy is the requirement for first, an elaborate control system, second, an automatic, fast acting shutdown system and third, an independent and separate slow acting shutdown system to be provided in each nuclear plant to shut down the nuclear chain reaction, should this become desirable. In addition to control and safety systems provided to manage the nuclear chain reaction of the reactor core, in such a way as to prevent an accident, engineered safety features are nevertheless always provided to mitigate the consequences of an accident should it occur. Such features might include, for example, filters or sprays to remove radioactivity from the reactor building atmosphere.

Another way of thinking about defense-in-depth is to consider the barriers that impede release of radioactivity from the fuel to the environment. First, even when molten, not all of the radioactivity escapes from the fuel itself, and the fuel cladding, when intact, retains most of the activity. That activity that does escape to the primary coolant is retained in the primary system, unless there is a significant breach in that system. That activity which escapes the primary system will be retained, to a large extent, by the containment building (which is required in all cases). Finally, all plants are required to have an exclusion area, of some 600 yards radius or more, and are located in a low population zone of radius generally in the range of 2 to 5 miles, both of which serve to isolate the source of radioactivity from the public.

The third general concept, quality assurance, as used in the nuclear industry, is extremely important and is perhaps unique. It is defined to comprise "all those planned and systematic actions necessary to provide adequate confidence that a structure, system, or component will perform satisfactorily in service". As such, it is meant to require a rigorous discipline throughout design, construction, test, operation and maintenance of a nuclear plant on the part of all concerned to produce a very high degree of assurance that objectives are indeed met. Frequently, quality assurance programs provide a check or an audit by a separate organization of what is done by line organizations, to see that the job really is done and is done properly.

In addition to the three general concepts just discussed, the bases for reactor technical safety reviews include the following:

(1) General Design Criteria

There are 55 general design criteria listed in Appendix A to 10 CFR Part 50, which is the AEC's rule on facility licensing. These criteria are grouped in relation to overall requirements, protection by multiple fission product barriers, protection and reactivity control systems, fluid systems, reactor containment and fuel and radioactivity control.

(2) Radiation Protection Standards

These are provided in 10 CFR Parts 20 and 50 of the Commission's rules and are quite specific on allowable doses to people and releases of radioactivity to the environment.

(3) Site Criteria

Site criteria for reactors are provided in 10 CFR Part 100. No fixed requirements are listed, but rather procedures are described by which the suitability of a site is determined.

(4) Regulatory Guides

In many cases, requirements of the Commission's rules are general. To give additional guidance on implementation of the rules, regulatory guides have been developed (a) to describe and make available to the public descriptions of methods acceptable to the AEC regulatory staff for implementing specific parts of the Commission's regulations and in some cases to delineate techniques used by the staff in evaluating specific problems or postulated accidents; and (b) to provide guidance to applicants concerning certain of the information needed by the regulatory staff in its review of applications for permits and licenses. The guides are not intended as substitutes for regulations, and therefore, compliance is not mandatory. There are ten divisions of the guides dealing with different facets of the licensing process; Division I deals with power reactors.

(5) Industry Codes and Standards

The Commission strongly believes in the increased use of codes and standards to improve the quality of design and construction and to accelerate the safety review process. Where such codes and standards meet Commission requirements, this is explicitly stated in Commission rules.

(6) Letters of the Advisory Committee on Reactor Safeguards

These letters comment not only on the acceptability of specific designs, but also on generic matters, including such things as research and development.

(7) Decisions by the Atomic Safety and Licensing Boards, the Appeals Board, and the Commission

These decisions may very well deal with policy matters and therefore set precedents for future licensing actions.

Aside from the bases and requirements contained in the materials just described, licensing decisions are based on the technical analysis of the proposed plant design and proposed operation, to evaluate the ability of the plant to perform as designed for both routine and transient conditions. To evaluate the consequences of non-routine conditions, several "design-basis accidents" are postulated. The consequences of such accidents are calculated conservatively and the resulting calculated doses at the site boundary are compared with the guideline dose values that are provided in the Commission's site criteria and which are used in evaluating the suitability of a site. Thus, for a low probability, but a very severe postulated accident, such as the instantaneous break of the largest pipe attached to the reactor pressure vessel, the site and the reactor plant design must be such that an individual at the site boundary would receive a calculated whole body dose less than 25 rem and a calculated dose to the thyroid from iodine less than 300 rem, in the two hours following the accident. While these guideline dose values may appear large compared to background dose rates and to those acceptable normally from medical X-rays, for example, they are still in the range where biological effects can only begin to be observed.

How does the Commission assure itself that plant construction and operation really are as they are supposed to be?

First, the Commission, in an operating license includes license conditions, in the form of technical specifications, that are required to be met by the licensee. These specifications, for both nuclear and environmental control, include items in the following categories:

(1) Safety limits and limiting safety system settings.
(2) Limiting conditions for operation.
(3) Surveillance requirements.
(4) Design features.
(5) Administrative controls.

Second, the Commission has a program for periodic, and mostly unannounced, inspection of licensed facilities.

Throughout the construction period and throughout the lifetime of a nuclear plant the AEC makes inspections, on-site and at vendor and manufacturers' facilities, to ascertain compliance with Commission rules and regulations and to make on-the-spot assessments of the quality and safety of construction and operation. Such inspections include visits to the shops of the reactor pressure vessel manufacturers, observation of concrete and steel placement, review of welding procedures and inspections, witnessing of preoperational and startup tests and routine operation and maintenance, review of operating procedures and emergency plans and assessment of overall management organization and effectiveness.

2.4. Operating Experience

At the end of February of this year there were 42 licensed nuclear power plants in operation in the United States, representing more than 5 percent of the nation's electric generating capacity. The oldest plant started operation in 1959. The cumulative operating experience now totals approximately 175 reactor years. The safety record has been perfect, to the extent that no incident at a licensed nuclear power plant has caused any detectable exposure of persons outside the plant site. The operating record for nuclear plants has not been perfect, but nevertheless, has been quite comparable to that for other generating stations of the same size and age. For example, for units producing 600 megawatts electrical and up, the forced outage rates and plant availabilities have been nearly the same for fossil-fueled and nuclear-fueled power plants of the same age.

An indication of the kinds of equipment failures and malfunctions that have caused forced outages of nuclear power plants is provided in table 1.[3]

It should not be inferred that these forced outages were necessarily safety related. In fact, only approxi-

[3] Taken from AEC report 00E-ES-001, Evaluation of Nuclear Power Plant Availability, January 1974.

TABLE 1. *Equipment failures and malfunctions initiating forced outages.*

Portion of facility	Number of outages	Hours	Percent of forced outage time
A. *Primary coolant system*			
1. Valve & pump seats	20	1523	10.19
2. Steam generator tube leaks	4	1402	9.38
3. Jet pump failure	1	1322	8.84
4. Main steam flow restrictor	1	312	2.09
5. Main steam isolation valve	4	148	0.99
6. Recirculation pump flow control	4	68	.45
7. Installation of pipe restraints	1	36	.24
8. Recirculation system valve operation	1	35	.23
9. Miscellaneous	4	24	.16
Sub-total	40	4870	32.58
B. *Secondary system*			
1. Condenser tube leaks	8	2003	13.40
2. Turbine control & stop valves	19	737	4.93
3. Reheaters, deaerators & FW lines	11	233	1.56
4. Thrust bearing wear detector	3	201	1.34
5. Feedwater pumps	9	181	1.21
6. Feedwater control systems	13	163	1.09
7. Turbine-Generator oil system	6	162	1.09
8. Airlocks into condenser or turbine	4	107	0.72
9. Superheaters	2	29	.19
10. Steam jet air ejectors	1	28	.19
11. Turbine seal steam	3	7	.05
Sub-total	79	3852	25.77
C. *Control rod system*			
1. Control rod failed to insert	3	1113	7.45
2. Primary system leakage through seals	2	918	6.14
3. Control rod slipped into core	5	57	0.38
4. Control circuit	2	9	.06
Sub-total	12	2097	14.03
D. *Electrical generation system*			
1. Generator stator windings	1	1323	8.85
2. Turbine Generator control circuits	1	161	1.08
3. Generator bearing seal	2	106	0.71
4. Main exciter	4	96	0.64
5. Switch gear	3	24	.16
Sub-total	11	1710	11.44
E. *Containment system*			
1. Torus baffles	1	572	3.83
2. Inspect drywell	1	11	0.07
Sub-total	2	583	3.90
F. *Electrical distribution system*			
1. Rainwater in outside control box	1	259	1.73
2. Cable tray fire	1	182	1.22
3. Main transformer	3	59	0.39
4. Auxiliary transformers & control circuit	3	24	.16
5. Loss of Offsite power	1	18	.12
6. Substation Switchgear	1	2	.01
Sub-total	10	544	3.64
G. *Reactor protection instrumentation*			
1. Tubing in the steamline venturi	2	318	2.13
2. Miscellaneous	11	100	0.67
Sub-total	13	418	2.80
H. *Engineered safety system*			
1. High pressure coolant injection	3	68	0.47
2. Low pressure coolant injection	1	42	.28
3. Isolation condenser	1	14	.09
4. Breaker for ESS (DB–50)	2	7	.05
Sub-total	7	131	.89
I. *Process control instrumentation*			
1. Instrument transformer	1	85	0.57
2. Loss of power to instrument bus	1	14	.09
Sub-total	2	99	.66
J. *Auxiliary system*			
1. DC to motor generator set	1	79	0.52
Sub-total	1	79	.52
Grand Total	178	14,382	96.2

mately 45 percent had any potential safety significance and none of the forced outages in any way endangered the health and safety of the public.

The causes of the failures (not in any order of importance) included the following:

(1) Stress corrosion cracking of stainless steel pipes.
(2) Inadequate design for dynamic loading of pipes and fixtures.
(3) Fatigue failures from vibration.
(4) Short circuits from rainwater leakage into electrical systems.
(5) Packing and seal leaks in cooling water, steam and lubricating oil systems.
(6) Hydrogen explosions in waste gas treatment systems and fires in cable trays.
(7) Personnel errors, including failure to follow procedures, poor workmanship, inadequate design, installation errors and lack of adequate performance verification testing.

3. Conclusions

From a discussion of the design and operation of nuclear power plants it is clear that there are a great many systems and components whose reliable operation is essential to achieve a low level of risk to the health and safety of the public, either during routine operation or as a result of incident. From a discussion of the consequences of reactor accidents, it is seen that the character of a nuclear incident and the conceivable consequences are such that a very low probability of the severe incident is necessary to achieve an acceptable level of risk. Thus, very low mechanical failure rates of systems and components are required.

The previous discussion of prevention of failures described the philosophy and regulatory review procedures for assuring reliable performance of components and systems in nuclear power reactors. While this approach is thought to result in suitably safe reactor designs, it does not permit a quantitative estimate of the risk being assumed by the public as a result of the operation of nuclear power plants. Also, the present approach does not provide a systematic procedure for identifying the relative safety importance of systems and components. The current criteria for acceptance of mechanical designs of systems and components, the relative emphasis on certain systems and components and the incentives for considering alternatives have been the product of what I call "community judgments" of all those directly involved in the reactor safety assessment process. Primarily, those involved have been those qualified to understand the highly technical issues involved. These have been the designers, constructors, architect-engineers and regulators. More recently, as a result of the rather innovative procedures of the AEC, which provide for public hearings of an adjudicatory nature on each nuclear power plant, because of a positive policy of the AEC to make public its policies, procedures and decisions, because of the increased access by the public to government documents and proceedings under the Freedom of Information Act, and because of the National Environmental Policy Act which requires extensive consideration of costs and benefits of a particular licensing action in addition to assessment of environmental impact, many public interest groups and individuals have become a part of the overall decisionmaking process.

Speaking as an individual—and in no way reflecting the AEC position one way or another—I first conclude that reactors can be built and operated safely, and I further think that the risk to the public presented so far, by operation of a limited number of reactors for a relatively few years, has been minimal, and has been far offset by the advantages I believe have been or are certain to be derived. Two such advantages that come to mind immediately are the conservation of fossil fuels and the reduction of atmospheric pollutants that result from operation of nuclear rather than fossil-fired plants.

I also believe, however, that to an ever increasing extent, with hundreds of reactors to be built and each to be operated for twenty to thirty years or more, that it is important, if not essential, to know quantitatively (even if only approximately) what the actual risk is from operation of nuclear plants so that comparisons can be made with other risks, that in turn would permit either public (or "community judgement") decisions to be made on possible alternatives, or, overt acceptance of the nuclear risk, whatever it is. To make possible a more quantitative assessment, in my opinion, will require a coordinated effort to use more extensively in the nuclear power industry the reliability technology that has been developed recently in the aerospace and military arenas. In addition to using probabilistic and statistical methods for analysis, this will require greater effort to understand failure modes and mechanisms and to predict and measure the in-service performance of components and systems.

Finally, in the nuclear business, I believe there is a new dimension in which success is to be measured. With respect to the control of radioactive effluents, for example, the AEC regulatory policy for review of system designs for retention or processing of liquid and gaseous wastes has changed in the last few years from a policy of deciding whether systems met particular numerical criteria (i.e., "speed limits") to a policy of deciding whether systems were designed to do all that was reasonable and practicable (with the current state-of-the-art technology) to retain radioactivity within the plant. Analogously, it seems appropriate to me to argue that systems and components should be designed, constructed and operated to be as reliable, i.e., failure-free, as is reasonable and practicable with today's technology.

I recognize that there are arguments against this position. For example, on the surface such a policy seems to fly in the face of current emphasis on standardization. Also, such a policy, because of the imprecision of the definition of goals, presents uncertain and moving targets to the designer and owner-operator. Finally, there is the difficult problem of balancing the cost of achieving higher reliability or availability or safety against the benefits to be derived. Nevertheless, so long as any radiation exposure is considered potentially harmful (which is a basic, long-term assumption of all national and international radiation protection standards), it will be hard to defend not requiring the application of any available technology to achieve levels of risk to the public from operation of nuclear power plants that are as low as practical. It follows that reducing the frequency of, or preventing, mechanical failures will be an important part of this effort.

4. References

[1] WASH–740, Theoretical Possibilities and Consequences of Major Accidents in Large Nuclear Power Plants (March 1957).

[2] WASH–1250, Draft, The Safety of Nuclear Power Reactors (Light Water-cooled) and Related Facilities (July 1973).

5. Discussion

A. J. Babecki, NASA, Goddard Space Flight Center: Does the AEC have any research facility or laboratory that it is supporting which is looking at material problems or material failures in nuclear reactors, or is this work contracted out to private industry?

P. A. Morris: All of the AEC work is done by contract, usually in the national laboratories. The program that comes to mind first, is the one on heavy section steel plates at Oak Ridge. The most uncertain event in the safety of light water nuclear power plants is failure of the reactor pressure vessel. One worries about this because if you destroy the geometry of the vessel, you can no longer cool the fuel, and because of the decay heat, the chances of rupturing the containment vessel eventually are very large. So it is extremely important to understand the technology of the heavy section steel in reactor vessels. There are other programs, primarily carried out by manufacturers on Zircaloy because of the problems with Zircaloy cladding of the fuel.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Mechanical Failures and Public Expectations of Safe Transportation

Henry H. Wakeland[1]

National Transportation Safety Board
Department of Transportation
Washington, D.C. 20591

Although, in the interaction between society and technology, it is desired to minimize the incidence of mechanical failure under all conditions, the practical ability to circumscribe and control failures in transportation (for example) varies widely. Depending upon identifiable variables in circumstances, society may effectively require a range of performance between complete freedom from failure and general acceptance of repeated failures of a life-threatening nature. An attempt is made to classify the variables of circumstance which seem to govern public expectation and to create a structure in which the degree of freedom from mechanical failure which will be tolerated by the public can be reviewed. The possibility of using this structure for estimating public reaction to failures is discussed.

Key words: Effects of mechanical failure on environment; failure prevention; government action; mechanical failure; safety standards; transportation safety.

Most scientists, engineers, or managers in failure prevention work are sooner or later required to decide how much failure to prevent, who must bear the risks or losses of mechanical failures, and who is to pay for the prevention work and receive the benefits therefrom. Such decisions, once hardly recognized or thought to be invisible, are now more often made in the open. Everyone recognizes that these decisions govern the gross losses felt by society from accidents and costly failures. But that is not all. The correctness of a great many of these decisions also governs the economic success or failure of the systems which we are developing or operating. Where scientists or engineers are making the decisions, the rightness of our actions will be judged by results and will affect the way that society evaluates our professions. Need I say also that correct decisions may somewhat affect our individual progress?

The public need for correctness and a broad base in these decisions is intense, whether the decisions are made by technical experts or public bodies. We now see that mechanical failures can govern our nation's access to natural resources and useful technology. Public concern for the effects of mechanical failures on the natural environment has played a role in delay to the Alaska pipeline, delays to offshore drilling for oil, and delays in the licensing of badly needed nuclear power plants. No one has calculated the effects of these delays on our energy position, but the effect is certainly not negligible. In each of these fields, public bodies have insisted on employing their own decisionmaking approaches, reviewing the arguments and taking the decisions out of the hands of technical decisionmakers, who formerly were seldom questioned. In the case of nuclear power plants, the questions were raised without the example of a single accident having major consequences. The public concern was for what *might* happen in the future.

Major consequences may also flow from wrong judgments of public expectations or unanticipated adverse public reaction. Adverse economic estimates, even those developed by careful analysis, may reverse safety gains or deny technically feasible safety. Congress is now interested in the detailed arrangements for promoting or requiring the use of safety belts and automobile crash injury prevention equipment, for example, and may reverse various safety regulations already in effect. Before the interest by Congress, the courts had also been asked to reverse various safety regulations. Overall, there is an impression that Congress or the courts will take over wherever the values to be served are not agreeable or do not, for some reason, meet public expectations.

Congress itself, of course, serves a variety of values in loss prevention. One committee will advocate strict discipline of all program decisions according to benefits and costs, while another committee seems not to hear that argument and insists that innocent school-children must be protected. Still another committee says that not

[1] The views expressed are individual views of the author and are not those of the National Transportation Safety Board.

one single nuclear power accident involving the public is tolerable. If Congress disagrees, how can scientists and engineers be expected to make predictive decisions that will meet public expectations? Yet it appears that we must make such decisions.

The different public consequences that can flow from countless possible mechanical failures are simply too numerous to be sent to Congress or the courts for adjudication. The need for decisions may also be more immediate than can be handled by the courts; very seldom do the courts permit themselves to be asked in advance whether a certain engineering decision is acceptable. Such answers are often not discoverable through legal processes until a failure has occcurred. Most legal processes operate in judgment only *after* the loss has been sustained. The judgment may apply to all *later* design. Thus scientists and engineers must learn how to cope with public expectations on a predictive basis as a matter of protecting investment.

Scientists and engineers are in the best tactical position to make the publicly acceptable failure prevention decisions. Our skills are used first chronologically. We are building better technical understanding of failures and failure prevention. We know better how to organize a new system technically to halt the effects of both mechanical and human failures. Given enough time and money, we would be able to quantify many of today's qualitatively defined failures and assemble them into failure prediction and loss estimate for system design. By studying individual accidents and through the use of statistical accident data, we can find the loss-inducing failures in old or existing systems. We can determine what changes will have the greatest effect and what the costs of failure correction will be. We could even determine what groups in society are bearing the risk or the loss, and even whether the consequences of failure can be countered by those involved.

What we lack for predicting public expectations, however, is something that scientists and engineers consider a basic tool for prediction. In the problem of public expectations there is no logical pattern to be probed. Instead, there are conflicting aims. The fatal weakness for a predictive effort is the lack of general theory. From our scientific perspective, absence of accepted theory precludes teaching and understanding, and tends to require each new decisionmaker to develop his own theories and approaches. Yet, in each field of knowledge there was a time of conflict before theories appeared.

Perhaps it is incorrect to say that there is no generally accepted theory of what is valuable in safety or failure prevention. Cost/benefit analysis is rather widely accepted. A feeling that loss reduction efforts can be justified only by showing that total benefits of a change will exceed total costs is clearly visible. Sometimes this view is accompanied by a corollary; that any decision in which costs exceed benefits is irrational, or mere "emotionalism." From this economic point of view, non-cost-beneficial decisions are acceptable only as temporary waste, accepted because the less enlightened who hold the power happen to be making a compromise. Yet many decisions in safety continue to be made irrespective of cost/benefit urgings and without cost/effective criteria. As will be seen, these essentially economic reasons for loss reduction effort are only part of a fairly large family of values respected in American society. They have a useful, but circumscribed, role in safety decisionmaking, and may be opposed to other values of society having far better recognition.

The oral delivery of this paper consisted of an illustrated discussion of a number of accidents, all involving mechanical failures, with an explanation of the public values which the author diagnosed as having been present. These were concerns or values implied in the background, circumstances of action, or followup actions. The illustrative accidents were among those investigated by the National Transportation Safety Board since 1967; however, the social concerns noted were not findings of the reports, but observations. Other situations in transportation safety in which the public values of safety problems were analyzable were also presented.[2]

A "Schedule of Concerns" developed in an attempt to classify such observations, was then presented, and is a part of this paper. This schedule seeks to list major classes of socioeconomic concerns and public values which have been the apparent basis for actions (or inactions) in mechanical failure safety decisions made by technical personnel or others exerting controlling responsibility.

The sources for these concerns and values were developed from the field of transportation safety and, in candor, are not limited in derivation to safety problems involving only mechanical failure. The field of transportation safety involves human failure as well as physical failures. Nevertheless, it is considered that all the concerns in the schedule apply to some circumstances or situations that could be found in mechanical failure loss. The concerns are not yet a full list for transportation safety. Some concerns or values are listed in section IV which are specialized to the interested viewpoint of the mechanical system designer. This was done simply to provide specific coverage to a field of interest to those in the Mechanical Failures Prevention Group. Some of the concerns are of very broad application as to the organizational situation; others are narrowly stated for application to rather specific circumstances.

The purpose of the schedule is to provide a rudimentary organization to assist the decisionmaker or the analyst of decisions in reviewing a safety situation and evaluating a variety of concerns or values. Word descriptions have been written which seek to grade each of the concerns. The user looks at the situation and the possible approach to failure prevention from the standpoint of each concern individually, and he estimates its grade according to the word descriptions. To record his analysis or estimate, the user marks the circle in line with his selection. The circles are placed so that marks to the right indicate generally increasing need or desirability of a failure prevention action, as suggested by the author's observations. These are believed to be generally agreeable.

The schedule is designed to aid decisionmaking, but it is also useful for questioning decisions.

---

[2] Schedule of Concerns for Safety by Mechanical Failure Prevention or Control of Consequences, MFPG 1974, H. H. Wakeland.

The first gain for the decisionmaker is an assurance that he has at least considered the major values and concerns in which the decision may be evaluated later by others. This tends to handle one of the problems of poor prediction, which is simply that of overlooking some objections or value system pitfalls. Other ways to use the chart will be noted later.

Four groups of socioeconomic concerns and public values have been classified:

I. Economic Efficiency in Control of Loss to Society

This group of values reflects the desire to reduce failure associated losses in American society as a whole according to economic factors or economic criteria. The six concerns in this group come into play according to the circumstances and the amount of information available for the decision.

These concerns are relatively new in the historical sense, and with the exception of I.C and I.E, they rely on failure consequence data or cost data that were hardly visible a century ago. At the present period, most analysts try to select safety changes (as distinguished from safe design of new systems) on the basis of economic cost benefit status. This is concern I.D. Benefits in losses avoidable by failure prevention are compared with the cost. At a minimum, a certain change is cost/beneficial if the benefits in savings (reduced to money values) are greater than the estimated costs over a long time period. Where all benefits and costs can be determined, the benefit/cost ratio can serve as a basis for priority judgment among competing changes. To the degree that a population at risk is accurately reflected in the loss data gathered, it can be argued that discussions based first on benefit/cost status of a proposal come close to an elusive ideal called "The greatest good of the greatest number." Benefit/cost would produce the greatest benefit with limited funds when workable.

The weaknesses of an exclusive benefit/cost criterion, however, are many. Data are difficult to gather and because data change, decisions can seem more logical for incremental improvement than for planning. The necessary summation of benefits must total lives saved, injuries avoided, and property damage prevented. Inevitably, money values are placed on human life and suffering, with illogical results on philosophical grounds. Methods of evaluating life tend to emphasize the individual's value as contributor to an economic system, and absurdities may result as soon as the analysis begins to try to include system failures that affect non-wage earners. Yet, when a high benefit/cost ratio is determinable, there may be wide agreement and few countervailing concerns seem logical.

Priority in cost effectiveness, item I.B, seeks the same result, but avoids the embarrassing effect of totalizing non-comparable losses. Effectiveness is judged by fatalities prevented, *or* injuries prevented, *or* property damage prevented, stated separately. The result is somewhat more open and easier to deal with on a policy level. Decisionmakers can still decide to prevent large property losses or numbers of costly minor injuries instead of a few fatalities. The difficulties of data and of application to new systems or long range plans are still present. Cost effectiveness decisions can be more easily adjusted to "human" values.

When loss data and cost data cannot be logically combined for cost effectiveness judgment, the next lower step on the economic judgment scale is the size-of-the-target. This concern is item I.A, size of total losses in continuing accidents. Among the economic concerns, society at large easily grasps the idea that a field of large loss ought to be attacked. Since the loss is great, any effort has a better chance of being effective, it might be thought. Size of the target decisionmaking sometimes produces effective results, but heavy waste of resources can also eventuate. Where a field is completely new, however, size of the target is necessarily the first approximation to economic effectiveness. Size of the target is listed first because, in the public view, it is expected that something be done when failures are producing large losses. The demand to "do something" will be present, even when the absence of economic analysis suggests that results are uncertain. The need to take some action has traditionally been the first concern because the earliest knowledge of accident failure losses arises when it first becomes possible to measure the loss. This consideration does not, of course, say that the amount of effort must be proportional to the loss. That question has to be weighed on several bases.

The size of the total loss in continuing accidents is graded as a proportion of the losses which are within public recognition of the field of loss. The perception by the public or Congress of field of loss is quite important and has been closely related to the visibility of accidents, discussed in IIIA.

The absolute money cost of prevention measures has historically served to permit or limit safety efforts regardless of considerations of the efficient use of funds. Failure prevention methods can often produce an absolute cost-saving, especially where operating efficiency and safety can be improved by the same action. But in many cases, where cost must be incurred, funds are simply unavailable. Alternatively, costs incurred could even be financially disastrous to those involved. The grading of this item reflects apparent public expectations; namely, that there will be less public pressure of necessity where the cost is high. The losses may also be high; high enough to overcome the cost. Yet the cost must be considered. Note that this form of absolute cost assessment refers to every level of the economy. The absolute cost will be important to producers, middlemen, consumers, ticket purchasers, maintenance operations, public agencies.

The importance of an activity to the national economy is a de facto concern in loss or accident prevention work, but one which requires study more than specific argument. At the public level there are occasional tendencies to use safety or loss prevention arguments to bring about changes which are more economic than safety-oriented. There may be genuine situations in which the economic health of a portion of the economy is associated with failure prevention. For example, economic conditions in a major portion of the transport economy may cause deterioration of the equipment or a moratorium on safety improvements may be sought. Though there are few published examples of this concern as a real influence, known examples have had strong effect.

The role of importance-to-the-national-economy as a concern appears to change drastically after a system investment has been made. When there is a possibility of a major scope system change, the fact than an activity is important to the national economy can serve to magnify the scope of effort. Safety was introduced into the Urban Mass Transportation Administration under conditions before the major-investment. However, once an important investment has been made, costs of change for safety rise, and the costs may tend to be viewed as threatening the economic operation. When that circumstance is present, there may be less resistance to correcting failures if the system is not economically important. Item I.E applies to a local economy as well as to the national economy. Again, it may be wiser for the decisionmaker to be aware of the effect than to argue it.

In Item I.F the fact that a certain system has a low or high status of loss rates in relation to its competitors is noted. The fact that there are high loss rates may not generate a strong expectation of correction among the public unless that loss rate is known. The importance of loss rates is probably strongly influenced by concomitant concerns in equity and social justice and in public awareness concerns in Class III. That is, there may not be a strong public component of concern where those who are suffering the loss seem to be participating in economic benefits.

The Class I concerns for economic efficiency in loss control seek to treat society as a whole. The thrust of these concerns is the most efficient use of public funds to reduce the totality of loss. Where such judgments are not possible, the economic efficiency concerns still seek efficiency in the use of funds, but through other, less precise criteria. The economic efficiency concerns could

SCHEDULE OF CONCERNS

FOR SAFETY BY MECHANICAL FAILURE
PREVENTION OR CONTROL OF CONSEQUENCES

SOCIO-ECONOMIC CONCERNS AND PUBLIC VALUES WHICH MAY BE BASES FOR NECESSITY OF ACCIDENT LOSS PREVENTION ACTIONS BY DESIGNERS OR OTHERS EXERTING CONTROLLING RESPONSIBILITY.

INCREASING NECESSITY OF FAILURE PREVENTION OR CONTROL OF ACCIDENT CONSEQUENCES →

I. ECONOMIC EFFICIENCY IN CONTROL OF LOSS TO SOCIETY

I.A. Size of total losses in continuing accidents

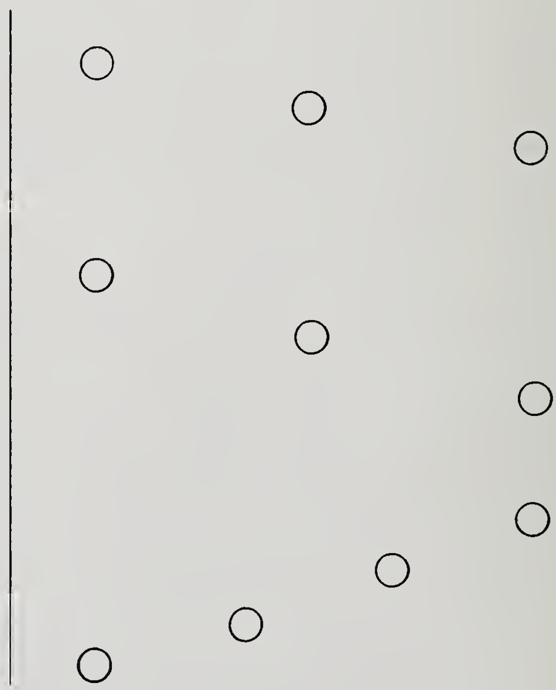
- o Small percentage of total losses in field
- o Significant proportion of losses in field
- o Losses large as proportion of field or in public recognition.

I.B. Priority of failure prevention in cost effectiveness

- o Preventive measure has low cost effectiveness in controlling risk relative to alternatives
- o Preventive measure has medium priority among alternatives on cost effectiveness basis
- o Preventive measure shows highest effectiveness for cost among alternatives

I.C. Absolute cost of preventive measure

- o Preventive measure has cost saving
- o Cost of prevention insignificant
- o Moderate cost
- o Cost of prevention disastrous to those who must bear cost.



also be credited with a strong thrust toward fairness. That is, it may be fairest to society as a whole to maximize the total benefits of failure prevention without respect to whether one element of society may be contributing more than another. Such an approach would seem logical for those social systems in which the good of the state is seen as paramount. All of the concerns for economic efficiency treat the losses as being undesirable because of their numbers. Numbers and costs confer priority. As noted earlier, the concern for benefits versus costs leads directly to evaluation of the human according to his contribution to the economy through earning power. This is most significant when seen beside the rule that only a benefit/cost ratio exceeding one can justify a safety change. These two features of judgment seem to say that the good of the economic system is the measure of safety effort, and that the individual is not to be valued beyond his economic contribution. Safety effort must show a profit for the economic system.

A good many safety proposals have actually been rejected in recent years without any stated reason other than a showing that their cost was greater than the present value of future earnings of those who might be saved. Such statements do not prove that our value system is narrow. The intent of such judgment is to save funds for some other safety technique that might save more lives. No analyst is known to have evaluated safety for retirees on the basis of present value of future earnings, nor has safety effort for newborns been limited to what might be justified by the investment made. Society is not really in danger of suspending safety effort should zero population growth become a goal. Our de facto social concerns extend beyond en masse evaluation. Let us try to formalize some of these values which arise from our status as individuals and groups.

SCHEDULE OF CONCERNS...

Increasing necessity →

I.D. Economic benefit/cost status of prevention measure

- Payoffs of prevention are other than economic
- Economic benefits and costs about equal
- Economic benefit somewhat greater than costs
- High payoff of expenditure in reduced loss or risk

I.E. Importance of activity to national or local economy

- Activity widely regarded as harmful to society or economy
- Activity supports economy, but could be abandoned without damaging effects
- Activity important to national economy
- Key economic or defense activity which must continue to operate

NOTE: The order of these grades is for design effort prior to system construction. Reverse the order for a system already installed in the national economy.

I.F. Status of activity's loss rates among competitors

- Activity has lowest loss rates of possible modes in activity
- Loss rate of activity near median of competitive activities
- Competing activities have much lower loss rates

II. Concerns for Equity and Social Justice

Mechanical failures involve risk and losses to people, and people are individuals. Safety regulations or other safety decisions based on economic efficiency often do not act equally on individuals and groups. The circumstances of failures place groups at risk in different degrees. Society has developed ways of avoiding inequities or unfairness, or of providing special protection in certain social problems. For example, laws provide that individuals can recover money for wrongful loss in an accident. Laws provide welfare payments for the disadvantaged, or special tax relief for the elderly or blind persons.

Men who make a career of flying military aircraft receive risk payments, effectively as compensation from the rest of society for presumed higher risk that benefits everyone. The cry of "Women and children first," effectively followed at the lifeboats of the *Titanic*, reflects one of the general principles of social priority decisions where risks are present. Ideas of equality, equity, and the avoidance of unfairness have been followed in Western society much longer than ideas of economic efficiency. In America it was declared to be self-evident almost 200 years ago that "all men are created equal." Above the entrance of the U.S. Supreme Court building are inscribed the words "Equal Justice under Law."

Yet there is still very little explicit recognition of these values when engineering designs are evaluated and decided. Standards and regulations do cover some hazards because they may be unfair or wrongful to some group, but little is said to make the reasons explicit. Many explanations can be given for this. There may be a lack of advocacy because groups are not organized or do not perceive the inequity in risk bearing. Historically, the raising of questions of group equity and fairness has temporarily threatened economic efficiency and produced controversy. (Consider, for example, child labor, mine safety, and the raising of crops with the aid of human slavery.) From a practical viewpoint, scientists and engineers are not trained to see instantly inequities caused by product design. Effort is required to discern these values in system design,

SCHEDULE OF CONCERNS...

II. CONCERNS FOR EQUITY AND SOCIAL JUSTICE

II.A. Share of persons at risk in general benefits of the Activity

- Persons at risk fully compensated
- Persons at risk receive benefit of employment
- Persons at risk gain benefits about same as general population
- Persons at risk bear risk almost entirely for benefit of others
- Persons at risk bear risk without compensation for benefit of same organization which controls their degree of risk

II.B. Voluntary-involuntary status of riskbearing by persons at risk

- Persons at risk bear risk voluntarily
- Risk could be avoided at cost or inconvenience
- Persons at risk cannot avoid risk-laden situation

II.C. Ability of persons at risk to know and avoid hazard

- Persons at risk fully appreciate hazard and can avoid
- Some appreciation of hazard, but degree of risk unknown
- Persons at risk cannot know the risk
- System tends to minimize awareness of risk or encourage unrealistic confidence

effort of a different kind than required to analyze effects in mass economics. When equity or equality are studied, however, it will be found that system design which efficiently reduces risk for the multitudes may concentrate the risk on the few. Scientists and engineers thus face a design decision not yet in the manuals. In actual experience, failures to recognize these values can result in reversals of technical decisions by policy-makers who are more directly trained in these matters. It is far preferable to detect these values at the design stage and perhaps make a different decision.

Let us review some examples in the six Equity and Social Justice categories of the Schedule of Concerns. We will draw upon problems treated in publications of the National Transportation Safety Board (NTSB).

Concern II.A provides a scale for varieties of situations in which groups of persons bear risk for themselves, for others, or for society in general. Unfairness is avoided by making failure prevention more necessary as the degree of riskbearing for others increases. Persons who live in the hazard zone along high pressure transmission pipelines bear the risk of a wide variety of mechanical failures. They bear this risk for the benefit of the remainder of society, and with no compensation. In a specific situation, it had been proposed that regulations for natural gas pipeline safety be justified by a showing that benefits in reduced losses would exceed the costs of preventing the failures. The NTSB pointed out that many of the fatalities and injuries occurred to persons living along the pipeline who were bearing the risks and losses, uncompensated, to provide the benefit of pipeline transportation to the rest of society. The Board said that benefit/cost analysis was not a basis for decision and that safety regulations should seek to equalize the risk between those who live near pipelines and the rest of the population. (The effect of balancing benefits and costs to adjudicate the risks of persons along the pipeline was to judge the safety of those persons as though they were within the pipeline economy.)

Category II.B describes the observation that society is willing to require greater protection against failures to persons who cannot avoid the risk. This observation

SCHEDULE OF CONCERNS...

Increasing necessity →

II.D. Protective attitudes of society toward deserving classes of persons

- Persons at risk regarded as taking risk recklessly or for profit
- Those at risk neither reckless nor especially deserving of protection
- Risk to persons in protection-deserving categories. Pregnant women, children, handicapped, elderly, otherwise disadvantaged
- Persons protected for symbolic or national interest reasons

II.E. Equal protection of groups by amount of effort

- Prevention efforts applied with approximately equal efforts to groups suffering loss, or at risk
- Prevention efforts not equal
- Prevention efforts appear very unequal

II.F. Apparent locus of operational responsibility

- Individuals considered primarily responsible
- Both individual and organizational responsibility
- Private organization regarded responsible
- Public organization regarded responsible

has been detailed for several classes of persons by Chauncey Starr.[3] The NTSB, in several documents relating to the safety of children riding schoolbuses, has cited their presence in the buses, not entirely of their own volition, as justifying higher priority efforts.

Category II.C covers the variety of situations as to knowledge of the hazard when there is ability to avoid it, if the hazard were known. Society's practice here is largely fixed by legal action after an accident has occurred, in which the failure to warn of hazards or possible failures can have a strong effect on liability. Recognition of this principle in the design of equipment not only protects against liability, but will also remove the unfair hazard. Examples at the high value end include booby-traps and practices which seek to fulfill liability criteria while submerging awareness of the hazard by vagueness or euphemisms.

Category II.D describes society's low estimate of the need for safety of motorcycle thrill riders compared to the actual heavy protection provided to persons in deserving categories. The social practice of providing this protection goes back at least to the prophet Jeremiah and is found to some degree in every civilized society. In a very recent example, the NTSB urged that the need to protect innocent children riding in schoolbuses outweighed the probable adverse benefit/cost ratio of a certain schoolbus safety improvement. After a period of time, the NTSB also recommended a way to implement a regulation without disturbing other priorities for use of government funds. Legislation proposed in Congress was opposed by the cognizant administrative agency on benefit/cost grounds, but Congress nevertheless enacted this form of special protection for children.

Category II.E parallels the idea of equal protection under law, translated to equal protection under regulation, and the similar idea of equal protection by technical effort. Notice that equal protection by amount of effort is an approach that will frequently require actions contrary to concerns for economic efficiency in loss reduction. This situation occurred in a study of Department of Transportation efforts directed to the protection of pedestrians, in which it was found that funded efforts for the direct benefit of pedestrians were of the order of $\frac{1}{10}$th the funded efforts which benefited motor vehicle occupants. After showing that the pedestrians were not the same group as the vehicle occupants, the NTSB recommended increased funding for pedestrian safety efforts, a recommendation which was carried out by the administrative agency.

Category II.F describes the variety of values based on private-public responsibility varying between low necessity of action where individuals are responsible and high necessity when the public is considered responsible. It is observable that Congress and the press respond much more strongly to failures involving public responsibility than private responsibility. An implicit example is seen in the difference in fatality rates between commercial and general aviation. Commercial aviation over a 5 year period has a low rate typical of all regulated carriers, in the region below 0.2 fatalities per hundred million passenger miles. General aviation (private flying) has a fatality rate between 16 and 20 fatalities per hundred million passenger miles, about 2 orders of magnitude greater.

III. Concerns in Estimating Public Acceptance of Change or Need for Change

Where failure prevention measures apply to the public or need the assistance of government, further concerns must be taken into account. Extremely strong demand for change or sympathetic attention to a subject grows and wanes under certain practical conditions. The concerns graded in Section III were arrived at by long term observation. They are practical public-opinion-generating forces, which affect the public necessity of change, and must therefore be concerns of the system designer or decisionmaker. These concerns tend to aid or to interfere with economically logical loss reduction decisions. They may be deplored by some scientists or engineers, and actions which take place are frequently called "emotional" or "irrational." There is, therefore, a strong tendency by scientists or engineers to reject these concerns rather than to make use of the motivations or changing tactics.

In Concern III.A it is observable that accidents of the type which attract heavy public notice, such as large explosions or commercial aircraft disasters, do attract public attention and interest which makes corrections more in the public eye and therefore more necessary. At the other end of the spectrum, public information must constantly be generated for highway problems which produce smaller losses visible only in statistics.

Concern III.B separates the amount of loss in a single accident from the public visibility as a differentiation to aid analysis. Some accidents will be highly visible, but produce little loss.

Concern III.C describes the apprehension of the public as a factor. This factor can be so strong as to completely determine actions, regardless of other factors. "Freedom from Fear" is of sufficient importance that it was named as one of the Four Freedoms in the Atlantic Charter. The scope of apprehension ranges from virtually zero, as in driving to work, to the extreme fear of exposure to radioactivity which delayed new nuclear power generation plants for a period of years during an energy crisis. The fact that highway accidents cause 150 fatalities per day, while there has never been a public fatality from a nuclear power plant demonstrates the role of public opinion as compared to statistical importance. The difference in effects is essentially one of differences in apprehension generated in the public by the activities.

Concern III.D, the impingement of preventive measures on personal freedom, reflects apparently deep-rooted but not necessarily universal American ideas of the proper role of government in protecting individuals. This is a cultural value and it must be considered. The full range of these concerns is not tested, but is classified here as a matter of structural order. Strong leadership can raise or lower the effects of these con-

[3] Starr, Chauncey, An Overview of the Problems of Public Safety, Paper presented to the Symposium on Public Safety—A Growing Factor in Modern Design, National Academy of Engineering, May 1, 1969.

III. CONCERNS IN ESTIMATING PUBLIC ACCEPTANCE OF CHANGE OR NEED FOR CHANGE

III.A. Public visibility of loss

- o Individual accidents attract short term attention
- o General awareness by public that losses are occurring
- o Heavy public notice of accident occurrences

III.B. Possible loss in single accident

- o Accidents never involve many fatalities
- o Typical single accident regarded as catastrophic
- o Large-scale loss to population in one accident

III.C. Public apprehension of activity to which failure is related

- o Activity generates little apprehension
- o Sufficient apprehenson to affect individuals' economic decisions
- o High degree of public apprehension

III.D. Impingement of preventive measure on personal freedom

- o Preventive measure violates supposed constitutional freedoms
- o Measure requires timely application and is restrictive
- o Preventive measure obtrusive, requires operation, but is nonrestrictive

cerns. At one end of the spectrum, motorcycle helmets were attacked as unconstitutional, but they are nevertheless a required form of protection in many States. In another case, the personal inconvenience of automobile safety belt interlocks is producing a rather strong effort against them in Congress.

A general observation of the concerns in Class III is that they are somewhat subject to leadership, but that leadership effects may be unpredictable. There is a tendency for the imperfect knowledge of hazards possessed by the public to produce concerns that run counter to logical loss reduction, but there are situations in which illogical concerns aid loss reduction.

IV. Concerns for Post-Accident Judgments of Design or Failure Control Decisions

This group of concerns is felt most strongly by persons who are responsible for design or regulation to control failures or failure consequences. The concerns are important because they do affect, or should affect, decisions as to whether mechanical failures must be prevented or guarded against. These concerns are understandable as written and do not require justification.

Decision Problems with a Schedule of Concerns

The purpose of a Schedule of Concerns is to aid decisions of failure prevention or control of consequences, not to make the decision. The task of the decisionmaker has always been to know the concerns present and to weigh them to meet a variety of public expectations. Thus the schedule adds a certain kind of order, but the task is still highly judgmental. Decisions still require balancing of values and concerns with less than a comfortable amount of knowledge. The schedule, for example, shows linear spacing of the circle marks, where nonlinearity is frequent.

The user first studies the situation in which the decision is to be made. He may begin by listing the various publics whose expectations may be touched, for that is important in estimating each concern. These may or may not include a consumer product public, a certain public which uses the transportation, a public of interacting administrative agencies, local or national public opinion, a professional peer group, legislative bodies, regulation enforcement agencies. The choice of the publics to be served can dramatically change the estimated concerns. Making specific choices of whose expectations are to be served is part of the value judgment.

The decisionmaker may then go through the schedule, studying each concern, perhaps initiating information gathering to clarify estimates. In so doing, he will almost be forced to consider many more factors than in an unstructured decision. He will be required to explicitly determine how various groups will be affected. The decisionmaker may very well find ways to decrease the necessity of controlling failure, as by opening a hazard to full knowledge rather than hiding it. He may find that responsibility should be in a different place. The decisionmaker will see directly some differences between decisions motivated by broad public concerns (Groups I, II, and III) and those which are needed to forestall his own possible embarrassment (Group IV). These concerns tend to be additive; it is not a loss to anyone if more safety is provided while protecting the designer's reputation.

When the schedule is marked with estimates of concerns, the conflicts between necessities inevitably stand out. The numbers of high and low estimates can be totaled, but that is not sufficient. The high-valued concerns can be assembled in a list. Take special note of the lowest valued concerns and list them. (Note, however, that some low-valued concerns are the result of ignorance while others are genuine contraindications of action.) When the lists are compared, the concerns in the broad judgment will be capsulized.

It will also be useful to write a separate description of the conflicts that appear. For example, Winston Churchill, in his decision at Dunkirk, knew explicitly that he was promoting overall economic efficiency by rescue at Dunkirk and permitting unfairness and sacrifice as smaller garrisons not to be rescued. It is important to know what concerns are served and what concerns must be sacrificed.

Thus are reached those ultimate conflicts, often suppressed, which require the exercise of knowledgeable judgment. There will also be happy discoveries where it

SCHEDULE OF CONCERNS...

Increasing necessity →

- o Preventive measure unobtrusive, requires no attention or operation

IV. CONCERNS FOR POST-ACCIDENT JUDGMENTS OF DESIGN OR FAILURE CONTROL DECISIONS
(Basis is not legal responsibility, but factors which influence society's judgments.)

IV.A. Predictability of loss by those responsible

- o Possibility of failure leading to accident is obscure or unpublished
- o Failure involvement in accident analyzable by known techniques
- o Failure involvement in accidents documented
- o Role of failure in accidents shown by statistical data

IV.B. Time since accident consequences of failure identified

- o Possible failure consequences pointed out recently
- o Possible failure consequences known for some time
- o Possible failure consequences pointed out repeatedly over period of time

IV.C. Involvement of deciding organization in consequences of accident failure

- o Deciding organization will not suffer money loss from accident
- o Deciding organization status re loss uncertain
- o Deciding organization must itself absorb money or personnel loss

is found that economic efficiency, equity, and public acceptance aid each other.

A schedule can be a wooden thing. The word implies, sometimes, a mindless or routine repetition. But this schedule is an early approach, the writer believes. Some additional steps are already visible. They include classification of recurrent application situations, detailed case studies of concerns as they apply to situations, search for patterns and combinations of concerns which have defined major failure prevention decisions of the past. Are there other concerns not generally identified? In whatever way they study the problem, scientists and engineers need to devote more attention to finding social concerns and values in their work, for public expectations are not based on optimizing one index, but on balancing many values.

5. Discussion

H. P. Jost, K. S. Paul Products, Ltd: It appears to me listening to Dr. Morris and Mr. Wakeland that, as a society we are more and more pressed and more and more willing to accept large numbers of fatalities. I wonder whether 10 years ago we would have been prepared to accept the possibility, if not the probability, of 350 people being killed in one air accident but with the use of jumbo jets we seem to be accepting it. Perhaps one day we will have to accept the possibility of 10,000 people being killed in one accident.

W. D. Compton, Ford Motor Company: The determination of the risk associated with a failure and the post evaluation of a failure is one thing; the translation of this information into standards or regulations to be used in subsequent designs is quite something else. Do you foresee this sort of thing happening? Do you think that a set of criteria or regulations will evolve, based on information from failure investigations, that can be applied to a broad range of situations in which the cost effectiveness may be quite different?

H. H. Wakeland: Yes, certainly it's going to happen. It is the nature of humanity to evaluate things in the way they are put before them, not in the way that a cost benefit analyst, even though he may be working under OMB's strictures or under the orders of the Joint Economic Committee of Congress, would evaluate them. It is very unlikely that the Silver Bridge investigation was a cost effective action. However, it fulfilled another of society's basic social concerns and needs. You have to remember that the first thing stated in the Declaration of Independence is not "cost benefit and maximum efficiency for all", but "all men are created equal". The sign above the Supreme Court does not say "maximum cost effectiveness for minimum fatalities in society as a whole". It says "equal justice under law". Equal justice under law means the sort of thing that you manufacturers reach in product litigation after an accident. That's entirely an equality and equity action. It may or may not come out to be cost effective. But society is going to go on asking for those things simply because that's the way the people think and that's the way Congress is organized.

W. D. Compton: Isn't it also appropriate to ask that there be something developed for a review of these criteria as technology changes so that one isn't faced with a long list of possibly obsolete restrictions?

H. H. Wakeland: It's very appropriate to ask for that. What should be asked for from organizations like Congress is a clear look at the priority system being used. Are priorities decided on a cost benefit basis or on a safety basis? These values conflict with each other and someone must decide which point of view will prevail. In the meantime, conclusions reached by the public on the basis of the sheer horror of a situation are not wrong conclusions. They may be uneconomical; they may steer us in the wrong direction; it's up to us to decide. I'll give an example. We were before the Transportation and Aeronautic Subcommittee of the House Interstate and Foreign Commerce Committee recently. We wanted to talk about soft cushions for the front of railroad locomotives for which the Federal Railroad Administration has a feasibility study that shows a ten to one benefit to cost ratio in lifesaving in a field which involves one-half of all the fatalities now assignable to railroads. They didn't disagree with us, but they wanted to talk about requiring money to be spent for inspectors to look at broken rails, even though broken rails are responsible for only about six fatalities per year. When we have our values organized and when they have their values organized, we can discuss the priorities intelligently.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Public Safety and Construction Surety*

Jacob Feld

Consulting Engineer
114 East 32nd Street
New York, New York 11016

Structures are conceived, designed and constructed for use by man and located in proximity to other structures and open spaces occupied by man. How safe are these structures when first occupied, and what degree of safety can be expected of them with age?

Who is responsible for the original condition of safety of the structures, and for the degradation of safety with age and use?

Who is responsible for the maintenance of sufficient safety to permit continued use and occupancy?

In addition to the broad aspects of the problem, the paper will describe avoidable use of incompatible materials that caused local distress or failure of the structure.

Public safety and construction surety depend on proper control of each factor involved in the design, construction, maintenance, and use of a structure. Any deviation on the part of any one factor makes the structure vulnerable to damage or even collapse. Absolute safety is not obtainable. The historical record of structure behaviour indicates the high reliability of present techniques and procedures.

Responsibility for defects and for the small percent of structures that do not perform properly is a deterrent to proper control, but considerable pressure is needed to further reduce the number of failures to insignificant numbers. It is the duty of the professions involved to provide that pressure, since legal and political action has not been successful, and to control the proper performance of the work in every factor and so provide public safety and surety in the construction industry.

Key words: Construction safety; construction surety; degradation of safety with age and use; maintenance of safety; public safety; safety responsibility.

1. Introduction

Structures are conceived, designed and constructed for use by man and located in proximity to other structures and open spaces occupied by man. How safe are the structures when first occupied and what degree of safety can be expected with age? Who is responsible for the production of safety in the structures, and who for the degradation of safety with age and use? Who is responsible for the maintenance of sufficient safety to permit continued use and occupancy?

Historical Notes: The problem of safety in construction must be separately treated for incidents, accidents and even collapse occurring during the production of the work, from the condition of the completed structure and its reaction to use and to natural causes of deterioration and reduction in its resistance against failure. Most of the accidents and partial collapse in structures come during the construction period, often with serious effect on the life and well-being of people involved in the work as well as the innocent public by chance within the area involved. This discussion is, however, limited to the accidents coming after completion when public use is begun and the people involved in the construction are away from the area.

The problem of safety has a long history and there are many references even in ancient regulations and discussion setting forth the limit of responsibility and of expectation. If the code of Hammurabi (about 2000 BC) is typical of ancient regulations, few engineer builders had the opportunity to learn from their own mistakes.

> "If a builder build a house for a man and do not make its construction firm and the house which he has built collapse and cause the death of the owner of the house—that builder shall be put to death."

Certainly such strict punishment for failures must have been a deterrent to shoddy construction practice.

*This presentation is based on a paper presented to Section of Engineering, New York Academy of Sciences, December 10, 1969.

Some warnings in the Hebraic literature against building on shifting sand may stem from experiences encountered when the Hebrews were the construction battalions in Egypt. That entire ancient city complexes collapsed from normal aging as well as from enemy action and tectonic disturbance, is found in the archeological exploration of separated layers of rubble. There is no record to indicate if any mortal was judged responsible for the collapses and the loss of lives.

In Greco-Roman times the construction industry was in the hands of trained slave artisans and successful work was rewarded by gifts of substance as well as of freedom. With no competitive financial or time schedule to worry the builder, work was well done and with apparent great success, as is seen in the many examples still existing. Similarly, in the Medieval period time was no object and the great successful religious and governmental structures have lasted for centuries. One doubts that our modern works will exist to comparable age.

In the common law developed in England, as is found in the 15th century court records from the reign of Henry IV, the rule was "if a carpenter undertake to build a house and does it ill, an action will lie against him." Of course, by "ill", is meant "not well", and if all work were done well, there would be no record of failure. So historically the burden rests on the construction industry to see that work is done "not ill" and each partner in the industry, from the architect who conceives the project to the foreman who directs the performance of a part thereof, must lend every effort to avoid and eliminate every possible cause of "ill" structures.

The Napoleonic code has become the basis of the common law wherever the original settlers were French. In many ways there is a greater responsibility placed on the designer and the professional in charge of the work as agent of the owner, to safeguard the investment and guarantee proper and adequate performance, than is expected under English legal procedure.

In the period from 1870 to 1900, construction became a large industry in the United States, chiefly influenced by the expansion of the transportation system with its necessary bridges, storage facilities, and industrial plants. Comparatively little is recorded of failures in buildings, most of which were heavy masonry wall bearing structures with sound timber floors. Bridge spans soon exceeded the capability of timber trusses and a great competition in the sale of iron spans resulted in many failures. These were spectacular and were almost daily newspaper headlines. Even foreign technical journals commented on the great number of unsuccessful designs of bridges in the United States. Engineering magazines from 1875 to 1895 were as full of reports of railroad accidents and bridge failures as today's daily newspaper reports of automobile traffic accidents. Even as late as 1905, the weekly news summary in *Engineering Record* described the most serious railroad wreck of the week, usually tied in with a bridge failure. The *Railway Gazette* in 1895 published a discouraging summary of iron bridge failures resulting from railway traffic, listing 502 cases in the period from 1878 to 1895 inclusive, collected by C. F. Stowell, and noting that the first 251 occurred in 10 years while the second 251 occurred in only eight years. In the years 1888–1891 inclusive there were 162 such accidents. These reports received wide publicity and attention and must have had considerable influence on the designing engineers as well as the bridge salesmen of those days.

In this discussion, the term "safety" is defined in the Webster Dictionary as:

(1) freedom from danger or risk
(2) freedom from injury
(3) harmlessness or being without hurt, loss or liability

and "construction" is the man-directed assembly of inanimate components to make up:

(1) shelter for man in his various activities
(2) tools to extend his power and mobility
(3) instruments to extend his senses.

2. Absolute Safety

Absolute and unqualified safety at all times can only exist if the designer has prepared a plan entirely free of any error in judgment or computation, the contractor has not built in hidden errors which void the intent of the design, the owner maintains the structure so that there is no degradation of strength in any component and the public does not exceed the expected loadings, both static and dynamic, nor modifies the strength of the structure either by removal of support or by imposition of new conditions from adjacent construction or demolition.

One cannot economically design for zero probability of failure, nor for infinite life of a structure. There is always a risk in being alive or even in being; the only certainty against death of a living being is not to be alive.

The concept of a factor of safety in structural design came about a century ago. If the ratio of available resistance to the effect of the loading always remains greater than unity during the existence of a structure, safety is assured. But the factor of safety is merely a ratio of two assumptions and is therefore set by either regulation or agreement at values large enough to allow for considerable error in assumption. For normal design of buildings 2.5 was considered sufficient protection and lately has been reduced to about 1.6, for bridges 3.0, but for the stability of earth mass embankment or dams, 1.5 is normal, since factors of safety of 2.5 or 3 would bring the cost so high that these earth structures would never be constructed. Under extraordinary water pressure loading, a computed factor of safety of 1.1 is often considered sufficient. One can readily see why the relative incidence of failure is greater in earth structures than in structural frames or mansonry buildings.

The proper choice of a factor of safety is often dependent not only on the physical design procedure and uncertainty concerning material strengths and construction control, but also on sociological and psychological effects of visible cracking which may not be structurally significant, on political grounds to avoid public criticism if failure occurs, and on economic grounds when the cost of extra strength is evaluated.

Where failure, should it occur, would not affect life, many designers and owners are willing to take a calculated risk in degrading the normal value of the factor of safety. In foundation design and performance, Professor George F. Sowers of the Georgia Institute of Technology has warned that more care is necessary since designs are becoming increasingly more daring, modern structures are now less tolerant to movements from foundation settlement, sites for new construction are becoming poorer in bearing value and populated areas, with their load imposition on the land, are taking up a larger proportion of the total available areas.

Separate designs of foundation and superstructure, each with ample factors of safety, in combination can prove insufficient if the two are not compatible in the elastic and plastic strains which will result. Absolute safety requires complete evaluation of strains which come from loading and provision for control of the strains to within the safe limits of the various materials used in the structure.

Similarly, bridge design over water courses cannot economically provide resistance against a 1000 year flood for a bridge expected to exist 100 years. During the 1964–5 floods, scour of piers and foundations caused some 76 million dollars of damage. If all the structures were designed for a 50 year flood expectation, the risk taken could be economically justified. Yet a 1000 year flood has a 2.5 percent chance of occurrence in a 25 year period and a 10 percent chance in a 100 year period. Even if we could predict the magnitude of a 1000 year flood, this would not be a certainty of safety in the 100 year life of a bridge.

Guaranteed safety of unusual bridge designs in the U.S.S.R. is proof tested with the design engineer and the construction supervisor standing under the bridge during the test. This may sound as fiction, but a Soviet engineer of railway bridges at an International Prestressed Concrete Congress held in Rome in 1961 showed a photograph of such a test with two lines of locomotives on his design with the constructor and himself standing under the bridge during the test.

Structural integrity under all possible forces requires continuity in the structural design, complete compliance in construction performance and soundness of the finished structure. Failure of structures during earthquake loading is often blamed on lack of continuity which would provide alternate support after local failure. For every construction failure there must be many near misses with some small excess strength permitting the structure to weather the storm. Can these structures be classed as having absolute safety against the next storm?

The criterion of public safety was analysed by David E. Allen of the National Research Council of Canada from data collected in both the U.S.A. and in Canada of fatal accidents as a ratio of exposure. For one hour of exposure, either occupancy or use, the death rate times 10^{-9} is listed as:

Failure in completed structure	less than one
Construction worker	218
Failure of temporary structure element	35
Bus or rail travel	80
Walking along a road	420
Automobile travel	1040
Air travel	2400
Swimming	3500
Motorcycle	4420
Alpine rock climbing	40,000

The incidence of fatality in a completed structure indicates almost absolute safety, yet failures do occur. Allen estimates that to reduce the probability of failure from 10^{-2} to 10^{-6}, the cost of completed structures would increase by 6 to 7 percent, but the cost of structural elements and assembly methods would increase by 20 percent.

3. Reliability or Insured Safety

Reliability stems from four separate controls in the construction program. The designer assures that his analysis and choice of materials will produce a safe structure. The contractor assures that his assembly conforms to the design and that the materials furnished conform to the desired strengths. The owner assures that the building will not be used under floor loadings exceeding the basis of the design and that the maintenance program will keep the materials and assembly in substantially the original state. The public represented by the governmental agencies assures that adjacent conditions will not be altered so as to degrade safety of the structure.

A perfectly produced design and carefully controlled construction does not guarantee perpetual life of a structure. Maintenance must be provided to prevent change in the material characteristics. Some of the natural agents causing deterioration and loss in safety margin are oxidation, solution, dehydration, solar radiation, thermal reversals, fatigue from load reversals, plastic flow, internal stress corrosion, bacterial action and many others. They require special built-in protections, such as timber preservatives, anodic corrosion protection, control of thermal effects, sealers and paints. But even the protections require renewal with age.

As an example of public responsibility, some simple cases of failure caused by extraneous conditions are cited. With the advent of steel bridges, many failures of bottom truss members were attributed to dogs. The steel members under the Riverside Drive Viaduct near 125th Street were eroded almost to the point of failure by the contact with acidic nesting accumulation from the local pigeons. After major repairs, the underside was sheathed in aluminum with no ledges on which the pigeons can roost and nest. Air contamination, especially the sulfur content, modifies and weakens cement mortar. Dewatering of adjacent areas causes a loss of foundation support, usually non-uniformly, with tipping and tearing apart of structures.

In 1846, Robert Stephenson, then president of the British Institution of Civil Engineers, advised the members to meticulously study failures of structures to determine limits of design reliability. Freudenthal, in his 1950 book on "Inelastic Behaviour of Engineering Materials and Structures", warns that a factor of safety in the design is only a measure of uncertainty, since it compares design stress, an intellectual concept, with measured material strengths. The designer should analyse the variability of all influences, during the expected life of the structure, on available resistances and on the service loads and external conditions.

Designs are based on assumed failure patterns. Carl J. Turkstra in discussing choice of failure probability provides as a good approximation for the best economic design the criterion that no further strength is needed when the increment to the basic design cost divided by double the economic loss resulting from failure is effectively zero. Just how close to zero is a decision based on experience and judgement.

At the Washington International Conference on Structural Safety and Reliability, held in April 1969, certain basic warnings were sounded. There is a limit to statistical application of structural reliability prediction. Structural failure is an emergency, whereas electronic failure is a contingency, hence more susceptible to statistical analysis. There is great need for study of competence in predictions of reliability of materials. In the case of fighter aircraft, an empirical accumulation of data, statistically analysed, shows areas for more intense inspection and control. No such procedure can be of great value for structures. There is at present no testing procedure for design and life estimation of fatigue sensitive structures. Witness the collapse of the Silver Bridge over the Ohio River in 1967. Engineering designers concentrate on technical matters that they can control. But the investigation of failures shows many parameters outside of direct control, not physical in nature, which should be considered.

The reliability of materials as delivered to the site for assembly is one of the questionable facets of producing safety. There are three classes of available materials: those proven by years of use, recent improvements of older materials, and completely new materials. In the second class are such acceptable items as new steels, better cements, laminated timber, Plexiglass and aluminum alloys. Newly developed materials are rapidly marketed with little historical proof of suitability. A large amount of research is needed to prove any new material as capable of providing required long time resistances.

Considering soil as an engineering material, it must be kept in mind that the design depends on results of laboratory tests on small samples that one hopes are truly representative and also that the interpretation of the tests is structurally applicable.

The failure mechanism chosen for determining safety may not necessarily be the one with largest probability of failure. In both soil and structure problems, distortion can be tolerated without collapse, but the effects of fatigue, wear and corrosion will then set a shorter life to the structure.

As an example of the basic problem, reliability of welding can be cited. It is recognized that flaws in the weld, just as flaws in any example of glue adhesion, have a detrimental effect on strength. Various test procedures have been developed for nondestructive testing of welds, such as radiography and ultrasonic scanning. How and who interpret the test data? Radiography shows a two dimensional picture of a three dimensional problem. Ultrasonic waves reflect from surfaces at a discontinuity, which may not be in the weld volume. The old welding specifications required welds to be "sound"; we are still looking for a foolproof test of soundness. D. A. Olsson of the Bethlehem Steel Company tabulates failures from welding in order of frequency as:

ship hulls—fatigue, gross failure, brittle failure.
cargo handling—brittle failure, gross failure, fatigue.
bridges—brittle failure, fatigue.
buildings—gross failure due to loading above yield point.

The marked reduction in hull failures after the National Research Council investigation and imposition of criteria for ship assembly following the World War II experiences, shows that research and control can provide greater reliability.

Reliance upon compliance with established regulation as proof of reliability was subjected to some question in 1969, when the revised New York City Building Code increased the minimum wind load to be resisted by multi-story frames. What about the reliability for safe use of the thousands of structures designed, approved and constructed under the requirements of the previous code, all of which bear legal certificates of occupancy. Are they now of questionable sufficiency? Similarly the 1969 Code of the American Association of State Highway Officials, the recognized criteria for design of all highway bridges, modified all earlier editions in a considerable reduction of safe stress for members or portions of members subjected to tensile force if any welding, even of such minor items as clips, brackets or studs, is performed. This reduction is listed under the heading of fatigue from change in stress level under live loads, not necessarily requiring reversal of stress. The effect is to degrade allowable traffic loading on bridges designed under the provision of earlier codes, or in other words, a sudden reduction of reliability of the expected strength of these structures. If reliability is referenced to consistency with code provisions, the date of the code becomes significant.

4. Responsibility for Safety

The total responsibility for safety of a structure is therefore the sum of the warranty or surety which can be given by the four separate entities involved. The designer can only assure of a safe plan for assembly of safe components; the constructor, of proper materials safely assembled; the owner, of control of loading to the maximum design assumption and of continuous maintenance to prevent degradation of material strength; and the public, of control of extraneous conditions which would void the proper assumptions of the design. Each factor must assume responsibility for its duties. Recent legislation in New York State crystallized this responsibility and forbids by agreement or dictate, the transfer of responsibility to another.

Various decisions by courts of the several states control the legal responsibility of the designer. In the case of Gage vs. Buttram, a California decision is based on: "The general rule is applicable that those who sell their services for the guidance of others in their economic and financial and personal affairs are not liable in the absence of negligence or intentional misconduct. . . . Those who hire experts are not justified in expecting infallibility but can expect only reasonable care and competence. They purchased service, not insurance."

And in another case, the decision reads: "For defects in the original plans and the approval of detailed plans arising from negligence on the part of the architect, liability resulted". So the Architect-Engineer is liable for negligence and for breach of an implied-warranty. The period for legal liability has been limited by some states to an average of eight years (from 4 to 12 years in different states) usually from the date of completion of the agreed services to be performed. Several states, including New York, have no time limitation for such liability.

The Architect-Engineer has three areas of legal exposure in the practice of his profession:

(1) Liability to the owner for defects in plans and specifications or for negligent supervision.
(2) Liability to third persons, not involved in his contract for services, for personal injury resulting from negligence.
(3) Liability to the contractor for interference with his rights in performing his work, and other miscellaneous types of indirect liability.

The general definition of negligence liability often quoted is by Professor George M. Bell in "Professional Negligence of Architects and Engineers" (Vanderbilt University Press, 1960). "An Architect-Engineer does not warrant the perfection of his plans nor the safety or durability of the structure any more than a physician or surgeon warrents a cure or a lawyer guarantees the winning of a case. All that is expected is the exercise of ordinary skill and care in the light of the current knowledge in these professions. When an Architect or Engineer possesses the requisite skill and knowledge common to his profession and exercises that skill and knowledge in a reasonable manner, he has done all that the law requires. He is held to that degree of care and skill and that judgment which is common to the profession."

The question often arises as to the safety of an existing structure and how can it be evaluated. An answer has been suggested by Committee 437 of the American Concrete Institute whose mission is the study of Safe Loads for Existing Concrete Buildings, as either a strength evaluation by analysis or a load test for any structure that has undergone general or local damage, excessive loading, impact, blast, fire, earthquake, war damage or deterioration caused by the environment. To set the criteria for such evaluation, full scale loading tests to failure are planned on existing structures scheduled for demolition, in abandoned exhibitions and in redevelopment areas. Some valuable empirical data has been collected on degree of compliance with the plans, and on the extent of deterioration.

As to recent new construction, similar indictment of the financial interests recently came from such diverse parties as Engineer A. A. Golder in Canada and A. C. Jenkins, then President of the American Institute of Architects, for allowing the erection of sub-standard buildings with too much emphasis on aesthetics for rentability and too little for utility, as affecting public health, welfare and safety.

Under the subject of "Social Benefit and Technological Risk" Chauncey Starr in "Science" of 19 September 1969 describes the cost of accidental deaths from technological developments for public use, especially since engineering developments are seldom tried out in small, localized areas test out the risks. There is now a very rapid acceptance of new designs and voluntary risks are accepted readily by the public, usually induced to do so by advertising and general awareness only of the advantages.

The problem is not limited to the United States. In 1966, after a rapid succession of structural failures, mostly of novel designs, an editorial in the British "Construction News" asked "When is it going to stop?" All of these happenings prompted people to ask whether or not there was not some underlying common reason or a vital link missing in the safeguards. The editor suggests two possible reasons:

(1) The system and method used for checking both design and construction.
(2) The allocation of responsibility.

England does not license professional engineers. Within the London County Council there are 28 district surveyors with absolute authority, who check structural calculations and deal with the details of construction stage by stage. Outside of London, local enabling legislation and Building Regulations control all new construction but with little personal contact with the work. A common national system is being studied with the background that almost all the failures were outside of the London area.

Sweden places responsibility for quality control and adherence to design and specifications on the contractor's engineer and this function is taken seriously. France, Holland and Belgium place all responsibility for checking design, material strengths, construction methods and performance on groups of experts hired by insurance carriers. The owner pays the premium and is insured against collapse, defect or abnormal future maintenance costs, both during construction and after completion. Insurance is not available to every project, only where the designer and constructor are considered qualified.

In Rio de Janeiro, after the collapse of an eleven-story apartment, some 30 minutes after the city engineers had inspected the building and declared it safe, the "O Cruzeiro", a leading paper in Brazil published a list of 20 buildings that collapsed in Rio between 1947 and 1957, and also identified a number of major structures on the main streets, Rio Branco and Presidente

Vargas Avenues, which were eight to twelve inches out of plumb. Many tenants including government agencies moved out. However, the City did nothing to institute better control.

A local newspaper in Italy, after the collapse of a new apartment in Barletta in 1960, taking 58 lives, said: "Here in Italy we prefer to weep for our dead rather than perform an autopsy and find out why they died." Five stories had been put on top of an existing one story garage building, avoiding all construction laws and by some passing of graft, the chief health officer issued a permit for use and occupancy before the work was even finished. The trial of a dozen engineers, contractors, health and construction officials resulted in a total of 70 years jail sentences.

The worst recent failure, also in Italy, was the Vaiont Dam disaster, claiming 2300 lives. In spite of continuous day and night inspection of conditions, some 700,000 cubic meters of rock slid as a mass into the full reservoir causing a tidal wave over the dam and into the villages below. Investigations and litigation went on for several years to pin the responsibility for not heeding the warnings of incipient slide failure.

5. Summary

Public Safety and Construction Surety depend on proper control of each factor involved in the design, construction, maintenance and use of a structure. Any deviation on the part of any one factor makes the structure vulnerable to damage or even collapse. Absolute safety is not obtainable. The historical record of structure behavior shows a high reliability from present techniques and procedures but with too many failures. Responsibility for defects and for the very small percent of structures that do not perform properly is a deterrent to poor control, but considerable pressure is needed to further reduce the number of failures to insignificant numbers. It is the duty of the professions involved to provide that pressure and control the proper performance of the work to provide public safety and surety in the construction industry.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

The People Considerations of Automotive Service

F. J. Uhlig

GM Technical Center
General Motors Corporation
Warren, Michigan 48090

The presentation will cover some of the human factor considerations in servicing automobiles at GM dealerships. Emphasis will be placed on the communication channel between customer and mechanic and possible attendant breakdown that can limit the effectiveness of repairs. Specifically, the repair challenge will be described relative to:

(1) Vehicle relation and consumer expectations
(2) The customer item-service writer interface
(3) Technical service information for internal service shop communications.

In addition, the presentation will cover some recent developments of GM service research which address themselves to the above problems for example:

(1) Service writer check sheet—a vehicle simulator oriented approach to customer order write up.
(2) "STAR"—A computerized reading comprehension method for use in improving the readability of service manuals used by mechanics.
(3) "USDA"—Universal Symbol Diagnosis Approach—that uses graphic symbols in a logic tree format to aid in vehicle system problem diagnosis.
(4) Shop operation management system.

Key words: Automotive repair; automotive repair effectiveness; automotive service; customer-mechanic relationship; human factor considerations; universal symbol diagram approach.

In keeping with the theme of this symposium on failures, I would like to address some consequences of failure associated with a complex mechanical system called an automobile. Specifically, I wish to focus on some people considerations of automotive service after highlighting the statistical impossibility of eliminating mechanical malfunction.

According to the U.S. Office of Consumer Affairs, motor vehicle service is the primary area of consumer dissatisfaction in the nation. Studies have also been made by the subcouncil on performance and service of the National Business Council for Consumer Affairs. In this report, data were presented from the Harris-Life Survey, conducted in February of 1972, that indicate the extent of consumer concern over keeping his automobile running; 40 percent of the respondents expected difficulty in obtaining repairs for their automobiles and 31 percent believed auto repairs were made improperly, up from 16 percent in 1967.

It is generally agreed that the failure rates of most current consumer durables is lower now than in years past. However, the sheer volume of consumer items in use results in a large volume of consumer complaints being registered with the manufacturers, dealers, service agencies, and public and private consumer protection organizations.

The National Business Council report went on to say that "no matter where complaints are first directed, manufacturers are usually blamed for product sales or service deficiencies." The manufacturers' challenge then is to reduce the probability of malfunction.

By specification, motor vehicles are required to perform well under a great variety of operational and environmental conditions. Due to these conditions and the inherent variability of our production process, there are times when the product does not perform its intended function, that is, it fails in some manner.

Specifically, we have a situation which can be illustrated as shown in figure 1. On the left of the figure we plot a curve which can be used to estimate the probability that a user will impose a given load or requirement on a part or system. On the right, we show a curve which can be used to estimate the probability that a part or system can handle a particular load or require-

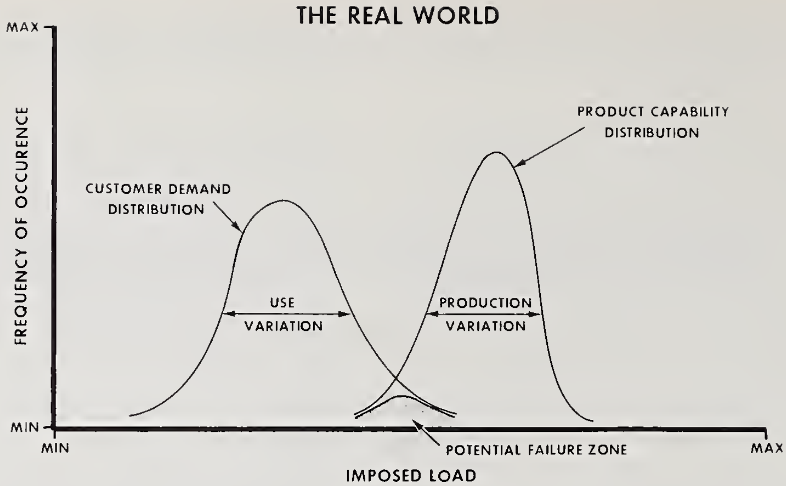


FIGURE 1.

ment. As shown here, some few users with higher-than-average requirements may encounter some few parts with lower-than-average capabilities. When this happens there is a possibility that failure will occur. This is represented by the potential failure zone. One might suppose that all we have to do to control or eliminate failure is to control or eliminate variation. Just how feasible is this?

First, let's consider customer demand. There is no product I know of that is subject to a wider range of uncertain use conditions than a motor vehicle. These include variations in road surface, climate, loads, operation time, operator skill and maintenance. The combinations are infinitely varied and are continually changing. The ability to have a thorough understanding of customer use variation is a challenge to the motor vehicle manufacturer all by itself.

Now what about the product? We make about six million vehicles a year in about 300 different models. There may be up to 15,000 parts in each vehicle, as many as a hundred variables for each part, thousands of manufacturing and assembly steps in putting all the parts together. and millions of people involved in its manufacture and maintenance. Someone has estimated there are one and a half million places for variation in a typical vehicle. Variation is simply the name of the game.

Obviously, we can't eliminate variation, but we perhaps can control it. A fundamental way to reduce the failure zone is to reduce the curve overlap. The spread of the customer variation curve could be reduced or shifted to a lower range by government regulations which affect drivers, road conditions or other use factors. But there is little the manufacturer can do about customer use variation.

On the other hand, the product variation curve can be altered by the manufacturer in two general ways. One method, as illustrated in figure 2, is to reduce the

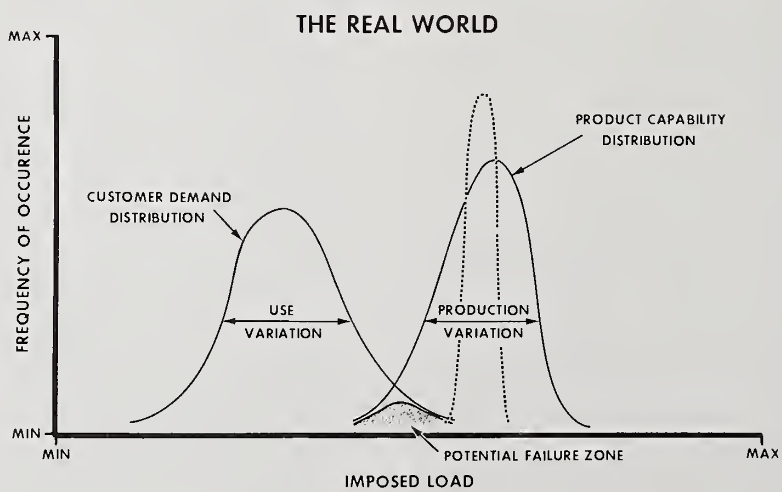


FIGURE 2.

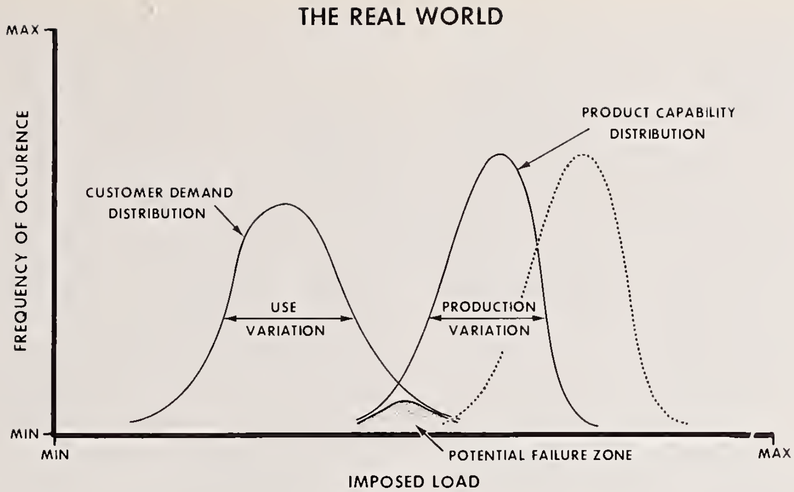


Figure 3.

spread by holding very close tolerances, limiting product variety or providing many tiers of inspection and screening.

Another method as shown in figure 3, is to shift the distribution to a higher capability range by possibly using heavier structures or more exotic materials.

Now that I have touched on some of the technical realities involved with automobile usage. I would like to focus on some human factors or people considerations that we judge are deserving of examination–human factors and reactions that become consequences of a vehicle failure incident.

In our studies of the automotive service system, we have been able to identify some of the vulnerable points that can cause the system to breakdown from a human factors standpoint. Psychologists tell us that one of the most important and pervasive ingredients in maintaining effective interpersonal relationships is communication; without effective communication, negative human factors consequences will result.

I would like now to treat some specific communication elements or links within the dealer service system that we judge are in need of help. In addition, I would like to explain what GM has done, or is doing on a trial basis, to increase the signal to noise ratios in these communication links, and thereby help to mitigate further possible adverse consequences.

In looking at the communication networks within a service shop, and there are many, three main channels are deserving of special attention.

In order for a mechanic to perform effective repairs, it is required that he has an accurate description of the owner complaint. Since this description is initially prepared by the service advisor in consultation with the owner, it is imperative then that the service advisor capture this vital information on the repair order in mechanic style language.

One effort we have made to improve communications in this vital area is the development of a rather simple device called a Service Advisor Check List. This printed form covers both sides of an 8½ by 11 in sheet of paper, and is a rather complete list of common automotive service problems along with their probable cause. Major classifications of service problems such as "noise in engine compartment" or "air conditioning" are grouped together on the form. Within each major classification a list of typical symptoms or problem descriptions are shown. The object of this list is to help the customer and service advisor arrive at the most accurate description of the service problem.

By going through a question and answer series, a general complaint of hard starting might be narrowed down to a specific problem like "hard starting in wet weather;" a particular situation that may not be evident in a general check-out in the service shop. Thus, the mechanic can be guided to the specific problem area with a high probability of correction the first time.

In practice, the service shops using the check list also have enlargements of the list displayed near each mechanic's work area. This allows the repairman to see the list of questions that led to the service instruction on the repair order. The information shown on the check list was developed by General Motors Institute for GM Service Research by interviewing scores of service managers, service advisors and technicians. The appearance of the chart has been modified several times, and further refinement is anticipated after more field experience. A field trial of this communication device is underway in select dealerships to determine the practicality and merit of daily use.

The second major communications channel that we judge could be improved involves technical service information generated specifically for use by the mechanics. This type of information most often is in printed form, and the most general example is the shop service manual. Ostensibly, the primary purpose of the shop manual is to provide the mechanic with a compilation of the necessary procedures and specifications to enable him to diagnose and repair all major systems of a car.

As a communication tool then, it is expected that the mechanics are able to comprehend what they are reading in the manuals and apply the information in conjunction with tools to the repair of automobiles. We took a look at the writing contained in many of our manuals, and found that much of the written content was not expressed clearly in simple terms. We reasoned that if a mechanic cannot understand what he is reading in a manual, then a high probability exists that repairs could be done incorrectly leading to customer dissatisfaction. Ineffectual communications to the mechanic, therefore, could produce subsidiary adverse consequences. Readability of service manuals became a target for improvement.

Research into ways to measure reading ease showed that several measuring methods had already been developed. Of these methods, the one developed by Rudolf Flesch best fitted the needs of GM Service Research. Dr. Flesch based his Reading Ease Formula on sentence length and syllables per word. He found that the higher the syllable count and the longer the sentence, the harder it is to understand the sentence meaning. Dr. Flesch's formula gave the research group a means to measure the reading ease of written material. Tests on sample text confirmed that shop manuals are hard to understand. It was found, in fact, that the equivalent reading grade levels required ranged from 12th to 16th. We questioned whether mechanics could really read on that level.

At present, the only available reading skill measurements are on future mechanics—those students now enrolled in vocational school programs. The reading comprehension of 239 students in the twelfth grade studying automotive repair and diesel engines was tested. The students were given a comprehensive reading test with the following results: only three percent were comprehending at the twelfth-grade level, the majority at ninth grade and twenty-five percent of the class sixth grade or less. We must remember, however, that these students are in the vocational arts and usually represent the lower half of the twelfth grade in reading achievement. A textbook written to a twelfth grade level would be understood by about thirteen percent of the class. To reach at least half of the group, the text must be written to a ninth grade level. If this is any indication of the reading ability of our mechanics, books for their use must be rewritten to just below ninth grade level.

In our initial analysis of shop manual readability, each measurement had to be done manually by having someone count words, syllables and sentences, and apply these numbers to the Reading Ease Formula. A quicker way to test was needed, that is how the Flesch Reading Ease Formula and the computer came together. The result is a program designed for a common computer that can be used by anyone with a little experience. The computer program provides a fast, useful measurement of reading ease. We have called this program "STAR", an acronym for A Simple Test Approach For Readability.

In working out this theory of reading ease, Dr. Flesch was convinced that two factors in written material tend to reduce understanding. They are long sentences and long words. Flesch saw that modern English writing has come a long way in reducing sentence length; but that it has lessened this gain by using longer words. This tends to cram too much meaning in too small a unit. This "density" of language is now the main problem in efforts to write clearly.

In using STAR to sample the readability level of written material, the text is entered into the computer through a keyboard-type terminal tied into a time-sharing system. Unlike other programs, plain text is used without any additional program markers. There are only a few special instructions used when typing the text into the time-sharing computer. As an example, a portion of our automative air conditioning manual was found to have an equivalent reading grade level of 16. To reduce the reading level, the sentences were shortened by breaking up clauses and substituting "low calorie" words for heavily loaded words; cooling for air conditioning . . . car for automobile . . . and inside for interior. Rewriting the portion reduced the grade level by 4 down to a 12th grade level.

Splitting up sentences and finding replacement words creates a real challenge to the writer trying to communicate. It cannot be an automatic process just to change words . . . or as some have suggested, a computer program to do just this substitution . . . it requires a writer who knows his subject and has a good working vocabulary. We cannot replace a writer with a machine, nor is it recommended.

General Motors is now in the process of applying STAR on an increasing basis to all of our service manuals. It is our goal to make the language in shop manuals as effective as we can in communicating the service information to the mechanic. General Motors is making this program available to all interested parties through a booklet called "S.T.A.R." It is available through GM's Public Relations Department, Detroit, Michigan. The booklet contains background information and the complete computer program instructions written in basic computer language. This will help a computer programmer to place the STAR Program into action. You may be interested in knowing that this presentation was analyzed by STAR, and the required reading comprehension grade level is thirteen.

As we continued to study the communications effectiveness of our service manuals, it became apparent that more was necessary than just improving the readability. The construction and format were judged hard to follow for a mechanic, and had complicated approaches for presenting diagnostic instructions. Our interest was directed to work that numerous highway departments have done in standardizing road signs using graphic symbols to communicate important information. Examples of these symbols are familiar to all of us. Once a person is made aware of the meaning of each symbol, one no longer needs to read the words; the meaning is conveyed at a glance.

Last year we decided to try this approach on a diagnostic procedure for a new system that was being introduced on 1974 models and would be common to all General Motors' car lines; namely the seat belt starter

STARTER CRANKS WITH SEAT BELT UNFASTENED

TEST CLOCK FUSE AND DIRECTION SIGNAL FUSE. REPLACE IF NECESSARY. CHECK FOR GOOD CONNECTION AT THREE WIRE CONNECTOR.

- STARTER DOES NOT CRANK → SYSTEM NOW OK
- STARTER CRANKS → REMOVE BY-PASS RELAY
  - STARTER DOES NOT CRANK → DEFECTIVE BY-PASS RELAY
  - STARTER CRANKS → GROUND OK. GREEN WIRE AT THREE WIRE CONNECTOR
    - STARTER DOES NOT CRANK → PROCEED TO SEAT SYSTEM DIAGNOSIS
    - STARTER CRANKS → CHECK CONTINUITY OF DK. GREEN WIRE FROM INTERLOCK RELAY TO THREE WIRE CONNECTOR. CHECK FOR 12 VOLTS AT INTERLOCK RELAY PINK WIRE. REPAIR WIRES AS NECESSARY. IF WIRES ARE O K, REPLACE INTERLOCK RELAY.

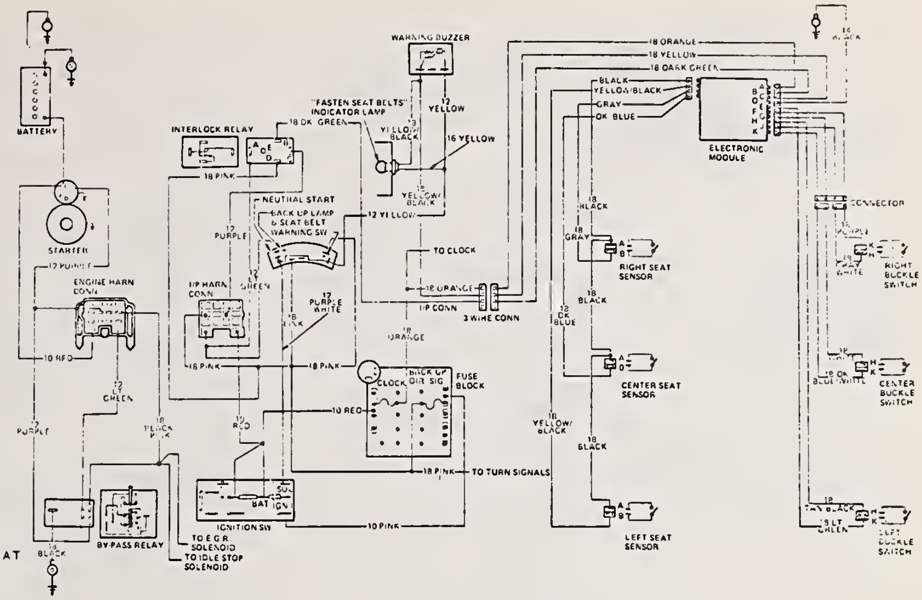


Figure 4.

interlock system. Figure 4 shows a portion of the diagnostic procedure for this system that, if put in the usual format, would have looked like this in service manuals. The failure mode is that the starter motor cranks with the buckle unfastened. In looking at the procedure, we determined that the graphic symbols shown in figure 5 might be applicable to the starter interlock procedure and be readily understood by a mechanic.

Our first attempt was made utilizing both symbols and pictures to guide the mechanic into the proper repair as shown in figure 6. You will note that the pictures are as the wiring and connectors would appear to the mechanic working on a car and not just a picture of a component out of its environment. You will also note that the system failure complaint is shown in symbolic form at the top of the page. It was determined at the beginning that words should be used along with the symbols until enough familiarity could be developed to eliminate them.

Various attempts were made to develop a format that seemed less confusing and easier to follow than the one you just saw. Again, selected words were added next to the symbols to make it more understandable. The final approach at formating is shown in figure 7. This is what has proven to be the most usable, the most flexible and the most understandable. It is the format as it appears in our 1974 Service Manuals.

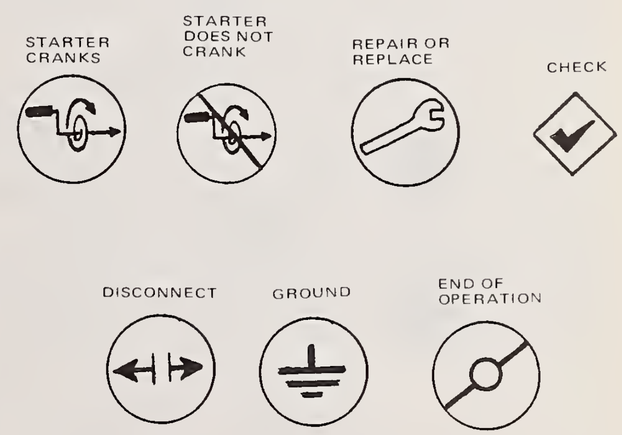


Figure 5.

This symbolic system was tested on mechanics possessing varying degrees of experience and skill. We found that highly skilled mechanics had no objection to using it and found it effective. Also, mechanics with moderate skill, who had difficulty in following the old word-tree format, were able to use this one satisfactorily. And most importantly, mechanics of low skill level, who

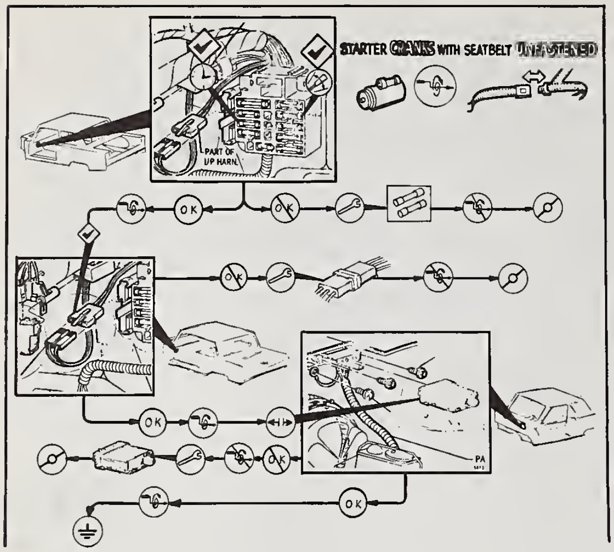


Figure 6.

found it impossible to use the old word-tree system, were able to follow the new procedure and, in fact, fix the car with little difficulty.

As an interesting sidelight, one mechanic with little experience, was still having difficulty following even the new procedure, because as we found out, he didn't read well. He was still stumbling over the words. When this became apparent, the words were scratched out and the man was then able to follow the procedure with virtually no problem.

So ultimately, we hope that diagnostic and repair procedures all might look something like this. A minimum of words, an uncluttered sequence and one that is easy to modify since adding or rearranging steps is obviously quite easy. All the reports from the field have been very positive. It is well accepted and seems to be working.

The ultimate critique, we believe, was received from one mechanic who did not like the new system. He said, "This makes it possible for anybody to fix it."

The third and last communications area I would like to talk about is a very broad one, the intrashop communications network. After we have effectively communicated with the owner through the service advisor

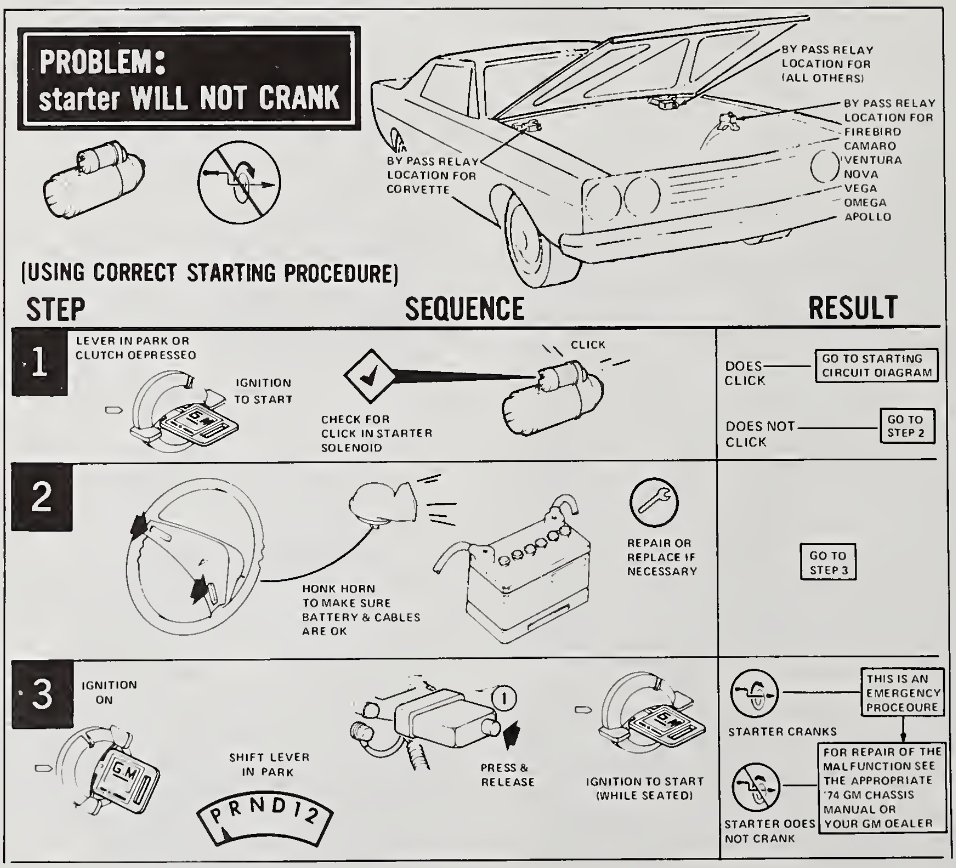


Figure 7.

and have determined why the owner is at the shop, the service challenge becomes one of expeditiously repairing the car. The vital elements of the ideal repair situation converge on the ability to fix the car the first time, on time and with minimum inconvenience to the owner. A service shop may be fortunate enough to have the most empathetic and understanding service advisors, competent mechanics, the latest in service tools and equipment, the most modern facilities; but if its shop scheduling system is inadequate, the potential exists for additional breakdowns in the customer-dealer human factors area.

Customer surveys we have conducted, specific to automobile service, indicate that the service shop's ability to return a repaired car to its owner at the promised time is of great importance. For a service shop to be able to give an owner a reliable promise time, it is required that the service dispatcher have an accurate means of determining the incremental time elements associated with the vehicle to be repaired and the available shop resource that can be assigned to perform the repairs.

Specifically, the dispatcher should be able to determine the total vehicle work content as derived from discrete time estimates of routine maintenance services, possible diagnostic time required and the estimate of total repair time. He will then compare his available shop resources such as mechanic specialty, equipment availability, floor space and parts availability to the work required and arrive at a promise time. To accomplish this task effectively requires a communication system with a high degree of information feedback.

Most dealerships employ a manual entry scheduling system that captures all in-process repair work on one large sheet of paper. The ability to keep track of all cars in the repair flow and provide current in-process status information hinges on the skill and expertise of each dispatcher. Smaller service shops, processing less than 50 repair orders per day, do not generally experience severe scheduling problems. Customer promise times are usually met.

In shops handling over 100 repair orders per day, scheduling problems resulting in missed promise times do happen more frequently. Keeping track of all those cars in the repair system is really quite a task for a human being. When these scheduling breakdowns do occur and promise times are not met, customers understandably become annoyed and angry. When this happens to the customer, another layer of adverse human factor consequences has been added to his initial experience with his vehicle failure.

Recognizing the importance of vehicle scheduling, our organization, Service Research, is now developing a system for improving dealership service shop operations. Utilizing the power and speed of the computer, the proposed system will provide assistance to the following dealership service operations personnel: service advisor, dispatcher, parts, cashier, and dealer management. I will touch on only the service advisor and dispatcher portions.

With the aid of special equipment connected to the computer, the service advisor will be more capable of analyzing vehicle problems and communicating this information to technicians. He will also be assisted in: preparing time estimates, that are realistic to meet; diagnostic repair problems, aided by the service advisor check sheet through computer logic trees and passing the information on to the technician via a technician write-up sheet; preparing repair cost estimates; providing vehicle status information, instantly and accurately; and in reducing paperwork.

For the dispatcher, the system will provide assistance through; scheduling of work loads in an efficient and orderly manner according to technician skill availability and promise times; reducing paperwork; job status information instantly without searching long lists; reduction of phone calls by providing service advisors and cashiers with accurate and instant job status.

The computer will schedule repair orders into the service shop work load flow according to the technician skills required and the promise time of vehicle completion. As technicians become available for service work, the system will assign properly sequenced jobs to the technicians. The dispatcher has the ability to override any computer decision.

What we have presented here is a service shop scheduling system representing state-of-the-art technology. A system of this magnitude would quite understandably not be required in all dealerships. Therefore, in recognition of this fact, we are also developing a manual and/or semi-manual system for dealer service shops which will offer many of the improvements of the more sophisticated computer system.

In summary, I have attempted to outline some special types of human factor consequences that can result from primary vehicle failures. I have also touched on some partial solutions that may decrease the owner aggravation index. The challenges facing those of us concerned with minimizing or preventing human factor breakdowns in the service system are great. The type of failure consequences we have dwelt on are people orientated and, as such, require people oriented solutions.

We must continue to re-evaluate our traditional approaches to solving these types of problems. In addition to the technical aspects of automotive service that include tools, equipment and the vehicle, we must focus on the service environment as it relates to human behavior. This we must do since environment shapes behavior and attendant human factor consequences. Constructive modification of the automotive service environment from a human factors frame of reference is a worthy goal.

Discussion

J. M. Karhnak, Jr., U.S. Army, Fort Belvoir: Those of us who are shade tree mechanics generally recognize the difference in quality level of things such as tools. We generally go into the life cycle cost which was mentioned in some of the other talks. We compare the price we pay for different things with the useful life. I have not noticed the same kind of practice in the automotive industry. You mentioned the stainless steel car which we probably couldn't afford, but there might be two or three different price levels of a given automobile that we might be able to buy with different life expectancies.

F. J. Uhlig: It's an interesting thought.

H. H. Wakeland, National Transportation Safety Board: I spent 5 years in the automobile industry myself and often like to ask people the embarrassing questions that I couldn't ask when I was an employee. Early in your talk you said that you had trouble defining the early performance envelope in which the vehicle would run and that you had difficulty telling the customer how to use it. When we look at the speedometer of a modern automobile which shows the figure of 120 mph on it, does that mean that it's safe to operate that vehicle at 120 mph idenfinitely, or is this an example of a piece of implied information that's been given the customer irrespective of your view that not much could be said about it?

F. J. Uhlig: I think you're right.

H. H. Wakeland: Is the area of a rated maximum speed for automobiles established by the manufacturer as part of his performance envelope in which he makes his design predictions desirable or practical?

F. J. Uhlig: We could design automobiles for a maximum speed of 55 mph. But passing maneuvers require a little more speed. So ultimately, the top speed of a vehicle has to be a little higher because of the passing ability.

N. Glassman, Naval Ship Research & Development Center: I got the impression that all the diagnosis that was occurring in this interaction at the automobile service department was the opinion of the driver and the opinion of the advisor. Isn't there a place here for some diagnostic techniques?

F. J. Uhlig: Yes, there is a place and I'll have to say quite candidly, that there isn't much really good diagnosis being employed. In the smaller independent service shops, where the mechanic can actually interface with the owner on a one-to-one basis such that the owner can describe his problem and the mechanic can go through a line of interrogation to determine the symptoms, the owner will probably get his car fixed. But as greater and greater the distance in the communication link gets between the owner and the mechanic a lot of misinformation is introduced. The service advisor check list that we mentioned was a means to sharpen up the dialogue between the customer and the service advisor. Many times a service advisor isn't even a mechanic, it could be anybody. Many of our dealerships employ ladies that don't really have much of a working knowledge of automobiles.

G. McPherson, University of Missouri, Rolla: You mentioned using a computer to schedule the flow of repair work. Is that going to be a mini-computer located on the premises or it is a time-sharing system? How large a work load is required to justify this?

F. J. Uhlig: We're now studying the communication system itself to answer that question.

H. P. Jost, K. S. Paul Products, Ltd: It might interest the audience to know that, not in the United Kingdom, but in some continental European countries, diagnostic means of checking cars are becoming more and more prominent now and, in my view, will become the thing to do in a few years time.

IMPLICATIONS OF MECHANICAL FAILURES

Session V

Chairman: R. M. Thomson
National Bureau of Standards

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Mechanical Failures: Implications for Science

John P. Hirth

Ohio State University
Department of Metallurgical Engineering
Columbus, Ohio 43210

The early scientific advances by Griffith for continuum cracks and by Orowan, Taylor, Hall and Petch, and Hirsch and co-workers for effects of microstructure have contributed to our current understanding of mechanical failure. However, this understanding remains incomplete. Here, a brief historical account and survey of the present status of various fundamental aspects of the failure problem are presented. Specific unsolved problems and areas for needed research are suggested.

Key words: Continuum cracks; continuum mechanics; failure mechanisms; mechanical failure; microstructural effects on mechanical failure.

1. Introduction

There are several approaches that one can take in considering contributions of science to failure prevention. One could argue that the most important problem is that of flaw detection which would imply that advances in instrumentation and experimental devices are points of greatest potential for science. Alternatively, one could make a case for statistics since flaw distributions, the likelihood of failure at a given flaw, and the local material properties themselves are all statistically variable. In the present article, however, we take the position that the greatest potential for science in failure prevention is in providing a better understanding of mechanisms of failure.

For convenience in discussion, the scientific advances which have contributed to the understanding of mechanical failures can be categorized roughly as arising in the disciplines of continuum mechanics, materials science and solid state physics. The first category comprises elasticity and plasticity, the second the relation of properties to microstructure at the size scale of about 1000 Å and larger and the third structure at the atomic scale. There has been considerable progress, particularly for metals, in the first two areas and some in the third. However, a number of important problems remain unsolved so that not much can be done at present in predicting the mechanical failure of materials from first principles. Moreover, there is some lack of communication among disciplines with regard to those advances which have been made.

In this article, we present in sequence, (a) a brief historical perspective, making no attempt to be comprehensive, (b) a progress report on our current state of understanding of the problem, and (c) a listing of topics of needed research. A general discussion of all types of mechanical failures would require too much space. Hence, for brevity, we primarily consider the room-temperature failure of structural parts. Also, the discussion deals mainly with metals and materials exhibiting nominal ductility, and for the quasi-static or low-cycle-stress cracking failure mode. Brief discussions of brittle materials and of high-cycle-fatigue cracking are presented at the end of the paper.

Considerations of polymeric materials and composites would require a far lengthier treatment. One should regard the following development as a typical set of problems for one portion of the general class of mechanical failures. There is some direct carry over of the following ideas. but other considerations would also be required in other cases. For example a whole new set of problems associated with viscoelastic-plastic behavior arises for the polymer case.

2. Historical Perspective

The concept of a stress concentration at a flaw in a material was illustrated by the early work of Inglis [1] [1] on the stress distribution about a cylindrical hole with elliptical cross section in a body under uniform remote loading. This idea was extended to the reversible extension of a planar brittle crack by Griffith [2]: the balance of elastic energy released with the free-energy

[1] Figures in brackets indicate the literature references at the end of this presentation.

dissipated in forming the crack surfaces for reversible crack propagation was involved in his work. Subsequent developments on energy release rates in cracking and on stress intensity factors for flaws of various shapes were made by Irwin [3] and others; compilations of stress intensity factors are given by Sih and Paris [4]. In view of these developments it is highly questionable to treat flaw problems in terms of, say, net section stress instead of stress intensities or concentrated stresses. For purposes of discussion, we mainly restrict ourselves to the mode I crack under uniform remote loading, figure 1, producing forces acting normal to the crack surface;

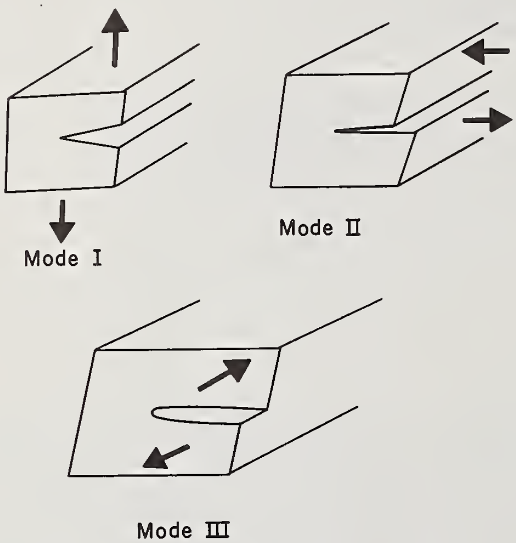


Figure 1. *Relative crack surface displacements for three modes of cracking.*

this is the usual practical failure mode of engineering materials. In this case the stress intensity factor is given by

$$K = C\ \sigma_r \sqrt{a} \qquad (1)$$

with σ_r the uniform remote stress, a the crack length, and C a geometric factor, generally near unity. In turn the near crack tip local stress field σ_{ij} has the form

$$\sigma_{ij} = f_{ij}\ (\Theta)\ K/\sqrt{r} \qquad (2)$$

with r and Θ cylindrical coordinates fixed on the crack tip.

Certain unique integral expressions involving energy conservation concepts for virtual reversible processes have been derived by Eshelby [5] and by Rice [6]. The former applied the integral to the force on an elastic singularity in an external stress field and in its most general form to his so-called energy-momentum tensor. Rice related his path-independent integral, the so-called J integral, to the energy-release rate in crack propagation. A most important development in the latter work was the result that the J integral applies not only to propagation of a brittle crack in an otherwise elastic material, but also to an elastic-plastic crack in the limit of small scale yielding as discussed in a following section.

Of course, numerous advances in the theory of plasticity have been useful in the development of elastic-plastic solutions for cracks. Particularly relevant to the present discussion, as revealed in a subsequent section, is the theory of characteristic slip fields in plastic zones, i.e., lines along which the plastic strain is purely shear. The influence of discrete microstructural features on the applicability of continuum plasticity models has been discussed by Drucker and co-workers [7].

Equivalent advances in materials science might be regarded as beginning with the finding by Rosenhain and Ewing [8] that plastic deformation of metals occurred in discrete slip steps. Subsequently, important work included studies of crystal plasticity by Taylor [9] and others, the discovery of dislocations by Orowan [10] and Taylor [11], and various developments of dislocation theory. Particularly pertinent to the crack problem was the work of Orowan [12] showing that flow stress was inversely related to the mean free-path between microstructural particles which strongly impede dislocation motion and of Hall [13] and Petch [14] in showing that flow stresses of polycrystalline materials were universely related to the inverse square root of grain diameter. Finally, we note the development of transmission electron microscopy by Hirsch, Horne, Howie, Whelan [15] and Bollmann[16] which has led to the enormous number of observations of deformed structures at scales down to tens of Angstroms.

In the solid-state area, work relevant to the cracking problem mainly involves the construction of atomic potentials. Rather than discussing this topic in detail we cite the numerous reviews in the recent book edited by Gehlen, et al [17].

3. Current Status

3.1. Contained Ductile Fracture with Planar Crack

Consider the case of cracks with planar cracked surfaces and straight fronts in nominally plane strain or plane stress fields. A recent review of the mathematical methods of analysis of such cracks is given by Rice [18], while a summary of such results together with a discussion of cracking mechanisms and of the role of microstructure is presented by McClintock [19]. The mechanism of fracture of many high-strength materials, e.g., heat-treated AISI 4340 steel, is one that we have termed contained ductile fracture and is well established by numerous metallographic results. We call this fracture mode contained because the plastically deformed region adjacent to the cracked surface is largely confined to a thickness equal to the spacing of microstructural features of the material. In this mechanism holes nucleate at inclusions or second-phase particles in the triaxial tensile stress region ahead of a crack, plastic flow occurs in the crack tip region with the region of large plastic deformation roughly limited to a size equal to the hole spacing, and crack advance proceeds by failure of the ligaments left between the locally blunted crack tip and the growing holes. This failure can occur by completely ductile necking down to a point, by a repetition of hole nucleation and growth on a finer scale and associated with finer scale particles in the microstructure, or by plastic instability as discussed in the next section.

On a macroscopic scale, the quantitative analysis of the process is fairly firm [20]. Figures 2 and 3 show plastic zones around blunt and sharp cracks in Fe–3 percent Si alloys [21]. The zones are revealed by etch-pitting which occurs below 5 percent plastic strain [21]. In both cases the shape of the zone agrees quite

FIGURE 2. *Plastic zone at a sharp crack tip in an iron—3 percent silicon alloy revealed by etch-pitting of sample loaded to* $(K_I/Y)=0.38\ \sqrt{in.}\times 200$.
Courtesy of G. T. Hahn and A. R. Rosenfield.

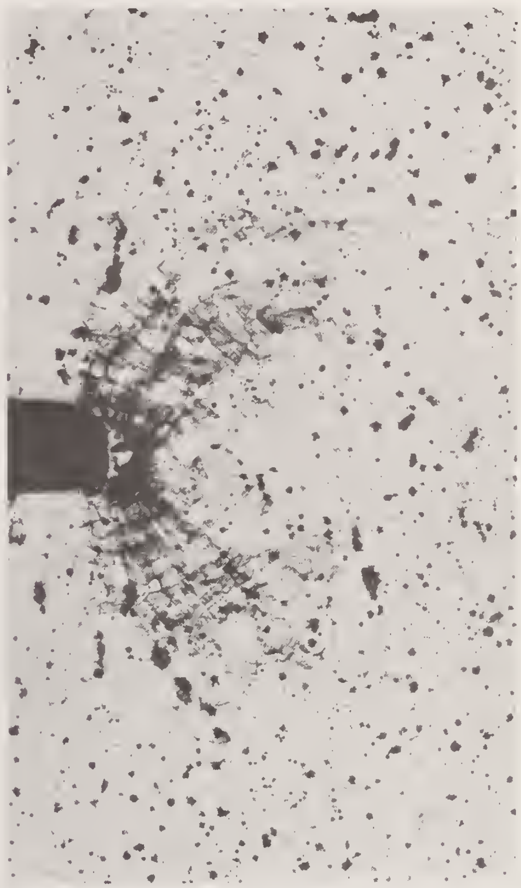


FIGURE 3. *Plastic zone at a blunted crack tip in an iron—3 percent silicon alloy revealed by etch-pitting of sample loaded to* $(K_I/Y)=0.21\ \sqrt{in.}\times 54$.
Courtesy of G. T. Hahn and A. R. Rosenfield.

well with expectation from elastic-plastic stress analysis (see Rice [20]). The details of plastic flow at the dislocation or grain size scale do not markedly affect the gross form of the plastic zone. Moreover the *J* integral approach provides a basis for calculating the energy release in crack propagation. Consider the case of a plastic zone small relative to other characteristic dimensions of the problem. In the limit the field obviously becomes essentially an elastic one except in an infinitesimal region near the crack tip, so that the *J* integral approach applies. Even with an appreciable plastic zone, the small-scale yielding approach can be used wherein the linear-elastic sharp crack field is used as a boundary condition at the elastic-plastic boundary and the plastic field, including blunting, is calculated [22]. The result agrees fairly well with complete finite element plasticity results. Provided that the proper constitutive equation is known, whether nonlinear elastic, work-hardening plastic or other, the *J* integral approach still applies. The plastic or elastic "blunting" essentially leads to the result that the long range field is that of a sharp crack equal to the true crack length plus the plastic or nonlinear elastic zone radius. The basic requirement for the solution to apply, in addition to the usual ones for a virtual reversible process, is that the form of the field fixed on the crack tip remains unchanged as the tip propagates; otherwise path independence is not retained. This is the essential reason for the breakdown of the model when the plastic zone size approaches the crack length or either approach the specimen size.[2] The same basic concept of a virtual reversible displacement would work in such a case, but the work would specifically have to be calculated for the particular cut surface which forms the new crack surface.

Although there have been the above successes in treating cracks a number of problems remain:

(1) Hole nucleation depends on either cracking of the inclusions or decohesion of their interfaces with the matrix and little can be predicted at this time about when such events occur. There are a few [23], but growing, number of observations in smooth tensile specimens of cracking and decohesion events in a locally uniaxial stress field which do indicate that typically the events occur at plastic strains roughly between 5 and 50 percent. One expects similar events in the multiaxial stress field in the plastic zone at the tip of a crack. If plastic strain is critical in the latter case, as in the former, these events will occur at some point within the plastic zone as the particle passes through it. Particularly if these events occur at the smaller of the above strains, the whole set of problems of deformation at the dislocation scale becomes important to the solution of the problem. These include anisotropic elastic effects both for matrix dislocations and for particle fields because of the different elastic constants of the latter [24], nonlinear elastic effects at interfaces and in dislocation pileups [25], anisotropic elastic and microplastic compatibility effects at grain boundaries and particle interfaces [26], interface structure including interface dislocations [27], interface cohesive strength, and local dislocation interactions including cross-slip [28]. There have been some advances in understanding of these individual deformation phenomena as discussed in the cited references, but by no means sufficient to treat the crack problem.

(2) The results discussed above suffice to rationalize correlations of crack velocity with stress intensity but do not suffice to predict a constitutive equation from first principles for crack growth in a material. The major problem again resides in an expression for hole nucleation. In some special cases, some progress has been made. For example, Rice and Johnson [22] supposed that particles decohere at a fixed strain ahead of the crack tip and that consequently both the zone of heavy plastic deformation and the crack-opening displacement should equal the particle spacing X. Then the critical stress intensity factor K_{IC} should vary as the square root of the particle spacing. These relations hold fairly well for manganese sulfide spacings in alloy steel [22]. For general particle distributions the same idea would lead to $K_{IC} \propto (XY)^{1/2}$ where Y is the yield strength. When the parameters X and Y are interrelated, the expression takes the form $K_{IC} \propto Y^n$. Such relations hold over limited ranges of Y but not at all over large ranges, as shown in figure 4. The probable reason is again the unknown variability in hole nucleation with stress, plastic strain and particle size. One factor which is established qualitatively is that particle cracking and decohesion occur at lower stresses and strains the larger the particle size [29]. This factor could account for part of the variation in figure 3.

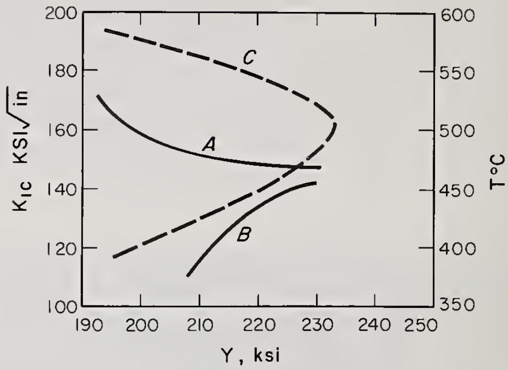


FIGURE 4. *Data for an 18Ni (250) maraging steel aged for 3 hr. at the indicated temperature and air-cooled.*
Curve C, 0.2 percent offset yield strength Y versus aging temperature, Curve A, K_{IC} versus Y, underaged, Curve B, K_{IC} versus Y, overaged.

(3) The above solutions are for planar cracks only. Branching has been treated only for special two-dimensional crack geometries [30]. Elastic-plastic solutions have not been developed for roughly branched nonplanar cracks, for planar cracks with wiggly crack tips, for cracks which meander three dimensionally or for planar cracks in the presence of random statistical distributions of growing three-dimensional holes.

(4) The above analysis leads to a prediction of a critical stress intensity and hence to a critical flaw size for a given loading, eq (1). However, the details of analysis of the nucleation of a crack at a flaw in a nominally continuous material are largely unknown for the same reasons as discussed for hole nucleation. All particles or inclusions may or may not crack, and the resultant cracks may or may not be equal in size to the particle size. In certain circumstances entire grains, whether flawed or not, can act as virtual flaws. Thus, one can only roughly associate inclusions and other particles or structural features with potential flaws. Hence, at present, one can only design the alloy structure to avoid microstructural features of certain sizes, assuming them to be potential flaws of that size.

(5) At present, the above theory does not include dynamic effects in detail. Particularly for rapidly growing cracks, problems arise in that the crack tip region is unlikely to remain isothermal because of dissipation of work in the plastic region, and damping effects may become important in the elastic region. Also, the details of crack propagation may involve localization of flow along characteristic slip lines in the plastic zone, as discussed in the next section. This localization is a rate process which perforce involves dynamic effects.

[2] While the path independent integral does not apply in such a case, a more general J integral relating the area under a load versus load-point-displacement to crack size is applicable and provides useful correlations with experiment [see various results of such J integral analysis by J. R. Rice, P. C. Paris and J. G. Merkle, STP 536, Am. Soc. Test. Mat. Philadelphia, 1973, p. 231].

3.2. Localized Ductile Failure at Shear Instabilities

Another failure mode that may or may not be associated in part with hole growth as in the preceding case is that of cracking arising within a thin lamellar region exhibiting plastic instability in pure shear. This mode has been discussed by Spretnak, et al. [31,32], and McClintock, et al. [33], who also provide experimental confirmation for the model. The shear instability along characteristic shear lines is distinct from plastic instability associated with necking. A clarification of this distinction is that in a constant-rate-of-extension tensile test of a round smooth bar, necking initiates at a load instability, and while plastic flow is macroscopically inhomogeneous at the size scale of the specimen diameter, it is microscopically homogeneous. The shear instability begins at a local maximum in true flow stress and is microscopically inhomogeneous, occurring in a region typically a μm or less in thickness along a characteristic slip direction. The local stress instability condition is

$$\delta\sigma = \left(\frac{\partial\sigma}{\partial\varepsilon}\right)_{\dot{\varepsilon},T} \delta\varepsilon + \left(\frac{\partial\sigma}{\partial\dot{\varepsilon}}\right)_{\varepsilon,T} \delta\dot{\varepsilon} + \left(\frac{\partial\sigma}{\partial T}\right)_{\dot{\varepsilon},\varepsilon} \delta T = 0 \quad (3)$$

and is evidently rate dependent. Here ε is strain, $\dot{\varepsilon}$ is strain rate and T is absolute temperature.

An example of such failure in an AISI 4340 steel is presented in figure 5. The specimen was in plane strain

FIGURE 5. *Mid-thickness crack configuration in U-notched AISI 4340 specimen loaded close to maximum load* [31]. Steel tempered to 178 ksi ultimate tensile strength. × 90.

so that the characteristic slip lines are orthogonal as indicated. Obviously, the path of the crack follows the characteristic slip lines, in one case actually switching from one set of characteristics to the other and back. In the case of figure 5, the microscopic propagation of the crack occurred by hole growth with holes initiated at carbide particles. In other cases, including some instances of single crystals of pure metals, the crack surface gives no indication of microscopic hole growth.

Phenomenologically, it has been established that such failures along lines of instability require that, (a) the rate of work-hardening be small (consistent with eq (3)), (b) that a free surface be present since initiation occurs there, and, (c) that a large plastic strain gradient exist.[3] These conditions were present for the case of figure 5. Pertinent to the previous section, they would be applicable in the ligament region between growing holes so that ligament failure could occur by shear instability.

Unsolved problems in this case include:

(1) What is the microscopic failure mechanism in cases where hole growth does not occur?

(2) No elastic-plastic solution is available for the ragged crack case such as that in figure 5.

(3) When hole growth accompanies failure along an instability line, does formation of the instability precede hole nucleation or vice-versa?

(4) What is the mechanism of initiation at a free surface?

3.3. Stress Corrosion Cracking

In this section, we discuss corrosion cracking in alloys such as austenitic stainless steel and alpha-brass which are in either an annealed or cold worked condition but which are at low strength levels relative to the elastic modulus compared to the high strength alloys such as high strength steel and precipitation-hardened alloys. The features of such stress-corrosion cracking (SCC) in comparison to the cases of the previous sections are given in figure 6 [34]. In higher strength ranges, SCC occurs at lower stress intensities than does ordinary cracking. Also, in SCC a regime of slow crack growth is exhibited at low stress intensities, stage I in figure 6b. Crack initiation occurs at surfaces and the useful life prior to failure is dominated by crack growth so that nucleation is not as important as in the previous cases.

In some cases stage I smoothly merges into the stage III rapid crack growth which follows the same v–K curve as ordinary cracking, figure 6b. In many cases a plateau stage II is also found. As expected, stage III usually corresponds to ductile cracking by hole growth as discussed above and with the crack proceeding transgranularly. Stage I corresponds to either intergranular or transgranular cracking with a much finer scale of features associated with plastic deformation at the crack tip, as would be expected at the lower stress intensity, and with a smoother fracture surface containing fine features resembling slip steps. Stage III corresponds to crack branching (meandering) or to rumpling of the crack surface (quasi-cleavage). Both effects would tend to reduce the local stress intensity and hence would

[3] Indeed it appears that crack nucleation is dependent on the attainment of a critical strain in this case in contrast to brittle cleavage crack nucleation within a plastic zone which follows a critical normal stress criterion [e.g., see R. O. Ritchie, J. F. Knott and J. R. Rice, J. Mech. Phys. Solids 21, 395 (1973)].

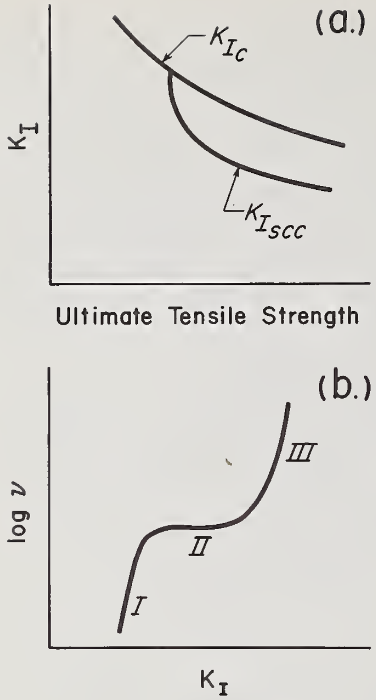


FIGURE 6. *(a) Typical plots of* K_{Ic} and K_{Iscc} *versus U.T.S. for high strength alloys.*
(b) Typical plot of crack growth rate v *vervus* K_I *in the SCC case.*

logically reduce the slope of the crack velocity versus average stress intensity plot.

While still somewhat controversial, the balance of the prevailing view is that SCC proceeds by a cyclic process of local plastic flow producing slip steps with surfaces freshly exposed to the solution, dissolution at these surfaces causing crack advance and increase of local stress intensities, causing the cycle to repeat. Other models mostly also involve the concept of a mechanical-plastic flow portion and a chemical dissolution portion of the cracking cycle.

The *J* integral, energy-release methods again apply to this case, but some of the unsolved problems are even more difficult than in the previous case:

(1) No elastic-plastic solution exists for the meandering crack which occurs to a marked degree for systems with stage II cracking.

(2) The problem in constructing from first principles an equation to describe crack growth rates for stage I cracking is more severe than for ductile cracking, even with known surface boundary conditions. The plastic zone size is smaller and hole growth has not been shown to be a factor in cracking (though it could occur on a near atomic scale). The plastic zone size is of the order of the spacing of dislocation sources, of dislocation cells, or of grains. Hence a description of the plastic flow involves all of the structural features discussed previously: anisotropic elasticity, compatibility, cross-slip, etc. One can balance the energy release with energy to create new crack surfaces and that dissipated in plastic flow, but since the latter is difficult to calculate, a constitutive relation cannot be constructed. Moreover, in most cases there appears to be a purely corrosion part of the SCC cycle producing a locally changing stress field at the crack tip, involving another rather poorly defined variable in the cracking process.

(3) Superposed on the previous problem is that of establishing surface boundary conditions. Local crack chemistry, diffusion in solution, rates of adsorption, degree of polarization or film formation, and electrostatic effects all affect the surface boundary conditions. These effects appear in the mechanical portion of the cycle as an effective surface energy appearing both in the work dissipation and as an influence on the emergence of dislocations at the surface and through local image stresses acting on near-surface dislocations. In the chemical portion of the cycle, they directly control the local dissolution rate.

(4) In the cases where SCC is intergranular, the question remains unsolved of why it is so. Possibilities for the effect include: grain boundary grooving which changes the local solution chemistry and influences the chemical step, solute segregation to the grain boundaries which influences either the chemical step or the mechanical step via local solid solution hardening, elastic and plastic compatibility effects which generate differential mechanical behavior at the boundaries.

3.4. Failure Following the Temper Embrittlement of Steels

A case with some features of each of the preceding cases is that of temper embrittlement of heat treated alloy steels. Cracking occurs at approximately a critical stress intensity, crack velocities are relatively high, and the nucleation stage is largely undefined as for the contained ductile fracture case of section 3.1. However, cracking in the embrittled condition occurs at prior austenite grain boundaries or at ferrite boundaries [35] and evinces no dimples or other features characteristic of a large plastic zone, in this sense resembling the SCC case.

The unsolved problems are mainly those mentioned above, but some new features merit emphasis:

(1) There is definite evidence that solute segregation to boundaries is an important factor with evidence for both transient [35] and quasi-equilibrium [36] adsorption effects. Mechanistically, the solutes could either cause a highly localized, contained elastic-plastic cracking, wherein the critical step is local plastic decohesion as in the previous cases, or could cause the critical step to become atomic bond breaking at the crack tip, with or without some minimal plastic deformation ahead of the tip.

(2) If atomic bond breaking is important, a number of possible unresolved mechanisms exist at the atom-dislocation scale. Does the breaking occur nearly reversibly with groups of bonds gradually parting or does it occur irreversibly bond by bond? If the latter occurs, then the mechanism is likely to resemble the nucleation and growth of double-kinks over the Peierls

barrier in low temperature dislocation glide. The phenomenology of such an approach has received some attention [37]. However, the elastic field of a kinked crack has not been determined and the form of the potential barrier is not known.

(3) Atomic interaction potentials are required for solute-solvent groupings in highly nonlinear elastic configurations to even begin to predict potential barriers from first principles. This problems is severe even for a crack in a single crystal of a pure metal. The progress that has been made in the physics of solutions has been for somewhat idealized. homogeneously strained solutions. The development of a potential for highly strained configurations is an outstanding challenge in solid state physics.

(4) For the case of cracking at boundaries, even with known atomic interactions it is important to know the boundary structure at an atomic scale, including the dislocation structure at the boundaries.

(5) With any irreversible portion in the cracking process, the problem in development of a first-principles relation to describe cracking is exacerbated. The energy released for a given crack advance is still known, but of the corresponding dissipation terms the irreversible portions usually are not. For example, if a crack advances by partly irreversible bond breaking with no plastic deformation, the thermodynamically reversible part of the energy released would appear as surface energy of the cracked surfaces while the irreversible part would be dissipated as elastic pulses eventually attenuated and dissipated by heat flow. Similarly with an elastic-plastic crack, part of the energy could be dissipated as elastic pulses in addition to that used to form the crack surfaces and the plastic zone. Then, independent of one's knowledge of the details of the plastic region, one still could not develop the needed equation without information on the irreversible release mechanisms. Acoustic emissions which have been observed in a number of cracking cases provide direct evidence for such irreversible terms [38].

3.5. Hydrogen Embrittlement

Hydrogen embrittlement includes both cases where alloys are exposed to hydrogen gas and cases of corrosion cracking where hydrogen is introduced as a consequence of hydrogen evolution being the cathodic reaction in the local electrochemical cell. Hydrogen embrittlement of alloys is manifested in two ways. At high hydrogen fugacities, the hydrogen forms molecules in voids and generates internal pressure which assists in the plastic growth of voids and subsequent cracking by the contained ductile fracture mode [39]. Alternatively, for high-strength alloys such as high strength steels, hydrogen at low fugacities (corresponding to pressures well below one atmosphere) leads to failure at stress intensities less than the critical value for a hydrogen-free alloy [40]. A three-stage v–K curve as in figure 6b is often exhibited. Scanning electron microscopy indicates that, as for SCC, stage I is intergranular and stage III is contained ductile rupture [41]. Hydrogen entry definitely has a role in the cracking process since oxygen inhibits hydrogen cracking [40], presumably by adsorption, blocking sites from hydrogen adsorption, while catalysts for hydrogen adsorption, such as hydrogen sulfide, enhance cracking [42].

Perhaps it will evolve that vapor embrittlement will be more appropriate terms for this type of failure. The work of Kerns and Staehle [42] has shown that slow crack growth can occur in a variety of gases containing hydrogen as a constituent and one, chlorine, that does not. Their work indicates that for AISI 4335 tempered to an ultimate tensile strength of 284 ksi, exposure to gases at 100 to 300 torr led to K_{ISCC} values decreasing from 80 to 40 ksi $\sqrt{in}$, respectively, for CH_4, C_2H_2, H_2, NH_3, HCl, HBr and Cl_2.

Again a number of problems are unanswered:

(1) For the intergranular mode, the problems of grain boundary solute adsorption, grain boundary structure, and compatibility effects exist.

(2) The mechanism is not known as in the previous case. Again the critical step in crack advance could be either the onset of decohesion in a contained plastic region ahead of the crack tip or bond breaking at the tip. Because interstitial hydrogen is highly mobile at room temperature there is an additional uncertainty in where hydrogen influences the cracking process. Plausible models have been advanced for hydrogen diffusing ahead of the crack into the contained plastic zone and aiding in decohesion [40]; for hydrogen adsorbing as a tightly bound surface layer hindering plastic relaxation [43], and thus decreasing the K value required for crack propagation, for electrostatic interactions of protons and electrons in fields arising from elastic-electrostatic field crossterms [44], for hydrogen aiding in bond-breaking (verified in an approximate atomistic computer simulation model [45]), for hydrogen influencing decohesion in the near vicinity of the crack tip [46], and for hydrogen influencing dislocation motion by forming atmospheres in the elastic fields of the dislocations [47]. Which model applies remains unsettled, although observations [42] of slow crack growth for the same steel in dry hydrogen or in dry chlorine tend to support the surface effect models, since chlorine solubility and diffusivity are much less than the equivalent values for hydrogen.

We do note that there are some inconsistencies arising from lack of communication between disciplines in the above models. However, some aspects of dislocation theory have not been considered in models involving solute atmospheres [48]; and, as pointed out by Rice [20], the region of maximum normal stress for the plane strain case has incorrectly been taken as equal to the tensile yield stress Y and to occur at the boundary of the plastic zone rather than the continuum mechanics result of a value $3Y$ at a distance roughly one-tenth of the extent of the plastic zone (in plane stress the maximum normal stress would equal Y).

(3) The need for interatomic potentials discussed in the previous section is also critical in the present case.

3.6. Brittle Fracture

Fracture with no plastic deformation, even on a microscopic scale is defined as brittle fracture. There is some question whether such fracture can ever occur. However, in principle, the problem of brittle fracture exists at three levels. First, there is the Griffith's crack in an otherwise perfect crystal. As discussed in earlier sections for cases where decohesion at the crack tip is controlling, problems exist in degree of reversibility of cracking (equilibrium surface energy as the dissipation step or that plus elastic wave release), planar versus double-kink type advance with lack of attendant three dimensional elastic field solution, and atomic potentials. A further problem is whether an atomically sharp crack can be retained in crack propagation without local microplastic relaxation by dislocation formation. A model three-dimensional calculation for deciding upon the relative likelihoods of these processes has been presented by Rice and Thomson [49]. Further development requires consideration of the local three-dimensional double-kink problem for both the crack and the generated dislocation.

A second case is that of cleavage fracture initiated after some plastic flow at a preexisting stress concentration [50,51]. There is evidence that in steels the requisite is for a critical stress to be achieved at a characteristic microstructural distance of about two grain diameters ahead of a preexisting crack. This critical stress is in turn related to a Griffith criterion with the flaw size equal to carbide particle diameters or inclusion diameters. In other instances initiation has been associated with dislocation intersections [52] or dislocation-twin boundary intersections [53]. The problems in both initiation and propagation parallel those of section 3.1, e.g., particle decohesion and cracking or non-planar cracks, and hence are not repeated here.

In connection with cleavage, the work of Hoagland, et. al. [54], is also of interest. They have initiated cleavage cracks, propagated them in a decreasing stress intensity field, and arrested the cracks at a critical dynamic stress intensity. The understanding of the arrest problem in a microstructural mechanistic sense is beset with the same problems as above together with various dynamic problems. The latter include the elastic-field of a moving nonplanar crack, dissipative heating in the crack tip region, and local strain rate dependence of the flow stress in the tip region.

At another level there is the problem of apparently brittle fracture in materials such as silicon nitride and silicon carbide now being considered for some structural applications. (Not all ionic or covalent crystals, or course, fail brittlely. Sodium chloride is ductile in compression and zirconium hydride fails in tension by localized plastic shear flow, for two examples.) These silicon compounds have quite complicated microstructures, which give rise to a set of problems in understanding:

(1) The fracture problem is primarily one of nucleation since growth is essentially instantaneous. Thus even with an expectation phenomenologically of cracking at a critical stress intensity—critical flaw size there is a most formidable problem in the statistics of extreme values in statistical distributions [55].

(2) There is a current need for characterization of microstructures to make possible even the assessment of typical flaw sizes for these materials.

(3) The calculation of local stress distribution is a difficult problem requiring consideration of anisotropic elasticity and of compatibility effects.

3.7. Fatigue

Space does not permit treatment of other fields of mechanical failure such as creep, failure under shock and impact loading, hot corrosion cracking and fatigue, all of which would require treatment equal to the above to reproduce even the above type of cursory survey. However, there are some direct analogies [56] in the particular case of high cycle fatigue which merit brief discussion.

Similar to SCC, fatigue cracking is almost invariably surface nucleated and proceeds by a period of slow crack growth. Surface stresses, surface finish, and surface chemistry thus influence crack nucleation. When failure does occur, however, most of the time to failure (roughly ninety percent or more) consists of the crack growth stage. During cyclic loading, work-hardening occurs in a few thousand cycles after which a nominal steady state is achieved in which cyclic plastic deformation occurs in persistent slip bands. Extrusions form at surface intersections by the slip bands with corresponding formation of internal porosity in the bands. Cracks nucleate at the interface between the slip bands and the matrix and grow. The internal dislocation structure in the bands comprises cell walls of multipole arrays which are relatively unstable under stress compared to the relaxed interwoven arrays characteristic of monotonically loaded specimens. Environment affects crack growth rates in analogy with the SCC case.

The local crack propagation event in some sense resembles the contained ductile rupture case [57]. This parallelism suffices to rationalize the phenomenological correlation between crack growth rates and stress intensity factors [57]. The analysis of this process in terms of continuum elasticity-plasticity is complicated by the occurrence of deformation in a highly localized slip band but, on the other hand, over distances (of the order of millimeters) large compared to the band width. Contrariwise, the scale of the deformation tends to rule out mechanisms based on dislocation events at the atomistic scale [58]. With these provisos, many of the problems (excepting those related to brittle fracture) discussed in connection with a monotonically loaded crack apply to the fatigue case.

4. Needed Research

The various problems in the context of the foregoing discussion suggest areas of needed research, which are enumerated here with little elaboration. Some similar listings have been presented recently for the cases of corrosion cracking and fatigue [19,20,48]. We note particularly that the following list includes those areas of potential interest which do not now seem to be receiving sufficient support or attention and does not include important research topics now being supported. The latter, which of course deserve continued support, include studies of solution and electrode kinetic effects in corrosion cracking, correlations of microstructure with crack velocity-stress intensity factor studies, dynamic effects for elastic cracks [59], and so forth. The listing follows:

1. Theory
   a. Continuum Mechanics
      (1) Elastic and elastic-plastic solution for a plane crack with a straight front in the presence of a random distribution of spherical holes.
      (2) Elastic and elastic-plastic solution for a crack meandering in three dimensions.
      (3) Elastic and elastic-plastic solution for a plane crack with a wiggly crack front or for a branched crack.
      (4) Inclusion of dynamic effects for intermittently irreversible cracks and for rapidly accelerating cracks, including both dynamic K–v and energy dissipation-v relations.
   b. Elasticity-Plasticity
      (1) Dislocation scale models of decohesion of a matrix-particle interface.
      (2) Elastic-microplastic compatibility effects at inclusions and second-phase particles including anisotropic elasticity.
      (3) Nonlinear anisotropic elastic treatments of dislocation pileups and interactions.
      (4) Elastic field of a plane crack with a double-kink on an atomic scale in an otherwise straight crack tip.
      (5) Elastic-microplastic treatment of a crack tip near a grain boundary with compatibility effects and a dislocation model of plastic relaxation.
      (6) Tensor treatment of dislocation motion in multiaxial stress states characteristic of crack tip elastic fields; interrelation with stress-invariant plasticity models.
      (7) Influence of surface films and electrochemistry on dislocation—crack interaction.
      (8) Development of a criterion for initiation of failure by shear instability and of a mechanism for cracking when hole growth does not occur.
   c. Atomic Structure
      (1) Models of grain boundary (interface) structures.
      (2) Models of grain boundary (interface) dislocations.
      (3) Estimations of interface cohesive strengths.
      (4) Estimations of the Peierls energy of an advancing crack.
   d. Interatomic Potentials
      (1) Potentials for pure metal atoms in highly nonlinear elastic strain fields characteristic of crack tips including electrostatic crossterms arising from electron redistribution and including free surface relaxations.
      (2) The same for solute-solvent interactions also including screening effects at solute sites.
   e. Statistics
      (1) Theories of statistics of extreme values for flaw distribution and flaw formation.
   f. Computer Simulation

Some comment is added here since some suggest that simulation is not worth while until better atomic potentials are available. Granted that simulations at present are not likely to exactly replicate real materials, the work is most valuable in studies such as those listed below where even order-of-magnitude results would provide better guidelines than presently available. In other words, actual mechanisms are tested, albeit in artificial materials.

      (1) Model of kinked crack as in item D–1–b–4.
      (2) Model of dislocation relaxation at sharp crack tip and test of theory [48].
      (3) Study of the role of solute adsorption, or adsorption at a sharp crack tip.
      (4) Study of the degree of irreversibility in propagation of a sharp crack.
      (5) Development of lattice Green's functions for a cracked body to permit the fixing of boundary conditions on an atomic region.

2. Experiment
   a. Deformation Studies.
      (1) As suggested by Rice [20], studies of hydrogen embrittlement for cracks with rounded tips of various radii to give regions of maximum normal stress at various distances ahead of the crack tip and hence to resolve the controversy about where the hydrogen exerts its influence.
      (2) Studies of interface decohesion of particles in the vicinity of similarly rounded notches to assess the process of decohesion in the presence of multiaxial stress states.
   b. Atomistic Kinetics
      (1) Study of adsorption and absorption in the presence of hghly nonlinear elastic fields by performing such measurements on whiskers and platelets which can sustain roughly five percent elastic strain.
      (2) Study of hydrogen diffusion in the presence of multiaxial stresses to determine the tensor nature of the stress dependence of the diffusion activation energy pertinent to crack-tip elastic fields.
   c. Microscopy
      (1) Characterization of solute adsorption-absorption at interfaces, grain boundaries, dislocations, and crack tips (Auger microscopy, atom probe).
      (2) Study of grain boundary structure (atom probe, electron microscopy of molecular crystals).
      (3) In situ studies by high-voltage-transmission-electron microscopy of SCC, hydrogen embrittlement cracks. (Studies [60] of ductile rupture of films from 0.1 to 100 μm show that some degree of triaxiality-hole growth is achieved in the center of the films so that the studies provide some semblance of plane strain conditions.)
      (4) Characterize the structure of complex materials (mostly ceramics) subject to brittle fracture.

The author is happy to acknowledge helpful discussions of the present topic over the past five years with colleagues in the ARPA Materials Research Council, in particular F. A. McClintock and J. R. Rice, whose work is drawn upon in the present article. The author is grateful also to R. G. Hoagland, J. E. Sinclair, J. W. Spretnak and R. W. Staehle for helpful comments, to G. T. Hahn, A. R. Rosenfield and J. W. Spretnak for supplying micrographs, and to the U.S. Office of Naval Research for support of this work.

5. References

[1] Inglis, C. E., Trans. Inst. Naval Arch. **55**, 219 (1913).

[2] Griffith, A. A., Phil. Trans. Roy. Soc. (London) **A221**, 163 (1920).

[3] Irwin, G. R., J. Appl. Mech. **24**, 361 (1957).

[4] Sih, G. C., Methods of Analysis and Solution of Crack Problems, Noordhoff, Leyden (1973); P. C. Paris and G. C. Sih, Fracture toughness testing and its applications, ASTM STP 381, ASTM, Philadelphia, 1965, p. 84.

[5] Eshelby, J. D., Phil. Trans. Roy. Soc. (London) **A244**, 87 (1951).

[6] Rice, J. R., J. Appl. Mech. **35**, 379 (1968).

[7] Drucker, D. C., J. Materials **1**, 873 (1966); T. T. Wu and D. C. Drucker, J. Appl. Mech. **34**, 195 (1967); T. W. Butler and D. C. Drucker, J. Appl. Mech. **40**, 650 (1973).

[8] Ewing, A., and Rosenhain, W., Phil. Trans. Roy. Soc. (London) **A193**, 353 (1899).

[9] Taylor, G. I., J. Inst. Met. **62**, 307 (1938).

[10] Orowan, E., Z. Phys. **89**, 605 (1934).

[11] Taylor, G. I., Proc. Roy. Soc. (London) **A145**, 362 (1934).

[12] Orowan, E., in Symposium on Internal Stress in Metals and Alloys, Inst. of Metals, London, 1948, p. 451.

[13] Hall, E. O., Proc. Phys. Soc. (London) **B64**, 747 (1951).

[14] Petch, N. J., J. Iron Steel Inst. **174**, 25 (1953).

[15] Hirsch, P. B., Horne, R. W., and Whelan, M. J., Phil. Mag. **1**, 677 (1956); P. B. Hirsch, A. Howie and M. J. Whelan, Phil. Trans. Roy. Soc. (London) **A252**, 499 (1960).

[16] Bollmann, W., Phys. Rev. **103**, 1588 (1956).

[17] Gehlen, P. C., Beeler, Jr., J. R., and Jaffee, R. I., Interatomic Potentials and Simulation of Lattice Defects (Plenum, New York, 1972).

[18] Rice, J. R., in Fracture, vol. **II** (Academic Press, New York, 1968), p. 191.

[19] McClintock, F. A., in Proc. Int. Conf. on Stress Corrosion Cracking and Hydrogen Embrittlement of Iron Base Alloys, Nat. Assoc. Corros. Eng., to be published, 1974.

[20] Rice, J. R., ibid.

[21] Hahn, G. T., and Rosenfield, A. R., Int. J. Fract. Mech. 4, 79 (1968); in Inelastic Behavior of Solids, Edited by M. F. Kanninen, W. F. Adler, A. R. Rosenfield and R. I. Jaffee (McGraw-Hill, New York, 1970), p. 673.

[22] Rice, J. R., and Johnson, M. A., ibid., p. 641.

[23] Several recent observations together with references to earlier work are presented in Proc. of the Third Int. Conf. on Strength of Metals and Alloys, Inst. of Metals, London, to be published, 1974.

[24] Eshelby, J. D., Solid State Phys. **3**, 79 (1956).

[25] Kuan, H., Ph.D. Thesis, Ohio State University, Columbus, 1973.

[26] Hook, R. E., and Hirth, J. P., Acta Met. **15**, 535, 1099 (1967).

[27] Bollman, W., Crystal Defects and Crystalline Interfaces (Springer-Verlag, Berlin, 1970).

[28] Hirsch, P. B., and Humphreys, F. J., in Physics of Strength and Plasticity, Ed. A. S. Argon (M.I.T. Press, 1969), p. 189.

[29] Klevebring, B. I., and Mahrs, R., Scand. J. Met. **2**, 310 (1973).

[30] Andersson, H., J. Mech. Phys. Solids **17**, 405 (1969); erratum **18**, 437 (1970).

[31] Griffis, C. A., and Spretnak, J. W., Trans. Iron Steel Inst. Japan **9**, 372 (1968).

[32] Chakrabarti, A. K., and Spretnak, J. W., Instability of plastic flow in directions of pure shear, submitted to Met. Trans., 1974.

[33] McClintock, F. A., Kaplan, S. M., and Berg, C. A., Int. J. Fracture Mech., **2**, 615 (1966).

[34] General features of stress corrosion cracking are discussed in Proc. Int. Conf. on Stress Corrosion Cracking and Hydrogen Embrittlement of Iron Base Alloys, Nat. Assoc. Corr. Eng., to he published, 1974.

[35] Ohtani, H., Feng, H. C., and McMahon Jr., C. J., Met. Trans., **5**, 516 (1974).

[36] Recent discussion including earlier literature citations is given in M. Guttmann, P. R. Krahe, F. Abner, G. Amsel, M. Bruneaux and C. Cohen, Met. Trans. **5**, 167 (1974); C. L. Smith and J. R. Low, Jr., ibid., p. 279.

[37] Hseih, C., and Thomson, R. M., J. Appl. Phys. **44**, 2051 (1973); see also ref. 46.

[38] Tatro, C. A., Liptai, R. G., and Moon, D. W., in Proc. Int. Conf. on Stress Corrosion Cracking and Hydrogen Embrittlement of Iron Base Alloys, Nat. Assoc. Corros. Eng., to be published, 1974.

[39] Smialowski, M., ibid.

[40] Hancock, G. G., and Johnson, H. H., Trans. Met. Soc. AIME, **236**, 513 (1966).

[41] Nelson, H. G., and Williams, D. P., in Proc. Int. Conf. on Stress Corrosion Cracking and Hydrogen Embrittlement of Iron Base Alloys, Nat. Assoc. Corros. Eng., to he published, 1974.

[42] Kerns, G. E., Wang, M. T., and Staehle, R. W., ibid.; G. E. Kerns and R. W. Staehle, Scripta Met. **6**, 1189 (1972).

[43] Gilman, J. J., ibid.

[44] Tiller, W. A., ibid.

[45] Markworth, A. J., Kanninen, M. F., and Gehlen, P. C., ibid.

[46] Oriani, R. A., ibid.

[47] Bockris, J. O'M., ibid.

[48] Hirth, J. P., ibid.

[49] Rice, J. R., and Thomson, R. M., Phil. Mag. **29**, 73 (1974).

[50] Wilshaw, T. R., Rau, C. A., and Tetelman, A. S., Int. J. Fracture Mech. **4**, 147 (1968).

[51] Ritchie, R. O., Knott, J. F., and Rice, J. R., J. Mech. Phys. Solids **21**, 395 (1973).

[52] Stokes, R. J., Johnston, T. L., and Li, C. H., Phil. Mag. **4**, 920 (1959); **6**, 9 (1961).

[53] Armstrong, R. W., in Deformation Twinning, eds. R. E. Reed-Hill, J. P. Hirth and H. C. Rogers (Gordon and Breach, New York, 1964), p. 356.

[54] Hoagland, G. R., Rosenfield, A. R., and Hahn, G. T., Met. Trans. **3**, 123 (1972).

[55] Coble, R., McClintock, F. A., and Thomson, R. M., in Proceedings of the 1973 Summer Conference of the ARPA Materials Research Council, Univ. of Michigan, Ann Arbor, 1974.

[56] Argon, A. S., in Corrosion Fatigue, Nat. Assoc. Corrosion Eng., Houston, 1972, p. 176.

[57] Paris, P. C., in Fatigue—an Interdisciplinary Approach, Edited by J. J. Burke, N. L. Reed, and V. Weiss (Syracuse Univ. Press, New York, 1964), p. 107.

[58] Avery, D. H., and Backofen, W. A., in Fracture of Solids, Edited by D. C. Drucker and J. J. Gilman (Interscience, New York, 1963), p. 339.

[59] Freund, L. B., Int. J. Engineering Sci. **12**, 179 (1974); see also ref. 18.

[60] Bauer, R. W., Lyles, Jr., R. L., and Wilsdorf, H. G. F., Z., Metallk, **63**, 525 (1972); R. W. Bauer and H. G. F. Wilsdorf, Proc. Int. Conf. on Dynamic Crack Propagation, Noordhoff, in press (1974).

6. Discussion

O. Jahari, IIT Research Institute: You showed a curve of volume fraction of inclusions versus the strain dependence. In such calculations, don't you have to take into account the size of the inclusions also?

J. P. Hirth: Yes, that was a correlation where the particles were all nominally the same size. It is true, failure is more likely to occur with larger particles. However, that correlation assumed that all samples failed independent of particle size. That was only a rough correlation with experiment. There's not enough detail to give a correlation with spacing and the size of the particles.

O. Jahari: But in general, except for certain types of alloys such as TD nickel, the inclusion size would vary a great deal.

J. P. Hirth: Yes.

O. Jahari: What happens in materials that do not have inclusions, like high purity zone refined materials?

J. P. Hirth: It's really not clear what goes on. For single crystals, of course, there are some cases where necking right down to a knife edge takes place; that is, no internal voids ever open up. However, recent electron microscopy studies by Wilsdorf of silver single crystals and large grained stainless steel indicate that internal voids open during crack propagation. Wilsdorf's specimens were not "thin" films in the sense of being thousands of angstroms in thickness, but they are not thick either. So there's a question of whether plane strain or plane stress plays a role here, but at least there is one case where a high purity metal fails by internal hole growth.

O. Jahari: We looked at high purity zone refined iron and found dimples, but at the bottom of the dimples, instead of finding particles we found what looked like holes. Unfortunately the resolution of the scanning electron microscope is not good enough to get further details.

K. G. Kreider, National Bureau of Standards: Would you comment on George Ansell's treatment of the role of inclusions in the fracture of ductile metals?

J. P. Hirth: There is some controversy here. Ansell's work is really most applicable to dispersion hardening systems and more related to work hardening rates. His idea is that Orowan type pile-ups around particles cause them to crack. Other people say that that cannot take place and instead there is dislocation generation and motion around the particles that cause a hole to nucleate near the particle. Experimentally, when people look around particles, they tend to see only one or two dislocation loops which would tend to discount the Ansell idea. On the other hand, the resolution of the instrument is such that there could be a number of loops right next to the particle that are not resolved. So it's really still an open question.

K. G. Kreider: Would you express an opinion?

J. P. Hirth: My own bias is that the dislocation relaxation takes place. But the evidence is still out.

B. W. Christ, National Bureau of Standards: In the Hahn-Rosenfield work on silicon iron, how do you get a kidney bean shaped plastic zone at the crack tip without lots of etch pits in the region where the plastic zone must have been when the tip got to the point shown in the slide?

J. P. Hirth: They got the sharp crack by propagating a fatigue crack to that point, which would have characteristically a much smaller plastic zone, and then they pulled it in tension.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Implications for Action for Engineering

H. W. Paxton

Carnegie-Mellon University
Office of the Director for Research
Pittsburgh, Pennsylvania 15213

The demands on materials for fail-safe operation have increased enormously since the days of "Galloping Gertie" and Liberty Ships. The necessity to use higher strengths has increased demands on selection, design, acceptance testing, nondestructive evaluation of structures, and manufacturing methods. Extremes of temperatures, thicker sections, and environmental effects have all accentuated the tendency for mechanical failure. This paper will discuss the current status of selected problems, and by assessing our fundamental understanding of certain important engineering parameters, will examine prospects for improved materials and techniques.

Key words: Environmental extremes; fail-safe operation; high strength materials; mechanical failure; potential service conditions; proliferation of specifications.

Many engineering structures are not "materials-intensive" and selection of material is made on economic grounds, all too frequently on a first-cost basis with little though to subsequent maintenance costs. Mechanical properties are important to a point, but since over-design is common, it is extremely rare to find failures from gross plastic deformation following overloads. Occasionally, massive flaws are present, for example, large internal cavities in castings, but this problem is operational rather than conceptual and will not be pursued further.

As requirements become more stringent, and the strength-to-weight ratio increases, there are two other causes of failure resulting from thinner sections which I will not discuss. One is buckling, usually plastic, and the other is general corrosion resulting in significant cross-sectional loss.

In the last two decades, the demand for materials to perform at higher strengths, greater thicknesses, and in extremes of environment such as high and low temperatures, various kinds of radiation, aggressive chemistries, etc. has grown enormously. It is in these areas where I would like to place most of my emphasis.

Fortunately, only a small fraction of failures occur in service, but in many cases these tend to be very expensive in terms of immediate damage (the "Silver Bridge"), lost or delayed production (e.g., several nuclear plant failures), expensive design precautions (e.g., much space hardware) and personal hazards or loss of life.

Let me now outline what I will cover. The points I would like to make really can be summarized as follows:

(1) How well do we know the service conditions?
(2) How well do we know the total properties of the material?
(3) What can be done in specifying materials and processes to make the final structure more predictable, and, in particular, less sensitive to unexpected conditions in service?

I will give some examples to illustrate these points. I will also apologize just this once for belaboring the obvious, but somehow the obvious is not always so.

There are many engineering structures which are sufficiently standardized that historical precedent can be utilized effectively and even an acceptable statistical chance of failure may have been established. It is the new and extrapolative designs which are of most interest here. Let me illustrate what I mean by an interesting example with (so far) a happy ending where prediction of service conditions was very difficult but also crucial. It involved the general concept of protection of a system at a critical point from the results of failure at another critical point. The Peach Bottom reactor of the Philadelphia Electric Company is a prototype of a very interesting competitor to the light water reactors. It uses carefully purified helium to extract heat from a graphite moderated core. This helium then is used as the heat source in a relatively conventional steam generator. The hot helium and the cooler return are in concentric pipes, and of course separating insulation is necessary. Provided the helium can be kept "pure," the problem is not particularly interesting because several layers of thin metallic sheet can be used to provide stagnant gas spaces in series. However, if we have any leakage from the steam generator thereby transferring water vapor to the helium, we have a potential problem. This water vapor reacts with the hot graphite in the core to form a mixture of CH_4 and CO which can be strongly carburizing to the FE–Ni–Cr alloys which typically would be candidates for the insulation. Carburization can produce an extremely brittle material which could have all kinds of unwanted effects. The probability of this sequence is small, but how small? As you might expect, we cannot solve this completely by analytical means, but it was possible to combine analysis with a few critical empirical data and produce a solution which has worked.

Another example of imperfectly known service conditions is from the bridge world. The "Silver Bridge" across the Ohio River failed catastrophically because load and the local industrial atmosphere caused slow crack propagation to a critical size, and this likelihood was just not forseen. A second general principle violated here was the lack of redundancy in the structure, a problem of increasing significance with materials demonstrating only fair resistance to crack propagation.

There are a number of other service conditions which sound trivial but are often tragic. The best design in the world is of no use if the concept in the brain of the designer is not in fact the one which actually goes into service. The problem is compounded if the designer is under the impression that it really did.

Two examples illustrate this point. Metallic implants to help repair skeletal damage are commonplace and essential. I have a number of thoughts on inadequate design and material selection which are not appropriate here. But I was shocked to find that a 316 L bone plate was being installed with vitallium screws. Ah, you say! But surgeons don't know any metallurgy and have perhaps never heard of electrolytic effects between dissimilar metals. I am sure this is truer than we would like, but in this case the surgeon was not to blame. The vitallium screws were in a package clearly labeled as stainless steel.

Another example concerns the failure of a large pressure vessel used in the manufacture of ammonia. The design was perfectly adequate but because of various time constraints an inspection for cracks in or near the internal surface of the welds was not made after stress relief. No cracks were found prior to this process and since reinspection was awkward and expensive, the judgment was made not to do it, even visually. Unfortunately, a crack did develop and under the influence of operating stress and a substantial partial pressure of hydrogen this propagated completely through the weld. The resulting shutdown and loss of production was of course very expensive. The designer thought he had designed a safe and proper system and indeed, in the absence of a crack and high pressure hydrogen simultaneously, in all probability he had.

The more general question which this audience may wish to debate is the cost-effectiveness of inspection. How much is enough? We have seen in recent years a new attitude on such matters as assigning economic values to environmental controls or safety of nuclear reactors. As scientists and engineers, we know that data on which such decisions are made lack the precision we would wish, but nevertheless action is taken and has legal backing. I think we have at the moment an inadequate basis for designers to assess the probable consequences of the differences between their brainchild and the structure actually in service. This is compounded as we work with untried designs, higher strength materials, and interactive effects (e.g., cracks and environment). An important factor in ameliorating this problem is to stimulate recognition of the importance of understanding the totality of a material to be specified—its "pedigree" if you will, since this word connotes not only ancestors from melting stock (sometimes of a standard questionable by the equivalent of the Kennel Club) but also off-spring as it in turn becomes re-melted and reused in the recycling which seems unavoidable in the near future. This is my next topic.

In their more pejorative moments, metallurgists are prone to refer to designers as "handbook engineers." There is some truth in this. Designers have a primary responsibility to design and this involves specifying a material. Certain properties are necessary and are in handbooks. In a great majority of cases the design works and everyone is satisfied. It is the relatively rare situation where selection of one property in ignorance of others causes problems and ultimately expensive failures.

I am reminded of a classic case where a very important material characteristic for a particular design was the thermal expansion coefficient. Consultation of physical property tables and price lists suggested that Type 430 (a 17% Cr stainless steel) was ideal. It would have been except that in service around 500 °C, as was planned, this material suffers a drastic loss of ductility and would rather soon have lost mechanical integrity. This fact is not in handbooks of physical properties, although it is reasonably well known in metallurgical circles. Many such difficulties can be avoided if everyone concerned asks the proper questions of the right people—the standard communications problem. Information transfer happens a lot of the time but when it doesn't, failures are more likely. And here designers get their own bark at the materials specialist! They might have a life cycle involving combinations of creep, fatigue and thermal cycling. So, they ask, what do I use? What stress or strain limits do I have? Are the welds likely to be better or worse than the base material? And so on. The truly knowledgeable metallurgist will be faster on his feet than Mohammed Ali in avoiding direct answers to such questions.

I have painted perhaps too extreme a picture, but, I maintain, one with substantial elements of truth. Let me discuss now some factors which may eventually contribute to a higher probability of using materials successfully.

A casual glance at a list of specifications, be they SAE, ASTM or whatever, really boggles the imagination. To an informed metallurgist, who has an idea of the function of the various elements listed in the chemical composition, there is a certain pattern. But we must forgive an uninitiated person for wondering if we are really serious about the need for all the compositions listed. We have perhaps heard of the old shop supervisor to whom there were only three kinds of steel, hot-rolled, cold-rolled and stainless. This is slightly apocryphal but I feel fairly certain that if these three were expanded to ten or twelve compositions, the great majority of our needs could be covered by more emphasis on careful processing.

A study of how these specifications arose is perhaps helpful. The chain of logic—composition → structure → properties—is something we teach to all our students. However, the degree of sophistication and understanding along this chain varies enormously. There are many situations where even on the research frontiers, a semi-quantitative appreciation is lacking.

Looking back, we see that variations in chemical composition have been a principal consideration in developing new alloys. This is understandable and in many cases logical. If we wish to have a stronger alloy, we know various ways to produce this—solid solution hardening, precipitation, cold work, and so on. Certain alloying elements can be used and it is a matter of relatively simple experimentation to find out how much. We thus get into the habit of creating a "structure" by compositional variations with the implicit assumption that the processing variables—melting, casting, working, quenching, ageing, etc.—will not be very different from those we have used in the past.

We will end up with a stronger alloy. Now we ask about the other properties—the more subtle ones like fracture toughness, stress corrosion crack rate, ability to resist cyclic loading, and so on. Here our predictive ability is much poorer and all too often we find empirically (and expensively) that our stronger alloy has marginal or unacceptable other properties. The next step is to try some other variation in chemical composition to produce our strength, and again measure our "subtle" properties. Eventually, we find the least harmful combination of increased strength and minimally impaired other properties.

I chose strength as the primary variable of interest in an alloy. Of course, others could be of maximum interest, but they do not change my argument. Many of the compositional specifications were designed to optimize one property. We then scramble as well as we can to avoid possible shortcomings in other properties.

We are obviously a long way from having the knowledge to design an alloy with a particular combination of properties. However, experience over the last decade or so in controlling processing variables more closely for a specific composition has proved to be a major step forward. We may note here particularly the pioneering work of Irvine and Pickering [1] [1] in simultaneously improving the yield strength and fracture toughness of low-carbon low-alloy steels, a development rapidly picked up by all the major steel producing countries. Over-simplifying slightly, this was done by recognizing the vital positive contribution of ferrite grain size and the negative contribution of pearlite to toughness. Chemistry still has effects but these are quantifiable separately.

A second example of processing control developed in the last few years is the ability to influence inclusion shape in steels, especially sulphides. Traditionally, the difference in toughness in rolled products, for example, between "longitudinal" and "transverse" directions is quite significant, and is convincingly explained by the presence of elongated MnS inclusions. Fatigue properties are also markedly different. There are other sulphides which have a higher melting point than MnS and thereby are stronger at rolling temperatures, remaining substantially spherical in the finished product. However, while the idea of using elements other than manganese to combine with sulfur is not new, earlier results were erratic, in part because oxygen and sulfur compete for these scavenging elements. When the proper techniques of addition were perfected [2], rare-earth silicides proved most effective and produce substantially equivalent longitudinal and transverse impact properties throughout a whole ingot. In higher strength steels, zirconium, (now reasonably economically available as Zircalloy scrap) seems to hold promise of achieving the same result.

In steels at the highest strengths conveniently used today, say in the range of 180 000 to 250 000 psi yield strength, one is struck by the large variations in toughness, crack growth rate and K_{ISCC} at a given yield strength, even for the same nominal composition What this means of course is that our understanding of the fine details of structure which control these properties is inadequate.

Some progress is being made in this area but it is, perhaps understandably, slow. Fast fracture of such steels occurs typically by the nucleation, growth and linking of voids nucleated at second phase particles. Hirth [3], in the preceding paper, has outlined some of the present theoretical approaches. Experimentally, detailed studies to provide quantitative understanding are fairly sparse. We note here an interesting unpublished study by Cox and Low [4] to compare behavior in a carbide-strengthened steel (quenched and tempered AISI 4340 at 200 ksi) with that of an 18 Ni maraging steel at the same strength level. In both cases, the first voids form at relatively massive inclusions (at the MnS—matrix interface in 4340, and by fracture of the titanium carbonitride inclusions in the maraging steel). These voids grow in the maraging steel until impingement occurs leading to coalescence and final fracture. However, an instability sets in earlier in 4340 in that void sheets nucleate at the cementite particles responsible for strengthening and lead to premature failure. Apparently the strengthening particles in the maraging steel do not nucleate these secondary voids, and thus we obtain the typical extra toughness of maraging steels. Since our knowledge of nucleation of voids in solids is still rudimentary, it is not surprising perhaps that we can still get large variations in toughness in apparently equivalent steels. It will require the most careful experimentation to document the subtle contributing factors.

In stress corrosion cracking the microstructural fea-

[1] Figures in brackets indicate the literature references at the end of this presentation.

tures are also very difficult to relate to such parameters as K_{ISCC} or crack growth rate. Some plausible reasons for this difficulty have been outlined by Hirth [3]. A variety of experimental studies in steels where a serious effort was made to vary one microstructural feature at a time was unable to uncover major effects [5,6]. Since these studies were on a single heat, another study was made to examine possible effects of variations in trace elements from heat to heat [7]. Again the effects were slight.

As we consider fatigue, one can find a similar story. Quoting from a recent NMAB report [8]:

> "Contributing to the complexities of using a fracture-mechanics model for fatigue-life prediction are various facts including:
>
> (1) Life estimates are subject to error since observed crack-propagation rates (da/dN) at specific values of the stress-range intensity factor (ΔK) show as much scatter as fatigue lives at specific values of the stress amplitude in conventional fatigue tests.
>
> (2) On the same metal, a scatter range of at least 1:3 in observed crack-propagation rates must be anticipated, while crack-propagation rates in the same class of metal may vary still more.
>
> "A similar uncertainty applies to the determination of the critical crack size. The rate of slow crack propagation is not a simple power function of the (constant) stress-range intensity factor (ΔK), as is assumed usually, but deviates significantly from such a function at both ends of the ΔK range."

The examples cited above, while representative rather than comprehensive, illustrate the point that even on "well-known" materials the properties of interest in failures are incompletely understood. The strength is fairly well understood and is reproducible within satisfactory practical limits, the role of undesirable but unavoidable second phases is understood in a qualitative fashion, e.g., the effects of different melting and solidification technologies on toughness, but there are other important properties, (da/dN,K_{ISCC}, etc.) which cannot yet be clearly identified with microstructural and compositional features. With some trepidation, I would propose two items for discussion:

(1) We try to arrive at a very limited number of materials (two or three at most) in each class with sufficient strength to be susceptible to occasional catastropic failure. A concerted effort should be made to characterize these materials so that we understand much better than now:

- (a) what factors are responsible for the variance in fracture properties
- (b) what processing techniques are necessary to manufacture a truly consistent product
- (c) what joining technique(s) will be effective
- (d) what characteristic defects can be found by present (and future?) inspection techniques.

The designer then will ultimately be able to prescribe with greater confidence and rigor a material to do his job. The first cost of a carefully processed material may well be somewhat increased, but I am confident that there will be many examples where these costs can be saved in maintenance, inspection and general "peace-of-mind."

(2) This item may not go over well but I hope it will be discussed vigorously and even bitterly rather than ignored. Very simply, I propose that we have far too many specifications which cause confusion and unfortunate choices. Can we devise a situation where any specification has a prescribed finite life and will automatically be delisted unless a sponsor or sponsors can be found for it? This sponsor would have certain rights and responsibilities which need to be worked out carefully, but more importantly, the lack of a sponsor would allow a marginal specification to die.

Let me summarize the points I have tried to make.

(1) In untried and sophisticated designs, an imaginative assessment of potential service conditions, including interactions, must be made by the designer at an early stage in conjunction with someone who understands the various possibilities.

(2) There are various pitfalls, some technical and some institutional, which can cause good conceptual designs to result in disaster-prone products.

(3) A serious effort must be made to reduce the variety of materials available which overlap each other in many properties. I recommend a concerted approach to provide a few well-characterized materials which are *reproducible* in all of their *significant* properties.

(4) We should rely on careful processing rather than changes in chemistry much more than we are accustomed to doing.

Finally, as a sting in the tail, let me add something which for reasons of time (and perhaps cowardice), I did not discuss in detail in the text. More and more, we will be held accountable for our actions and designs. We are, or at least insist we are, professionals. Therefore, we must know our levels of confidence, we must understand where our expertise fits in with that of others and we must not be embarrassed to admit ignorance when appropriate and to ask for advice.

Meetings like this are fine, if we keep up the cooperative spirit when we go home.

References

[1] Pickering, F. B., The effect of composition and microstructure on ductility and toughness, in Toward Improved Ductility and Toughness (Climax Molybdenum Development Company of Japan) 1971, p. 9.

[2] Wilson, W. G., and Wells, R. G., Metal Progress, (Dec. 1973), p. 75.

[3] Hirth, J. P. This volume, pp. 179.

[4] Cox, T. B., Thesis, Carnegie-Mellon University, 1973 (to be published).

[5] Stavros, A. J., and Paxton, H. W., Metallurgical Transactions 1 (1970) p. 3049.

[6] Proctor, R. P. M., and Paxton, H. W., Trans. ASM **62** (1969) p. 989.

[7] Proctor, R. P. M., and Paxton, H. W., Corrosion Science **11** (1971) p. 723.

[8] Application of Fracture Prevention Principles to Aircraft. NMAB–302 (Feb. 1973), p. 41.

Discussion

R. S. Fein, Texaco, Inc.: The last two talks were concerned primarily with the implications for work that needs to be done in developing new information. Yesterday we heard Dr. Jost talk about the transfer of technology that's already on the shelf and the tremendous financial implication that the use of this information has. It was estimated that about one percent of the gross national product in terms of energy or raw materials could be saved. I think the real challenge for engineers, particularly, is how to put this information in such a form that the practicing engineer can use it everyday and use it effectively. This is something that needs considerable effort on the part of knowledgeable people such as research engineers and research metallurgists.

H. W. Paxton: I couldn't agree with you more.

R. Thomson, Federal Energy Office: It seems to me that this may be a very laudible approach to try to decrease the total number of specifications and the total number of classes of materials. How do you think we'll get there? It seems to me that the natural tendency, with the pressure on to develop new materials, is to go just exactly in the opposite direction. Also, there is the tendency for every materials producer to get his own special material out. What kind of practical approach do you think we should take here?

H. W. Paxton: May I remind you, the title of this symposium is the "Definition of the Problem". I agree with you it's going to be tough, but evangelists must get out and sell the bill of goods.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Mechanical Reliability—Implications for Engineering, Manufacturing and Design

W. Dale Compton

Director, Ford Scientific Research Laboratory
Ford Motor Company
Dearborn, Michigan 48021

The automobile is a complex assembly of some 15 000 different components in a network of functional subsystems and systems. The definition of the problem of mechanical performance is complicated by the absence of a fixed vehicle performance envelope. Control over use of the vehicle is in the hands of the individual driver. This is substantially different from military and aerospace hardware for which there are specified performance envelopes, and from many consumer products in which the performance follows a fixed pattern, almost independent of consumer preference. This situation imposes special implications for engineering and design, and for testing.

Improved methodologies for predicting component durability and lifetime under real-life conditions, when combined with information on material and structure properties derived under realistic test conditions, can be of great value in developing designs with improved mechanical properties. Likewise, the high volume production rates associated with automotive components involve consideration of material processing and manufacturing factors if the proper levels of mechanical performance are to be achieved.

Key words: Design engineering; material deterioration; mechanical failure; mechanical integrity; mechanical reliability; performance envelope; predicting product reliability.

Although life was much simpler then, few of us would be willing to return to the era where communication was limited to the newspaper and the wireless, where the principal labor was devoted to the production of food, and where the upper level of education seldom exceeded the twelfth grade. The demand of the late-twentieth-century consumer has been for ever-increasing comforts and new products. And this demand has been met, but at the expense of increasing complexity. But since complexity brings with it the increased probability for failure, it is little wonder that the consumer and the supplier are continually caught up in the conflict between the demand for inexpensive products and the demand for products with high reliability. Zero initial defects, infinite life, and low initial cost are objectives that are very difficult to achieve in the same product. But, the buying public is really rather realistic about what it expects. While it doesn't demand infinite life at zero cost, it does want a product with increased mechanical integrity. Surveys of what bothers automobile buyers clearly indicate that they want a vehicle with greater reliability, durability and increased handling ease—all within the constraints of low cost and high operating efficiency. Althugh they tend to consider squeaks and rattles as a greater measure of mechanical integrity than is probably justified, they are clearly concerned by mechanical failures that impair the operation of the machine. It is essential that we continue to develop improved ways of providing an improved product.

It seems clear that the manufacturer that can reduce troublesome failures even in the face of the demand for increased complexity will have an advantage in the marketplace. How, then, do we proceed to try to satisfy this need, remembering always that the final decisions as to what is cost effective for the product will be settled in the marketplace?

In the automotive business there are five key elements to the process of assimilating those 15 000 or so discrete parts into the mechanical, electrical, chemical and physical system that we call an automobile. These elements are:

Design
Manufacture
Assembly
Customer Usage
Service

Since Mr. Uhlig has already discussed the matter of service in some detail, I will not dwell upon it here except to reiterate the point that service must be considered on a par with design, manufacture, and assembly in the total context of producing a failure-free system. It is important, though, to understand that attempting to anticipate how the customer will use (or abuse) the vehicle and how it will be serviced and maintained in the field has a significant impact upon the design, manufacture, and assembly of the total system.

My purpose here is to try to give you some feeling for the complexity of the problem of reliability in the automotive business, some insight into how we try to attack these issues, and some view of how we look at the future. First, for the complexity.

The design objectives in this area are reasonably straightforward: achieve the optimal cost design that will meet the objectives for the part while providing the required reliability. Of course the performance objectives of the part must be generated from the performance objectives of the subsystems and these, in turn, from the performance envelope for the total vehicle. Thus the tendency would be to start with a detailed performance envelope for the vehicle—an envelope that can be taken as representative of customer usage. But one immediately recognizes that the loads that are applied which lead to the forcible deterioration of a vehicle will be of varying amplitudes, will be dependent upon the type of usage that each customer imposes, and will vary with the environmental conditions. Thus, the performance envelope that the vehicle must satisfy is very broad. As will be seen shortly, this has a profound impact upon how the design process is carried out.

Clearly, design "prove-out" is a major task that involves extensive laboratory and field testing of components, subsystems, and vehicles under a variety of conditions. This is costly and time-consuming and often indicates that a design change is needed, with another "prove-out" required. Thus, if we can provide the designer with better techniques, we can look forward to more reliable products at lower design costs. But no matter how complete the design process becomes, it will always be necessary to verify experimentally that the functional performance and the durability levels are acceptable by utilizing tests that produce results that correspond as closely as possible to those arising from customer usage.

While it is not appropriate here to catalog all of the efforts that are under way to develop better information and techniques for the dsigner, I will mention two that are being developed and used. The first is a response to the question, what is the accumulated effect of usage in the field? In order to obtain a better measure of accumulated loads, a method is being developed that appears to have broad applicability in determining the structural loads that selected components will experience in service [1].[1] A number of *S/N* Fatigue Life Gages with various mechanical multipliers are bonded to a component. A permanent change in gage resistance is observed which depends upon the magnitude and the number of cycles of load that the part experiences. The relationships are schematically shown in figure 1. Under a proper set of conditions, these resistance changes can be used to reconstruct a good approximation to the

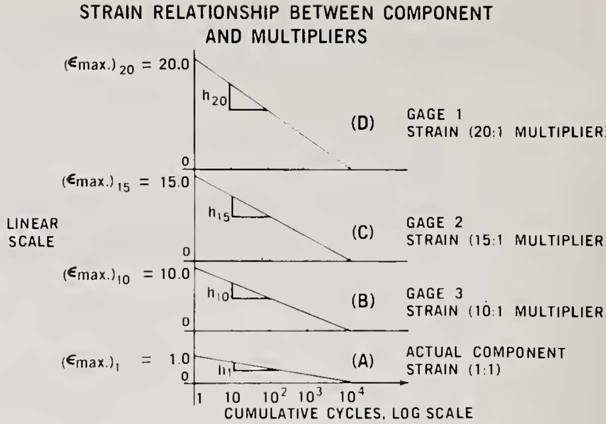


FIGURE 1. *Relationships between the strain experienced by a component and responses of strain gages having mechanical multipliers of 1:1. 10:1, 15:1 and 20:1.*
(Taken from reference [1], figure 7.)

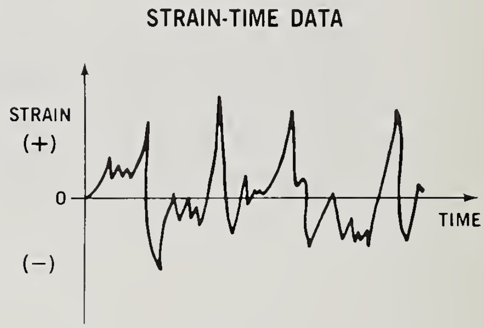


FIGURE 2. *A typical strain-time response for a component.*

load-range history that the part has experienced. For example, the strain-time sequence shown in figure 2 can be reduced to the form shown in figure 3, which is a "cumulative distribution" of the strain versus the "cumulative number of cycles" that produced that strain. Figure 3 also shows the laboratory verification of the method through a comparison of the actual load histories with the load histories estimated by this rather simple technique. The good agreement is obvious. The correlation between the estimated and recorded load histories then was made for a suspension component after 50 and 100 circuits on the test track. The results are shown in figure 4. The estimated load history derived from the S/N gage readings compares quite favorably with the actual strain history that was recorded during one circuit on the track. While this technique appears to hold considerable promise for developing a realistic measure of the loads that structural components have experienced in service, there will be a very long period of data taking before the actual performance envelope for many components can be reasonably established.

The second method is a response to the question, what are the localized strains that are present in various com-

[1] Figures in brackets indicate literature references at the end of this presentation.

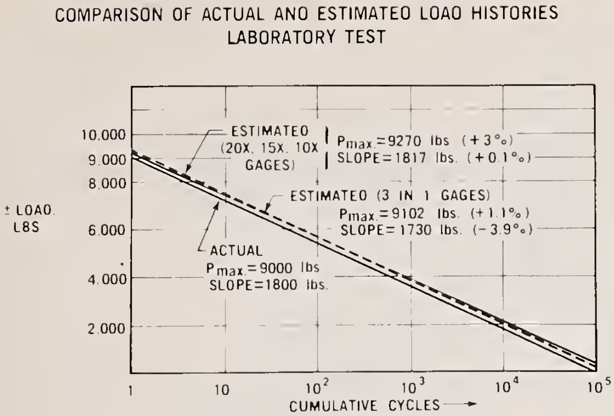


Figure 3. *Comparison of actual and estimated load histories for a laboratory test sample.*
Cumulative load is plotted as a function of the cumulative cycles that produced the strain. (Taken from reference [1], figure 16.)

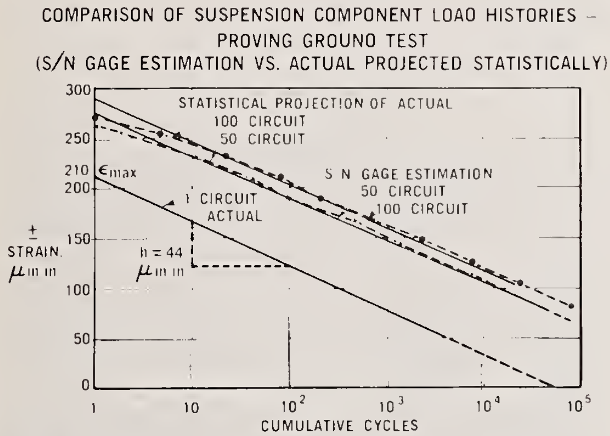


Figure 4. *Comparison of suspension component load histories as estimated by S/N gage and as projected statistically. Data accumulated from proving ground test.*
(Taken from reference [1], figure 19.)

ponents that are subjected to "typical" usage? This problem is approached, as shown in figure 5, by attaching strain gages to the component of interest and recording the output of the gages when the vehicle that incorporates this part is operated over a "typical" test cycle. The surface conditions of the test track have been determined through experience to subject the vehicle to a reasonably good approximation of the varied stresses that a vehicle will experience in normal service. The recorded strain dependence, as is shown in figure 6 for three different components, is then translated into a control algorithm that allows a computer to simulate the track conditions in the laboratory and subject a sample to a strain history that will duplicate that seen by the part in service [2].

Figure 5. *Component shown with attached strain gage.*

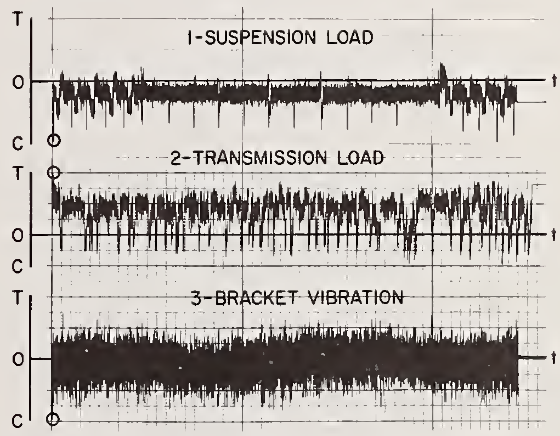


Figure 6. *Strain dependence as a function of time for three components as measured during a vehicle test.*

An expanded portion of the strain history of the central trace, as taken from the component on the test track, is reproduced in figure 7. This is the strain sequence that is duplicated in the laboratory. The relationship between the stress and strain is far from linear, as can be seen by a comparison of the strain sequence with the stresses needed to impose the observed strains on the test sample shown at the bottom of figure 7. The determination of such complex stress-strain responses provides an essential input into subsequent cumulative

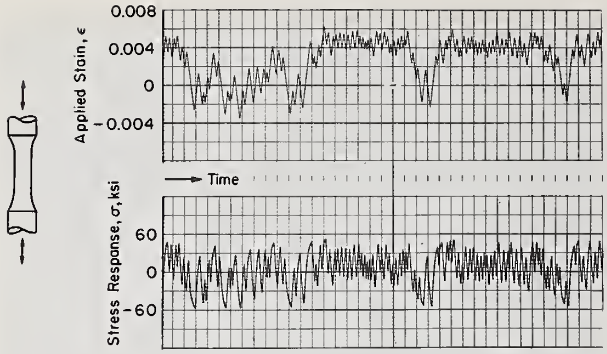


FIGURE 7. *Determination of stress response of specimen shown at left during complex strain sequence.*
(Taken from reference [2], figure 2.)

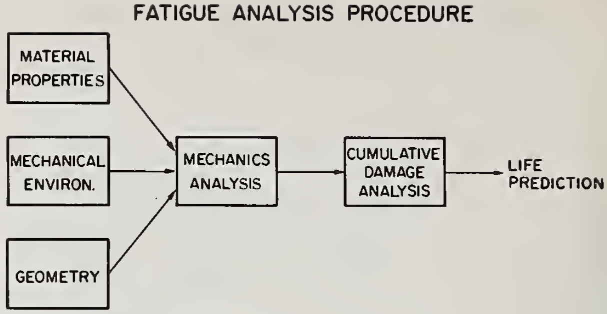


FIGURE 8. *Essential aspects of fatigue analysis procedure.*
(Taken from reference [2], figure 4.)

damage analyses and can be accomplished either experimentally, as shown, or analytically if the appropriate *cyclic* stress-strain properties of the material are known. Clearly, one is interested in the most critical elements of a component in such an analysis, so this technique is most often applied to such areas as those near holes, notches or welds. Not only is the strain spectrum as measured on the track an approximation of the true effects that will be seen in the field, it is further assumed that a succession of runs on this track segment will subject the part to an accumulated effect that is representative of long in-service effects and, therefore, can be used to accumulate damage typical of several years of vehicle usage while in the hands of the consumer. There is justification for this assumption, since experience has shown a good correlation with performance under average customer usage. It is clear, however, that the width of the performance envelope that is generated by the wide variety of usages is quite large. Thus, the accumulated load measurement technique mentioned earlier will give a better measure of the *average* loads experienced under customer usage rather than the distribution in the loads that arise from the wide variety of usages. Even so, these data will then make it possible to more effectively utilize the measured local strains in evaluating component design.

But the development of an improved performance envelope is only the first step, for the designer must then translate a desired performance into a design. At this point, all of the material properties that are of importance to the performance of the part must be available to him, or he must be able to estimate the important properties of the material from the limited data that may be available.

The design procedure must now proceed along the paths shown in figure 8 where a fatigue analysis, leading to a life prediction, is developed from the material properties, the mechanical environment, and the geometry of the part. The performance envelopes mentioned above, albeit very imprecise, must be introduced as a part of the mechanical environment to which the part will be subjected.

Let us look at the problem from another direction for a moment and consider some of the multitude of ways that a component can fail mechanically. These include fatigue, after a few cycles and after millions of cycles, fracture, both brittle and ductile, creep, gross yielding, and buckling. The loading mode, type of stress, and operating temperature strongly influence the type of material that can be used in a given application. Further, it must be expected that the material will deteriorate in use. It wears and galls; it oxidizes and corrodes; and it changes its metallurgical structure through diffusion and grain growth. With all the information available on material properties, it seems incongruous to assert that the absence of suitable material properties presents one of the serious deficiencies in developing an adequate life prediction for the part. But, unfortunately, this is true. The most easily determined properties are often not the most useful. This situation can be illustrated by considering fatigue analysis in a bit more detail.

Those properties of most direct bearing to a fatigue analysis are the cyclic stress-strain response of the material. These are much more significant for life prediction than are the more commonly available monotonic relations. Also needed for the damage analysis are material fatigue-life curves. Examples of each of these types of data curves are shown in figure 9 for a low carbon steel that is used prominently in the vehicle. For the design engineer, however, these curves must be supplemented with the sort of detailed specifications listed in figure 10. We have collected all of this type of data that we can find into a Materials Handbook for use by the design engineer. By way of emphasizing the limited extent of the information available, the present handbook contains data on only 41 steels, 7 cast irons, and 8 aluminum alloys.

As the design engineer considers the requirements for efficient energy management in a structure, requirements that arise from the need to protect the vehicle and occupant against damage and injury during a collision, it is necessary to know whether the various properties of materials depend upon the strain rate. The dependence of the yield and the ultimate tensile strength as a function of strain rate is shown in figure

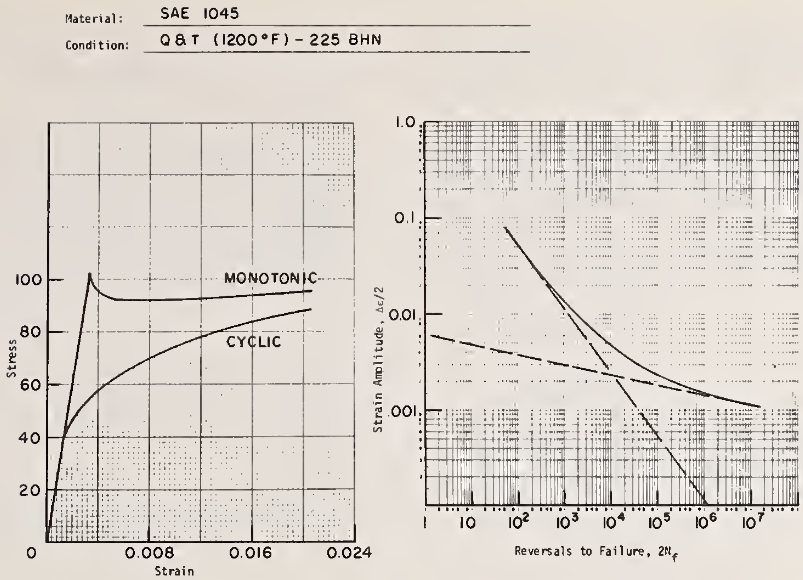


FIGURE 9. *Stress-strain characteristics and strain amplitude as a function of reversals to failure for SAE 1045 steel.*
(Taken from reference [3].)

MATERIAL CHARACTERIZATION SHEET — Sheet 1 of 2

Material: SAE 1045 — Specimen Ref. No.: IX
Condition: Q & T (1200°F) - 225 BHN — Test Cond.: R.T.

Monotonic Properties:		Cyclic Properties:	
Mod. of Elast., E	29.5 x 10^3 ksi	Yield Strength, 0.2% S_y'	60 ksi
Yield Strength, 0.2% S_y	92 ksi	Strain Hard. Exp., n'	0.18
Ultimate Strength, S_u	105 ksi	Strength Coeff., K'	195 ksi
Red. in Area, % RA	65	Fatigue Strength Coeff., σ_f'	178 ksi
True Frac. Strength, σ_f	178 ksi	Fatigue Ductility Coeff., ϵ_f'	1.0
True Frac. Ductility, ϵ_f	1.04	Fatigue Strength Exp., b	-0.095
Strain Hard. Exp., n	0.13	Fatigue Ductility Exp., c	-0.66
Strength Coeff., K	166 ksi	Transition Fat. Life, $2N_t$	10,000 rev
True Toughness, U_p	165,600 in-lb/in^3		

Comments:

CHEMISTRY

C	Mn	P	S	Si	Cu	Ni	Cr	Mo
.47	.77	.007	.035	.23	.08	.02	.02	.01

HEAT TREATMENT

AUSTENITIZED : 1500°F/ 30 MIN., BRINE QUENCH, TEMPER : 1200°F/ 1 HR.

Microstructure: TEMPERED MARTENSITE

Source: R. W. Landgraf (7) Date: 1/70

FIGURE 10. *Properties of SAE 1045 steel.*
(Taken from reference [3].)

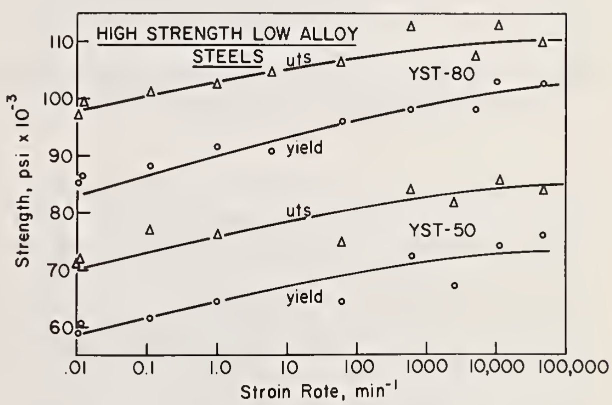


FIGURE 11. *Yield and ultimate tensile strength as a function of strain rate for two high strength, low alloy steels—YST 50 and YST 80.*
(Taken from reference [4]), figure 5.)

11 for two high-strength low-alloy steels. A strain rate of 5×10^4 in/min is equivalent to a 50 mph strain rate. By contrast to the behavior shown for the HSLA steels, the aluminum alloys tested to date show essentially no strain rate effects. Data are available for only a very few materials in the high strain rate regime.

When we look beyond the usual metal alloy systems to those materials that are finding increased usage (for example, plastics and elastomers), we find even less data on materials of direct interest, and less fundamental understanding of properties that can be used to make reasonable extrapolation of material properties. Phenomenologically, many plastics behave similarly to metals [5]. For example, the general response of ductile polymers to cyclic strain is a cyclic softening. There are, of course, distinct effects of the microstructure on the details of the softening mechanisms; for example, single phase structures, such as polycarbonate, tend to exhibit an incubation period before softening, whereas multiphase microstructures, such as nylon, show immediate softening with no incubation period. But for the highly heterogeneous systems that we use so frequently, for example the rubber-modified plastics and the glass-reinforced plastics, an overall cyclic softening occurs which continues throughout the life of the material. Figures 12 and 13 present the comparison of the monotonic and cyclic stress-strain relations for glass-reinforced nylon and ABS, respectively. Although these curves have the same general form as are observed for most metals, the magnitudes of the effects differ by an order of magnitude or more. The maximum stresses in plastics are substantially less than for metals, whereas the corresponding strains are much larger than occur for metals. The temperature effects are quite important for both plastics and metals.

What about fatigue failure of the plastics? Figure 14 gives a comparison of the fatigue life curves for pure nylon and for nylon loaded with 13 percent and 33 percent glass fiber. The form of the upper curve is characteristic of the fatigue of those polymers that do not craze. The shape of the lower two curves is characteristic of polymers that craze. In the heterogeneous polymer systems, the added material, such as the glass fibers, produces an effect on the mechanical behavior that is equivalent to crazing in a homopolymer. It appears that any perturbation that disrupts the bulk flow of the polymer and promotes a localized accommodation to deformation can be expected to introduce an abrupt transition in the fatigue response at short lives. There is no question but that the fundamental understanding of the response of heteropolymer systems to mechanical stresses is inadequate and deserves much more attention.

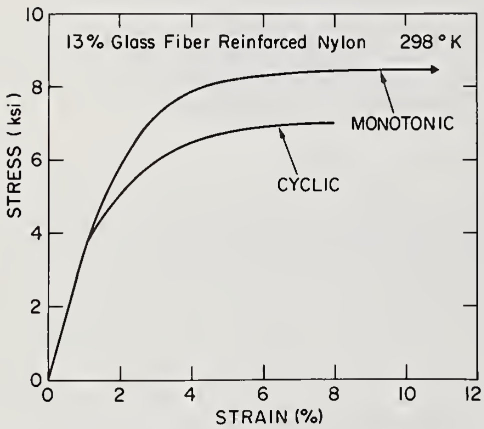


FIGURE 12. *Cyclic and monotonic stress-strain curves for ABS at 13 percent glass fiber reinforced nylon at 298 °K.*
(Taken from reference [5], figure 22.)

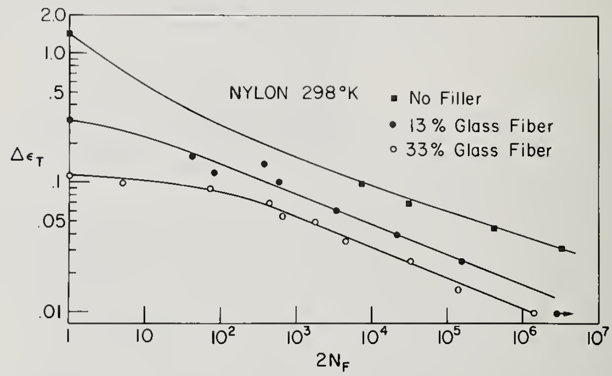


FIGURE 14. *Fatigue life data showing the effects of reinforcing glass fibers (chopped) on the fatigue life of nylon at 298 °K.*
(Lower two curves taken from reference [5], figure 40.)

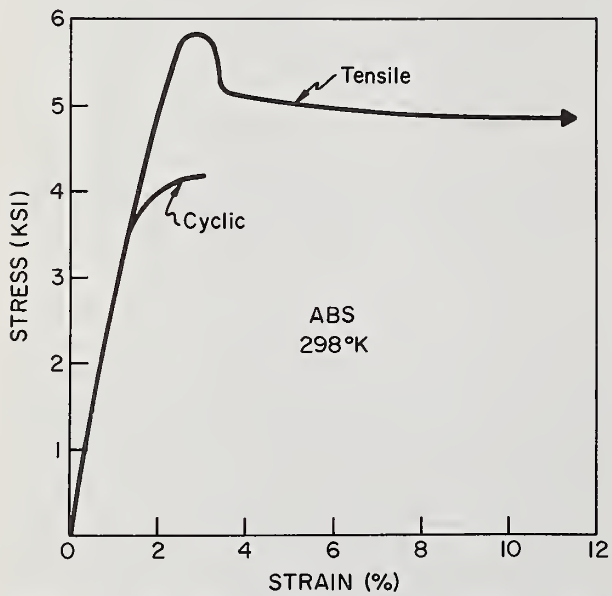


FIGURE 13. *Cyclic and monotonic stress-strain curves for ABS at 298 °K.*
(Taken from reference [5,] figure 23.)

The strain rate dependence of the glass-reinforced plastics is much greater than that of any of the metals studied thus far. The dependence of the ultimate tensile strength upon strain rate is shown in figure 15 for two types of commercially available glass-reinforced plastics. An 80 percent increase in the ultimate tensile strength occurs over the seven orders of magnitude change in strain rate.

When we turn to elastomers, the material problems become even more complex. Figure 16 is a plot of the number of cycles to failure against a parameter that is directly related to the strain energy. These data suggest that fatigue failure is essentially a cut growth process that takes place at small flaws or stress risers that are present in the material. The initial size of these flaws is estimated to be about 2.5 x 10^{-3} cm. Such cracks are not readily visible to the eye and will not become so

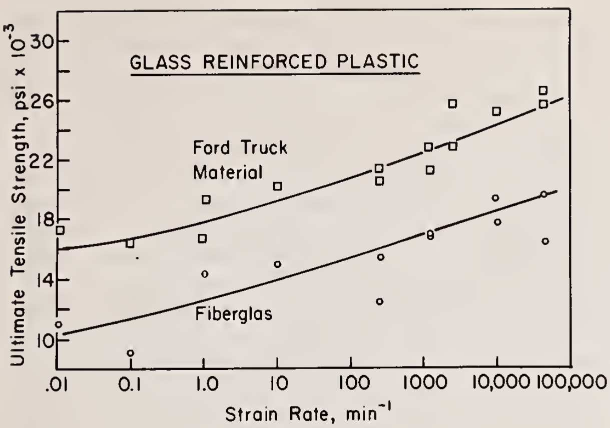


FIGURE 15. *Ultimate tensile strength as a function of strain rate for two commercially available glass-reinforced plastics.*
(Taken from reference [4], Figure 8.)

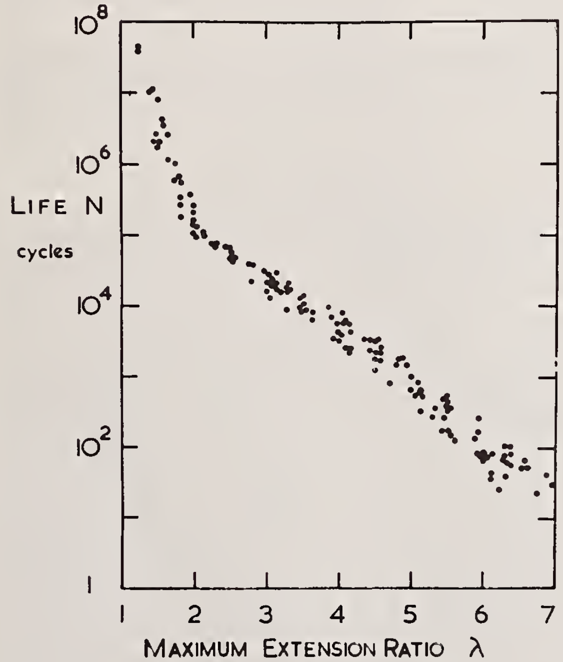


FIGURE 16. *Cycles to failure as a function of strain energy for die-stamped dumbbell test samples of vulcanized natural rubber.*
(Taken from reference [6], figure 7.)

until about 90 percent of the useful life has elapsed, in agreement with the general observation that failure may be imminent when the deterioration of rubber compounds becomes visible.

The combined effects of mechanical stresses and the atmosphere must be carefully considered in specifying rubber components. This is illustrated for unprotected natural rubber in figure 17 where the crack growth is seen to be dependent upon both the cycling rate and the presence of oxygen. The addition of an antioxidant to the material removed much of this frequency dependence.

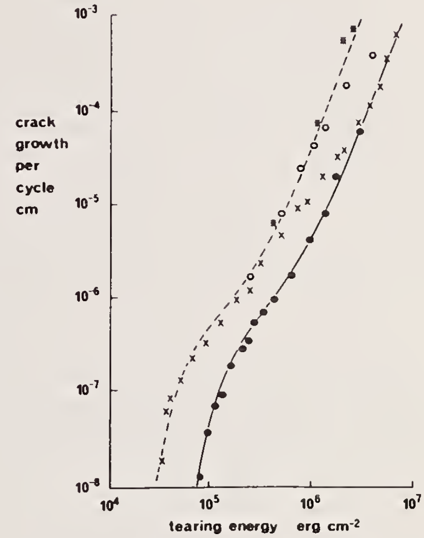


FIGURE 17. *Effects of frequency on mechanics-oxidative crack growth in an unprotected vulcanized natural rubber measured in laboratory atmosphere: (X)100 cycles/min.; (0)1 cycle/min.; (*)0.01 cycle/min.*
The solid curve was measured in vacuum at (o)100 cycles/min. Broken line is the same as the solid but displaced by a factor of 2.6. (Taken from reference [7], figure 16.)

I am compelled to observe at this point that it is an unusual design engineer who understands all of the key variables in such diverse materials as metals, plastics and rubbers well enough to be able to start from a design concept and arrive at an optimal design without a long series of "prove-out" testing. Just to emphasize again the point made by Harry Paxton in this Symposium, this is particularly true if there are no historic data available on a similar component.

There are many areas in which data are needed but are not adequate. A better characterization of the engineering properties of the more complex materials must be developed. An understanding of failure phenomena in these complex material systems is of vital importance. And this is particularly critical where we are engineering new materials. It seems to me that this represents a genuine opportunity for the science and engineering communities to cooperate in satisfying these needs. I submit that the engineering community needs more support from the materials engineers and the materials science community if it is to successfully meet the demands of the consumer for increased reliability with increasingly complex systems.

But as you well know, the choice of material, the development of a design, and the testing of the prototype components are only the first aspect of the design process. The component has to be manufactured, with perhaps a number of different companies involved, and then assembled into the system. Each of these steps presents opportunities for improper manipulation with a resulting reduction in overall reliability. The designer is confronted with a particularly difficult task when it comes to considering process specifications. It is usually preferable to specify performance rather than process, for in this way the processor will be able to exercise much more initiative and will produce a better product if he is given the maximum amount of freedom in developing a process that will yield a part with the prescribed properties. But still the designer has to be able to control the subtle differences in material properties that can result from even minor changes in the processing. This is precisely the reason why the process validation procedure is so stringent and can lead to what some people view as a reluctance on the part of the automotive industry to readily accept new processes. A new process can be introduced only when there is sufficient confidence in it to be certain that it will not adversely affect the reliability of the product. Establishing this may be a laborious process. Process variability, quality control, the statistics of sampling, the relationship of testing to field failures and the characteristics of the failures must all be established. When this is being done on a new design for which no field history exists, the problem is greatly compounded.

Although performance specifications are the most desirable from the processor's point of view, the sensitivity of some material properties to the process steps is so great that detailed processing specifications must be followed in some cases. As an example, I need only mention the importance of the close control that is needed to eliminate the possibility of hydrogen embrittlement on electroplated components. The ASTM has issued a detailed recommended practice for the preparation of high carbon steel for electroplating with a minimum of hydrogen embrittlement. This ASTM standard proceeds to note that "the test for the effectiveness of the procedures used to control embrittlement lies in the subsequent service use of the material." With no known rapid test for hydrogen in the part and no really adequate means of characterizing hydrogen embrittlement failures, the designer has no recourse but to specify the process steps.

I will not dwell on the effect of the assembly operations upon reliability. A general build-up takes place from the assembly of individual components into subassemblies. The final assembly operation then combines these into the complete vehicle. Testing and quality control are carried on throughout this operation in order to minimize assembly errors. The designer, however, has the task of developing his component and subassembly designs such that it minimizes the opportunity of inadvertent damage to the part during handling and reduces the possibility of having an improper connection, alignment or fixturing of the part. Obviously, the reliability is affected by the complexity of the final assembly.

What I have tried to indicate here is the breadth and complexity of the problem that confronts the designer as he attempts to consider the modes by which the component can fail. In an effort to help him prioritize the tasks that face him in his attempts to reduce the failures, we have formalized this process into what we call Failure Mode Effects Analysis (FMEA). This is a systematic approach to assessing the probability of occurrence of a failure *and* the effect that such a failure will have. Field experience is a great help in assessing failure modes, for such data are derived from systems that are in the hands of the consumer. But FMEA is particularly important for a new design where a particular component or part has never been evaluated by the customer and for which field experience is not available. In its simplest terms FMEA consists of describing the anticipated failure mode, the effect of the failure in terms of customer reaction, and the cause of the failure. The designer then proceeds by giving a numerical estimate to each of the following: the probability of occurrence of the failure, the consequences of the failure (which means the severity of the failure), and the probability that failure will be detected before it reaches the customer or will give the customer a warning that an incipient problem might lead to failure. These numerical estimates, when combined, offer a priority ranking among all the various conceived failure modes. Although this is, admittedly, a subjective process, it does require a careful examination of the trade-offs that are possible in designing, manufacturing, assembling, using, and servicing the part or system.

Throughout this discussion, I have made mention of the great benefits derived from field experience. While this is absolutely true, I must point out that the information channel between the customer and the manufacturer is oftentimes very noisy, appears to have numerous delay lines in it, and frequently has such an impedance mismatch at the ends that it is very nonlinear. The information content is oftentimes highly subjective, sometimes self-serving to the customer and may not reflect engineering reality. So we often develop multiple information paths. For example, we get information on warranty problems from our dealers; we support various surveys that ask present, past and, hopefully, future owners what they like or dislike about the product and what difficulties they have experienced with it; we survey repair data obtained on commercial fleets of vehicles; we collect information on our own fleets; we try to understand what product strengths and weaknesses affect the used car market; etc. All of this continues for several years for each model. Thus, a lot of data is developed and, imperfect as the individual segments may be, in the aggregate the information has been used effectively to improve the product. But as an illustration of the type of short-fall in the data, the time constant for developing high mileage field histories is, by its very nature, long, and the process of accurately verifying and correcting assumptions relative to customer usage is extremely difficult.

This brings me directly to the issue of this Symposium—Mechanical Failure—Definition of the Problem. I believe that the problem for the automotive industry can be rather succinctly stated in the following fashion: it is impossible now and for the foreseeable future to predict from first principles the absolute rate of failure for a mechanical system as complex as the automobile when it is subjected to the wide diversity of uses and to irregular maintenance that is characteristic of customer usage. Further, since the customer has the freedom to utilize his vehicle in almost any fashion, it is impossible to test against all contingent circumstances. Thus, we cannot expect to eliminate all mechanical failures. We can expect to minimize them, however. But in order to accomplish this, we will need a more accurate description of the performance envelope that the vehicle must satisfy, a much greater understanding of material properties, how these properties relate to mechanical behavior, and how the behavior is affected by process parameters and environmental effects. We need the continuing help of the universities, the national laboratories, and the not-for-profit laboratories in developing the proper models of mechanical behavior for materials that are not of the ultimate purity, ideal crystallinity, and of the simplest chemistry. Those aren't the materials we use now nor will we be able to use them in the future to any extent. And all of these material problems are going to be compounded as we search for alternate materials that can replace those in short supply, as we search for ways to reduce vehicle weight in order to improve fuel economy, as we search for materials that will survive the higher temperatures brought on by many of the emission control systems, and as we try to build vehicles that will withstand crashes at higher and higher speeds.

And finally, let us not forget that the cost that the consumer must pay for providing any solution must be commensurate with the effect that it produces.

References

[1] Sheth, N. J., Bussa, S. L. and Merces, N. M., Determination of accumulated structural loads from S/N gage resistance measurements, SAE Paper 730139 (Jan. 1973).

[2] Landgraf, R. W., and LaPointe, N. R., Cyclic stress-strain concepts applied to component fatigue life prediction, SAE Paper 740280, (Feb. 1974).

[3] Landgraf, R. W., Mitchell, R., and LaPointe, N. R., Monotonic and cyclic properties of engineering materials, Scientific Research Staff, Ford Motor Company (June 1972).

[4] Davies, R. G. and Magee, C. L., The effect of strain-rate upon the deformation of materials I—tensile behavior, (to be published in Journal of Engineering Materials and Technology).

[5] Beardmore, P., and Rabinowitz, S., Dynamic Factors affecting the structural uses of plastics, presented at the SAM Materials Symposium on Relationship Between Structure and Mechanical Behavior of Polymeric Solids, Chicago, October, 1973. Proceedings of the symposium to be published as an ASM seminar volume.

[6] Gent, A. N., Lindley, P. B., and Thomas, A. G., Cut growth and fatigue of rubbers. I. The relationship between cut growth and fatigue, to applied polymer science, Vol. **8**, (1964), p. 455.

[7] Lake, G. J., Corrosive aspects of fatigue, Rubber Age, p. 30. (Oct. 1972).

[8] ASTM Standard B242–54 (Reapproved 1971). See sec. 8.2.

Discussion

E. Passaglia, National Bureau of Standards: You've concentrated your discussion on design considerations in trying to prevent failures. Yet there are two things right now on all automobiles which, while they don't prevent failures, do indicate the approach of incipient failures. These are the ammeter and the oil pressure gage. Are there any thoughts of using predictive devices of this kind for other critical components in automobiles?

W. D. Compton: Yes, there is a considerable amount of research effort directed at developing better information transducers for use onboard vehicles. You have probably found the change from ammeters and oil pressure gages to lights that blink frustrating. That change was made because many people won't read a dial. But there are a lot of other indicators under consideration such as tire pressure indicators. If electronic transmission controls are employed, undoubtedly there will be some kind of torque indicator. This would be a direct measure of the condition of the transmission. As these kinds of sensors are added, the wiring harness becomes more complex, and the probability of an electrical failure is increased.

R. E. Maringer, Battelle Memorial Institute: You stated that a service history is of great value to the designer in correcting defects. Are there any serious attempts in the industry to analyze the cause of failure in a catastrophic accident?

W. D. Compton: We expend a lot of effort in research activity doing just that. Obviously, it has to be done in all cases of litigation and it is done in a large number of others. The service organization does a great deal of this also. One of the problems is that sometimes a part or component is so badly damaged that little information can be obtained without an unbelievable amount of work. The analysis of a failed transmission, which had not been abused after failure, revealed a problem with one of our material suppliers. If the owner of the car had abused the transmission by continuing to drive the car after the transmission failure, this problem probably would not have been discovered.

H. Corten, University of Illinois: You were talking about your performance envelope. Once you have this, I am sure it's going to show that there are a few vehicles that are abused rather severely, a few more that are abused less, and the remainder which are in the normal usage range. What philosophy are you going to use to decide how much abuse to design for?

W. D. Compton: The design is based on evaluation of cost effectiveness. That will get us right into the argument as to whether the governmental regulation should insist that you design for all possible uses or for average use. The clean air amendments dealing with the control of vehicle emissions require the emissions of all automobiles to be below a specified value. It costs the consumer a substantial amount to prevent any vehicle from exceeding that value. We believe that this kind of regulation must be approached with the realization that there are going to be statistical distributions and that there should be allowance made for some excess if it's balanced by the proper distribution below the average. That's a somewhat easier thing to discuss in terms of vehicle emissions than it is in terms of safety, because the vehicle emissions contribute to an overall ambience whereas safety devices protect the life and limb of one individual.

A. J. McEvily, University of Connecticut: I was interested in the emphasis you placed on strain rather than stress determination in connection with fatigue analysis. Does that carry over to the idea that low cycle fatigue due to abuse at the high level of hitting pot holes, etc., is controlling and causing fatigue failures, when they do occur, rather than high cycle fatigue in which case you might use stress or strain interchangeably?

W. D. Compton: There is an effort to try to eliminate the assumption as to what the relationship is between stress and strain and to use directly in the laboratory that which is measured on the component in the field. If the stress-strain relationship is known there is no reason to use one in preference to the other.

J. M. Karhnak, U. S. Army, Fort Belvoir: You mentioned that transducers were built into some components to give information outputs at periodic intervals. Would you comment on that?

W. D. Compton: These are accumulated load gages. It's simply a use of three or four of these gages in conjunction with mechanical multipliers, so that there are several levels of sensitivity as you go up the line. The outputs of these devices can be used to produce a histogram of the total amount of accumulated damage. There's a SAE paper on this subject that is referenced in my paper.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Implications for Action—Industry

John E. Ryan

General Electric Company
570 Lexington Avenue
New York, New York 10022

This paper will review product trends for industrial and consumer durables; the foreseeable environment of production and use of these products and systems; and the implications for preventive action in this industrial area based upon the growing body of knowledge of mechanical failure mechanisms and the increasing concern for economic and social consequences of mechanical failures.

Key words: Consumer products; marketing; mechanical failure; product performance; product reliability; product testing.

My position on the program, as a representative of industry, implies that we are the ones who will "put it all together." The public—which is but a collective term for our customers—expects this; the title of my paper "Implications for Action—Industry" underscores it; and industry gladly accepts responsibility for achieving integrated, optimized results in the products it supplies.

Let's pause and consider what the word "failure" really means before we charge off on action plans. Running through all the shades of meaning listed in the dictionary is the common thread of being unsuccessful, inadequate, falling short of the goal. All of these imply that there *is* a goal, a bench mark performance, for goodness. Indeed there is, but it is often a largely subjective one in the mind of the user of a product and likely to be considerably different from one individual (or group of people) to another. Add to this variability the fact that the bench mark is often a moving target and you will appreciate that an industrial company has to be a pretty good marksman to avoid failures of the product or the business.

Failure of products, we all know, is of two distinct types: (1) the slow deterioration of performance as in bearing wear; (2) the sudden, often catastrophic breakdown of a part or assembly such as occurs in fatigue fracture. The first type can often be mitigated by proper maintenance but with consumer products—and many capital equipment products too—users have come to expect a relatively maintenance-free product. So we marshall all available technology and try to give them what they want even though we may have difficulty in communicating to each other an acceptable definition of reliability. Failures of the second type—the sudden breakdowns—can, of course, be minimized by careful design but the very word "design" implies a certain kind of usage, a certain intensity of loading. Here's where a manufacturer and a user may have quite different ideas—each in all sincerity.

This is all part of what has been called the Space Syndrome. The consumer says, in effect, "If you (meaning industry) can put a man on the moon, why can't you make a really reliable product for me?" Not only the individual consumers say this; the big industries and utilities sing the same tune, sometimes with concern for substantial loss if failure should occur.

The reasons behind this attitude are quite understandable. On the one hand, we are becoming increasingly dependent upon our machines, both in the home and in industry. On the other hand, we cannot afford the extra cost of standby equipment, at least not to the extent that we once thought was prudent and justifiable.

In a more fundamental sense, such attitudes seem to stem from a growing concern for the rights of the individual. Arthur E. White, in last October's issue of the "Public Relations Journal" calls this phenomenon "the galloping psychology of entitlement"—the process by which a person's wants and desires become converted into presumed rights. He points out that this is "indeed a very old trend, long recognized as part of a worldwide revolution of rising expectations. But in recent years it has accelerated and assumed new political and institutional forms."

The earlier papers in this Symposium considered quite thoroughly what implications the facts of mechanical failure have for science, engineering and manufacturing. Trying to figure out what I should cover, I wrote a little equation.

$$\text{Industry} - \text{Science} - \text{Engineering} - \text{Manufacturing} = \text{Marketing} + \text{Finance} + \varepsilon$$

I threw in the epsilon to cover ingredients I might have overlooked but the more I thought about it, the more I realized that epsilon contained a mighty important entity—general management.

The Marketing function can, and certainly should, describe in functional terms for the business, what various segments of their market want in a product and what value (plus or minus) these potential customers attach to each product characteristic—at least stochastically. The scientists, engineers and manufacturing men respond, preferably in a give-and-take iterative fashion, while the Finance fellows organize the cost and price consequences of the various alternatives into a meaningful array. Then comes the moment of truth, when the general manager and his functional associates, as a team, choose the course that will be followed. The inevitable trade-offs are made, including those involving level of performance and reliability. At these critical junctures, the voice of the engineering manager must be clearly heard, not only as a technical witness but as an active advocate.

However intuitive such final product decisions may be, once all measurable facts are on the table, the guiding principle is to maximize the ratio of total value—as the intended customer regards it—to cost, the total cost to produce, sell and use it. Since price can usually be set in relation to value, at least as the intended customer perceives value, that all-important difference between price and producer's cost, viz., PROFIT, is the crucial consequences of these general management trade-offs that I have been talking about. "What's new," you say, "that's elementary economics." Yes, you're right; at least in theory, product decisions have always been based on what the management of the business believed was a maximum ratio of perceived value to total cost. Note that I said "perceived"; I mean "as the customer perceives it." A reputable, ethical manufacturer knows that perception extends far beyond the day of purchase. He intends to stay in business so he needs repeat business. That means that actual performance must coincide pretty well with represented—or implied—performance; otherwise the disillusioned customer will go elsewhere.

"Elsewhere" implies that there is indeed a better source. We fervently hope that the time will never come when further improvement is not possible, when no one ventures to improve lest he fail. However, under the increasing pressure for "more" that exists today, there is a grave but understandable temptation for manufacturers to weight less heavily such characteristics, like reliability, that are not so easily or so immediately perceived and rewarded by the customer. To be sure, such cutting of the corners, either deliberately or through oversight, eventually catches up with those who play that way but the time lag between purchase and disillusionment introduces instability in the marketplace and a frustrating distrust and hostility among all parties concerned. We believe the only way to arrest such deterioration is for industrial companies to subscribe to, and live by, professional principles from the top to the bottom of their organizations. In the dictionary sense "professional" means "pertaining to a vocation that requires specialized knowledge for its practice." I mean that—and more. As one writer put it: "Professionalism is characterized by obedience to the unenforceable."

As manufacturers we must expect our customers to "ask for the moon," so to speak. For the most part, they have only a general idea about what product characteristics are difficult to obtain simultaneously. Consider the following opposed pairs, both elements of which a customer might expect in a product offering.

Versatility versus Operational simplicity
Accessibility versus Safety
Compactness versus Serviceability
Special features versus Reliability
Speed or response versus Stability.

These rising expectations in the marketplace, along with rising costs, mean that manufacturers will have to rely increasingly on the best technology available to deliver durable, reliable products. Those comfortable "factors of ignorance", euphemistically called "factors of safety", are going to have to be reduced because of both economic forces and the social pressures for conservation of natural resources through the more efficient use of engineering materials and energy. We shall have to use not only more complete and exact theory; we shall have to do better planned, more truly indicative testing—and do it more efficiently, in less time.

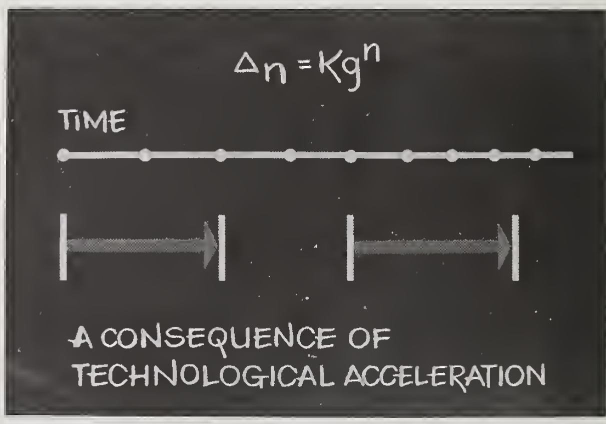


Figure 1.

This is a big order, but you'll see from figure 1 why it is inevitable. We all acknowledge that technological developments are occurring at an accelerated pace. Let's quantify that; let's assume that advances in a given field occur in quantum steps of uniform magnitude but, as I said, at an accelerating rate. These I have represented as dots on a uniform time scale, the time increments decreasing according to the equation:

$$\Delta_n = Ke^{-gn},$$

where n is the increment number and g is a positive constant. If a given dot is plotted at the point in time when it emerges as an established research finding, consider the consequences of a fixed length of time between that moment and the time when it is embodied in a proven product, ready for sale. Some while ago such a gestation period would have spanned, let us say, two technological milestones; in other words, only two as-yet-unapplied units of knowledge would be in the "mill" of reduction to practice. Contrast this with a later period in which, perhaps, three such units would be advancing through the development, design and test sequence. Note the advantages that a fast-moving but sure-footed manufacturer has in such a situation. With proper technical assurance, he can elect to run with a later, more advanced and presumably more productive technology and arrive at the market place one quantum jump ahead of his competitors.

Recognizing that the winner always sets the standard for the next try, it is obvious that means must—and will—be found to accelerate development, design and test activities. I have chosen to comment mainly on testing because it is often the most time consuming, not only for the final product, but for components all during development and design.

Testing products or components for degradation or failure in service is so time consuming that efforts have always been made to accelerate the degradation process, relative to real life,

(1) by making the cycle rate faster and
(2) by increasing the intensity of the various stresses (temperature, force, velocity, etc.).

Simple as this sounds, it can lead to erroneous results, as every engineer who has lived a while knows. The main traps I believe, are these:

(1) Increasing the cycling rate may produce secondary effects, such as higher temperature, resonances not present in real use, etc.
(2) Mutual effects of several applied (or environmental) conditions in real life may be unexplored if the common "one variable at a time" approach is taken for the sake of procedural simplicity.
(3) Intensifying the independent variables, singly or in combination.
(4) Extrapolating limited data without knowing, or worse yet, knowing but not applying the available theoretical knowledge of the failure mechanisms at work.

Because linear relationships are eaiser to handle mathematically than nonlinear ones, we are all too prone to linearize phenomena even though we know that most cause-effect relations are, in fact, nonlinear. But bcause that nasty reality makes superposition of experimentally determined effects questionable, if not downright wrong, we tend to ignore nonlinearities.

Another testing foible, found in academic as well as industrial circles, is to do very thorough and precise testing on variables for which the testing is easy or where our laboratory's superdeviometagraph can be employed. Other equally important variables may be considered only sketchily. We may even be following meticulously the official test procedure of the ASKY (!) but applying it to a situation that could not have been anticipated when that hallowed document was written!

I cannot leave this matter of accelerated testing without commenting on extrapolating limited data without benefit of theory. Among the practical engineers of industry there is still a tendency to lean on intuitive experience more heavily than on theory when assessing limited data. Fortunately, theory is gaining in credibility every year, not only with the younger engineers, fresh from school, but with the older ones as well. The famous English consulting engineer Urwick, speaking after years of experience, stated the point very well when he said:

> "We cannot do without theory. It will always defeat practice in the end for quite a simple reason. Practice is static. It does, and does well, what it knows. It has, however, no principle for dealing with what it doesn't know—Practice is not well adapted for rapid adjustment to a changing environment. Theory is lightfooted. It can adapt itself to changed circumstances, think out fresh combinations and possibilities, peer into the future."

Fortunately, we have available today some techniques based on sound theory, that can help us to hew closer to the line in the use of materials while predicting with quite satisfactory accuracy the risks of failure. The list includes:

Failure mode and effects analysis
MTBF and MTTR analysis
Decision analysis
Design optimization methods, extended through computerized procedures
Probability applications of many types.

So far I have discussed only testing in the manufacturer's laboratory, how to accelerate the process without impairing validity, etc. How about testing in actual service? When a product goes into service, a long and very real experiment begins, whether we call it by that name or not. Unfortunately, we all too often do not take even cursory data on it. True, a given product unit may be thousands of miles away and we can't expect our customer to take its pulse regularly for our benefit. But when malfunctions do occur, especially within the warranty period, we do investigate with our own field service people. Resourcefully, these men fix the trouble, write a brief report (no time to communicate more than the highlights!) and go on to the next problem.

Unless such failure reports and diagnoses get organized and classified somehow as they come into headquarters from every direction, much of their value is lost. This need not happen. For more than a decade computers with large memory capacity have been available and, in fact, have been successfully used by many companies to store, then sift the field service reports so that corrective action, both in theory and practice, can benefit future products. These companies have discovered that, though the significant facts had always been present in those file drawers of reports, the design engineers saw only glimpses of field performance until the computer brought its searching and organizing power to bear. This improved procedure is far from universal today. We in industry who have experienced the benefits must urge them upon our more reticent fellow manufacturers.

In concluding, I want to say a few words about the role of government in all of this. Certainly it is proper for government—federal, state or local—to enact and enforce laws that protect the citizen's person and property from (1) risks that he, in general, cannot be expected to detect and understand, and (2) those that he cannot control even if he does know of them and understands them. But inherent in all regulatory authority is the tendency to prescribe in even greater detail how a certain degree of reliability or safety must be achieved rather than what is to be achieved. This tendency toward the "how" in codes and regulations occurs because it is often difficult to define a result or effect in precise quantitative terms. Yet it must be done or we in industry who are subject to the regulations will be caught on one of two horns of the dilemma: (1) We will have our opportunity to innovate stifled by detailed prescriptions based on past technology, or (2) we will be subject to the whims of bureaucrats as they interpret such weasel words in the law as "*substantial* hazard" and "evidence that *reasonably supports* violations". Ethical industrialists have nothing to lose and much to gain by assisting our lawmakers and administrative agencies in making regulatory legislation fair, properly aimed and precise in language. Unless we are willing to get into the game, we cannot, in good grace, criticize from the sidelines. As Thomas Jefferson once said, "For evil to prevail, the only requirement is that good men do nothing."

IMPLICATIONS OF MECHANICAL FAILURE

Session VI

Chairman: E. E. Klaus
The Pennsylvania State University

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Provention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

The Implications of Mechanical Failures for Consumer Product Safety—Vice Versa

Lawrence M. Kushner

Commissioner
Consumer Product Safety Commission
Washington, D.C. 20016

In years past, manufacturing a safe product (rather than one that is simply marketable) for the consumer was regarded as a nice, but not necessarily imperative, thing for a manufacturer to do. Recent legislation, however, mandates consumer product safety, thereby putting additional responsibilities on product designers and engineers and the research scientists who support them. Experience in operating a regulatory product safety program discloses that materials failures, many of which are mechanical in nature, are an important factor in consumer product safety. A discussion will be presented of the implications of the Consumer Product Safety Act for materials engineers. Some specific examples of mechanical failures in consumer products that have created serious safety problems will be given.

Key words: Consumer product safety; failure prevention; government action; government responsibility; mechanical failure; safety standards

I believe that what I have to say will be both interesting and encouraging to you in your efforts to attract the level of attention to mechanical failures—their causes and prevention—that the subject merits.

I'm personally sympathetic with any group of scientists and engineers who is trying to motivate its research in terms of detailed knowledge of real-life problems. For that group of scientists and engineers to be concerned also about transforming the increased knowledge resulting from its research into specific developments which benefit the public is an additional plus. It is my understanding that both of these are consistent with the objectives of the Mechanical Failures Prevention Group.

As far as mechanical failures are concerned, I have a personal scientific interest in the subject which goes back to my days with the Metallurgy Division here at the Bureau. But more recently, my interest in mechanical failures has been heightened by realization of both the implications of consumer product safety for the study and prevention of mechanical failures and, on the other hand, the implications of success in dealing with mechanical failures for the safety of consumer products.

Indeed, this interest is more than a personal one. The agency which I represent today—the Consumer Product Safety Commission—has acknowledged its interest in the purposes and programs of the Mechanical Failures Prevention Group by its cosponsorship of this meeting along with NBS and six other organizations.

The principal way in which the national concern for consumer product safety is going to impact on the objectives of MFPG is through the provisions of the Consumer Product Safety Act of 1972.

The impact is going to be a beneficial one.

For those of you whose companies make consumer products or supply the companies who make consumer products, this law is going to heighten your management's concern about mechanical failures, and indeed about any other aspect of product design and performance which bear importantly on the safety of consumers when using their products.

For those of you who are not in consumer product industries, some of the programs that will be set up under the Consumer Product Safety Act—and the way in which they will be run—will be of interest and offer the opportunity for useful participation in technical matters of public concern.

Let me turn now to a brief discussion of the Act and detail further its implications for this audience.

In 1972, Congress and the President made a forceful commitment to the notion that consumer products should be safe. That commitment is embodied in the Consumer Product Safety Act. That Act offers opportunities—and responsibilities—for engineers and scientists.

The law heralds a new commandment—"Consumer Products Shall Be Safe." Its purpose is to introduce a safety ethic into consumer product industries—to put safety on an equal footing with the traditional concerns of top management—cost, style and fuction.

As a matter of fact, safety may well become a little more equal than the others. It isn't against the law for a product to be too expensive. Nor is it against the law for a product to be unattractive or not to work. But under the new law, putting an unsafe product on the market can lead to civil penalities as high as $500 000 or criminal penalties up to $50 000 and a year in jail.

The Consumer Product Safety Act established a new, independent regulatory agency, the Consumer Product Safety Commission, which is charged with reducing the hazard posed to consumers by unsafe consumer products.

The definition of a consumer product under the law is broad. The law says that a consumer product is any article that is manufactured for sale to, or for use by, a consumer in or around the home, at school, in recreation or otherwise. There are some exclusions—products that are already regulated by other Government agencies. These include foods, drugs, cosmetics, automobiles, boats, airplanes and firearms. But what remains is impressive, generally estimated at about 10 000 products including toys, gas and electrical appliances, power tools, recreational equipment, furniture, clothing, doors, windows and on and on.

The range of regulatory options available to the Commission is also broad. Which one is used in a particular case depends on the severity of the hazard. For "imminently hazardous" products, immediate ban and seizure are in order. For those requiring less immediate action but which are serious nevertheless, there are mandatory standards, nonconformance to which is a violation of law. For lesser hazards one can consider voluntary standards. And certain kinds of hazards may best be dealt with not through standards or other regulatory actions but through information and education programs directed to the consumer.

One of the features of the law requires that if a manufacturer, distributor or retailer obtains information that a consumer product either does not meet an applicable standard or has a defect in it and therefore could present a "substantial hazard" to its user, then that party has an obligation to inform the Commission immediately. The Commission can then require the manufacturer to recall, repair, replace, or repurchase the defective product. The Commission can require that the public be notified through a widely distributed press release. This is the kind of publicity any responsible company would like to avoid.

This spectrum of regulatory options plus the penalties that can be assessed should provide effective motivation for the development of a safety ethic in the consumer product industries. We should see the development of an environment in industry in which scientists and engineers of all kinds can express their latent interest in safety.

In this context, the prevention of mechanical failures should surely get high priority attention.

I have been told that before the Consumer Product Safety Act became law, the top management of many companies manufacturing consumer products believed that building safety into their products would make them too expensive to be successful in the highly competitive consumer product market. My perception now is that it may be too expensive for a company not to.

The Consumer Product Safety Commission has been called potentially the most powerful Federal regulatory agency ever created. One of the most imaginative and amusing descriptions of the Commission is that offered by the syndicated columnist, James Kilpatrick, last year.

> It must have been quite a moment for Dr. Frankenstein when his monster got off the table and walked . . . The Congress might well have put itself in the same fix with the Consumer Product Safety Commission.

Is the law overkill? After all, we're only dealing with stoves, toasters, lamps, power lawnmowers, matches, and other apparently innocuous products. Yet the accident statistics are appalling and do, in fact, describe a serious problem which merits national attention.

It is generally estimated that there are over 20 million consumer product-related injuries a year which require medical treatment. These 20 million accidents include over half a million hospitalizations, more than 100 000 permanent disabilities and 30 000 deaths. These are the dimensions of a problem to which industry must now direct attention.

For those of you in industry, the challenge—and the opportunity—are clear. And hopefully the challenge will be transmitted to the academic community as well. The opportunities for imaginative, exciting, and useful research and development are plentiful. Particularly when one considers the cost constraints that apply. After all, this R&D is being motivated not by "cost plus" contracts with the Government, but by the demands of a competitive market—an extra constraint, but an additional challenge.

The opportunity for scientists and engineers goes beyond laboratory R&D. It extends to the development of effective, technically valid consumer product safety standards—both voluntary standards under private sponsorship and mandatory ones being developed for the Commission. The motivation for industry people to be heavily involved in developing these standards is clear.

But there is a long-standing complaint that the interests of consumers have not been adequately represented in the development of even those few voluntary standards that apply to consumer products. We hope that more voluntary safety standards will be developed. But in order for such standards to be credible to consumers, there must be technically competent consumer representatives on the committees which develop them. And this is equally true for mandatory standards.

The basic problem is a financial one. When engineers from industry or the Government serve on standards committees, their expenses are paid by their respective companies or Government agencies. But there has been no one to pick up the tab for technical people to serve on these committees on behalf of consumers.

The Commission is going to do two things to help improve this situation:

We're going to develop a list of qualified, technical people who are interested in serving on standardization committees as consumer representatives. The list will be compiled following a public solicitation addressed to the technical community and will be available to any standards-developing groups seeking qualified technical experts as consumer representatives. More importantly, the Commission will pay the expenses of such representatives in the development of standards in which the Commission has a particular interest. This applies to both mandatory and voluntary standards.

I think you will agree that the Consumer Product Safety Act provides many avenues through which the technical community can address a whole set of safety-related problems including mechanical failures.

Are mechanical failures an important source of consumer product hazards? I believe the experience of the Commission amply demonstrates that they are *very* important indeed.

I mentioned earlier in the talk the requirement of a manufacturer, distributor, or retailer to report to the Commission any time he has reason to believe that a product in commerce could, because of a manufacturing or other defect, present a substantial hazard to a consumer. In the year that the Commission has been in existence, we have received well over a hundred such defect notifications including, but not limited to, bicycles, TV sets, gas ranges, power mowers, cigarette lighters, air conditioning system controllers, garden tractors, and even reclining chairs with heaters built into them. In approximately 40 percent of these cases, the hazard resulted from actual or potential mechanical failures. In some cases it was poor mechanical design. In others, it was improper materials selection. In others, it was a combination of both. In still others, the hazardous products appeared on the market because of a quality control breakdown.

Let me cite a few cases:

CASE A—The product is a heavy duty rotary lawnmower. The problem is the improper selection of materials for the unit which attaches the blade to the drive shaft. In this particular case the attachment assembly became embrittled after welding and subsequent carburization and heat treatment. The critical part suffered brittle fracture while the mower was in use, the blade separating from the shaft. The manufacturer had to replace the defective parts on approximately 450 production units that had already left the factory—some of which were in the hands of consumers.

CASE B—A ride-on lawnmower. Because of improper design of some spiral pins which fastened gears to shafts in the steering mechanism, under repeated high load stress reversals, the pins worked their way out and steering control was lost. Over 4,500 mowers were affected. All of the known purchasers of these units were contacted by phone or letter and provided repair kits at no cost.

CASE C—Defective design in a bicycle fork stem involving over 200,000 bicycles. Corrective action is being undertaken by the manufacturer.

CASE D—Another ride-on lawnmower—this one presenting the possibility of failure of a brake cable due to improper design and materials selection. Over 6,000 production units are involved. The company is identifying all of the customers and sending service bulletins and repair kits.

As a final item let me quote from the Product Safety Letter, a newsletter which covers the consumer product safety scene in the regulatory agencies in Washington. This is from the May 6, 1974 issue.

> Schwinn recalled 70 000 bicycles to replace imported pedals after spindles fractured due to faulty heat treating.

Mechanical failures are a *very* important problem in consumer product safety. The costs to manufacturers of taking the corrective action required by the Commission are high. Not only is one concerned about the immediate, out-of-pocket dollar cost of the new parts and labor, in the long run the unfavorable publicity that inevitably accompanies a manufacturer's recall may be far more costly.

One of the important components of any program to reduce mechanical failures is the development of good nondestructive evaluation techniques. Let me give you a specific example of their importance to the Commission.

The Commission has been involved in rather extensive debate about the incorporation of sampling plans in its mandatory standards. The debate is not yet over.

The sampling plans that are the subject of this debate would set the sampling procedures which a manufacturer must follow in testing his product for conformance to the standard. One side of the argument holds that sampling plans are needed in those cases in which the test embodied in the standard is destructive or, for one reason or another, 100 percent testing of all production items is unfeasible. A sampling plan in a mandatory standard would acknowledge the practical fact that short of 100 percent testing, no production process can be completely foolproof and that in any production batch there is some expected fraction of defective products. In other words, putting a sampling plan into a standard makes the standard applicable to a production batch or a production lot rather than to an individual product.

Now this position is great as far as manufacturers are concerned. And the use of sampling plans makes sense for those consumers such as DOD, GSA, or industrial purchasers who buy large quantities of a particular product and are willing to accept a statistical statement of quality.

But on the other hand, how about the individual private consumer who buys only one item and not an entire lot? If he gets one of the defective products, the likelihood of which the standard acknowledges if it has a sampling plan built into it, what protection has that consumer been given by the mandatory safety standard?

The answer to the dilemma is the development of nondestructive evaluation techniques which can be incorporated in the production process and could give the manufacturer confidence of 100 percent compliance, even though he may not have subjected 100 percent of his production to the proof tests in the standard itself.

It shouldn't be necessary to lower the level of protection afforded by a standard because of inadequacies in the technology of nondestructive testing. You people are in a position to do something about this.

In closing, let me point out that the real purpose of the Consumer Product Safety Commission is not to issue regulations. It is not to levy fines, nor put people in jail. Our real purpose is to get manufacturers—and indeed all of those who are involved in the sequence of events from product design through production and distribution—to think product safety.

In my view, this is an area in which technology that has been developed for sophisticated space programs and military programs can be converted to be useful in the consumer product industries. The challenge is to do it at a price that manufacturers and consumers can afford.

The threat of regulatory action in the consumer product industries is now so real that an environment conducive to achieving the objectives of the Mechanical Failures Prevention Group should exist. I hope you will take advantage of it.

The safety of consumers would be greatly enhanced by your success.

Discussion

J. E. Ryan, General Electric Company: I would submit that sampling techniques have long been established as being the best assurance within a cost benefit frame that we can get. The fact that the DOD has accepted sampling, after much study, as the best cost effective way of handling the quality of weaponry signifies that it is technically sound. To use the case of the DOD, consider the soldier who goes into the jungles of Viet Nam with an M16 carbine and because of the sampling plan there is a remote chance that his gun won't fire, so he gets killed. Are we going to try to shield our consumers beyond the point to which we would shield our soldiers?

L. M. Kushner: The fundamental point here is the difference between the DOD as a consumer and the individual consumer as a consumer. The DOD as a consumer is buying millions of rifles and, as I mentioned in the talk, it's reasonable for them to accept some statistical statement of quality control. I don't think that that is necessarily a reasonable thing for the individual consumer. Now I recognize that in instances in which the only test available is a destructive test, and 100 percent testing is 100 percent destruction and complete safety, 100 percent testing is not reasonable. I don't think there is a simple answer to the question. I think the way the Commission is ultimately going to come out here is eminently intelligent. We will use sampling plans where they have to be used and try to avoid using them where they don't need to be used. One of the ways that you can avoid having to use them is to have the necessary equivalent nondestructive testing that can be incorporated right into the production line so that every single item is subjected to the nondestructive test which, if passed, could give the manufacturer essentially complete assurance that the item would pass a destructive test if one needed to be performed.

J. E. Ryan: I will defer to the experts on non-destructive versus destructive testing. But I would rather take a destructive test sample plan than to temperize with a supposedly equivalent non-destructive test which inevitably, if it is to be more than a token, will greatly increase the cost on the line.

L. M. Kushner: As I mentioned, the challenge is to do it at a cost that people can afford.

H. Corten, University of Illinois: In looking at products we see a spectrum of strengths and a spectrum of loads and you have been talking primarily about the removal of defects in the spectrum of strengths. How will you decide what is a reasonable spectrum of loads and when something has been "used" as opposed to "abused" in judging failures?

L. M. Kushner: We will try to write into the test methods incorporated in the standard a range of loads that is reasonable to expect in terms of normal use as the product ought to be used and ways in which we can reasonably expect that the product is going to be misused or abused. I would put the same responsibility on the designer and the manufacturer when designing a product. They must take into account that the product will probably in most cases end up being abused or misused as many times as it is used properly. One of the messages that we are trying to get across to product designers is that it isn't going to be sufficient to make a product which is safe only if the instructions are followed very carefully, because it is known ahead of time that instructions are not going to be followed very carefully.

J. Rabinow, National Bureau of Standards: Returning to the subject of testing, isn't it a fact that if you do 100 percent testing, there is still no assurance that, in practice, the product will not fail? For example, in testing bottles, each bottle may pass the test, but some may still rupture later. In fact, the testing may increase the possibility of rupture. There is no guarantee that 100 percent testing will guarantee 100 percent reliability.

L. M. Kushner: I'll have to accept that, particularly if the test could weaken the product.

C. G. Interrante, National Bureau of Standards: Many of the primary metal producers have equipment operation policies that involve extensive safety programs. The principal benefit from these programs is lower accident rates. These programs have been effective to the point that it is claimed to be much safer for the husband to work in a steel mill than it is for the wife to be at home. Are there plans to make these kinds of programs available to consumers?

L. M. Kushner: Absolutely. Information and education programs are an important part of the total range of options that the Commission has to deal with particular subjects. Of the 20 million accidents that occur each year, only 20 to 25 percent of them could be eliminated by a change in the product. We're sufficiently sophisticated to understand that. In order to make dramatic inroads, that is approaching a 50 percent cut in accident rates, there is a tremendous educational and informational responsibility requirement.

C. G. Interrante: Do you have any idea why the media hasn't undertaken such educational programs?

L. M. Kushner: The news media is simply not really an educational device as such; its primary function is presenting news. Banning a product is much more newsworthy than telling people how to use it effectively and carefully.

W. R. Bozman, National Bureau of Standards: How do you feel about the safety of products which are used in a way that the product was not originally intended to be used? For example, the small shelf on a stepladder that is designed to hold a bucket of paint is frequently used by people to stand on. There may be a sign on the ladder warning people not to step on the shelf, but they will do it anyway.

L. M. Kushner: Product misuse such as that is foolish, but on the other hand it's to be expected and that is the situation that the Commission is trying to cope with. We think that manufacturers have a responsibility to deal with this problem too. Misuse is often predictable and in some cases a product can be designed to anticipate that kind of misuse or abuse. It's also true that the intelligent consumer is going to have to pay a little bit more because of a concern for the proclivities of the unintelligent one.

W. D. Compton, Ford Motor Company: What is your thinking about self-certification by industry in terms of compliance? Do you foresee establishing fixed numerical standards or fixed performance criteria that the manufacturers will have to certify having met before a product can be sold much as the clean air amendment now requires in the emissions area?

L. M. Kushner: The law does require that products subject to a mandatory standard of the Commission must be certified by the manufacturer as meeting those standards. I have no objection to self-certification in a mandatory program because there are substantial sanctions for violating the standard. If we find a violation, we can take people to court and we can impose penalties, so there are real incentives for everything being on the up and up. On the other hand I am a little bit more leery of the use of self-certification in voluntary programs because there are no sanctions to give people the assurance that there are sufficient motivations for the companies to be totally honest in the way the certification programs are used.

W. D. Compton: What would be the result in a certification program where the product was properly certified and approved by the agency but then was found to be faulty in the field? Is that still the manufacturer's responsibility?

L. M. Kushner: The Commission's action would depend on a whole set of circumstances other than just the specifics that you identified. Conformance with a standard, as far as private liability suits are concerned, does not grant any immunity in the courts.

R. F. Scott, Jr., Travelers Insurance Company: Are there any mandatory standards in effect at this time?

L. M. Kushner: Yes. There are no standards that have been issued under the Consumer Product Safety Act itself, but the Commission has been given responsibility for several older laws. The three most important of these are the Flammable Fabrics Act of 1953 as subsequently amended, the Hazardous Substances Act, and the Poison Prevention Packaging Act.

R. F. Scott: But you have not propagated any new standards of a mandatory nature?

L. M. Kushner: That's right, because, under the law, there are requirements in terms of the kinds of procedures under which mandatory standards can be developed for the Commission. We have taken a long time in writing regulations and procedures because we're trying to encourage many new kinds of groups to get involved in the sponsorship of standards development. Regulations have recently been published in the Federal Register and we are now in a position to undertake the development of mandatory standards. We will be doing that for architectural glass, power lawn mowers, matches, electrical extension cords, and one or two other products.

G. McPherson, Jr., University of Missouri, Rolla: Are you empowered to do research, to develop standards and perhaps develop non-destructive testing techniques, and is the Commission going to do the research or is it going to sponsor research in companies and universities?

L. M. Kushner: We have authority to conduct research and development programs related to the mission of the Commission. We have our own technical staff that by and large spends its time fighting fires. We haven't yet formulated a long term R&D program, but when we do, we will be seeking proposals from the entire technical community.

NATIONAL BUREAU OF STANDARDS SPECIAL PUBLICATION 423, Definition of the Problem, Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group, May 8–10, 1974, held at NBS, Gaithersburg, Md. (Issued April 1976)

Mechanical Failure—A Material Matter

Richard W. Roberts

Director
National Bureau of Standards
Washington, D.C. 20234

The Federal Government has long been concerned with minimizing mechanical failures. This concern has come about for two reasons. First, the Federal Government purchases and operates a great deal of equipment, much of it, as in the space program and in the DOD, equipment of extreme technological sophistication. The Government, in purchasing, using, and maintaining this equipment, has often led the way in devising new approaches to mechanical reliability. The second reason involves areas of broad public concern, such as safety. Here, Congress has granted Government agencies the authority to set mandatory standards that are designed to prevent mechanical failures in the private sector. It is clear that in the future both these activities will be continued.

In addition, several current national trends call for enhanced concern with mechanical reliability. These trends are related to the conservation of natural resources, the continuing need for new methods of improving product quality and maintainability in the face of rising labor costs, and the growth of new technologies, such as nuclear energy, requiring a new plateau of operating reliability.

These trends are analyzed in terms of future government policy and future modes of interaction between the government and the private sector.

Key words: Conservation of material resources; failure prevention; government action; mechanical failure; mechanical reliability; safety standards.

1. Introduction

Man is a great user of materials. With them he has created homes, offices, factories, cars, television, in fact, civilization as we know it. Creating, building, using are constants of the human condition; differences are merely of degree. While one culture constructs its dwellings of adobe and another uses steel and concrete, the intent is the same. Take away materials and our grandest plans and projects become meaningless. Take away materials and most of mankind would disappear.

We use enormous amounts of material, and the consumption rate is accelerating. In 1950 we used 1.2 million metric tons of aluminum; in 1970, over 4.5 million metric tons. The 1950 use of nonmetallic building construction materials was 500 million metric tons, contrasted to 1600 million metric tons in 1970. To put it on a more personal basis, about 18 000 kilograms (40,000 pounds) of new materials are used each year for every American. This figure includes 4000 kilograms of gravel and sand, 350 kilograms of cement, 550 kilograms of iron and steel, 7 kilograms of lead, 3500 kilograms of oil, and 2300 kilograms each of natural gas and coal. In terms of products, we used up and threw away last year 100 million tires, 7 million cars and trucks, and 7.6 million TV sets.

Not only is gross materials use rising, but there is a constant demand for new and better materials. For example, the search for energy alternatives is creating the need for high temperature, high strength ceramics for use in higher efficiency turbine generators. Engineering data on such new materials are needed in a hurry, as are reliable ways to fabricate them and evaluate their expected performance in service.

2. Mechanical Failures Do Occur

With the production of millions of products and structures, I don't find it surprising that some are subject to mechanical failure. Almost no system is foolproof. Rather I find it surprising that so many things work so well for so long. This is certainly a credit to American technology. But mechanical failures do occur, and our goal should be minimization of their frequency and consequences.

Mechanical failures are nothing new. The old nursery rhyme "London Bridge is Falling Down" is really a piece of history, celebrating the slow degeneration of that famous span. The cause of the London Bridge troubles is lost in antiquity, but when the Ashtabula Railway Bridge collapsed in 1877, killing 90 persons, the fault was traced to insufficient knowledge of the wrought iron used in its construction. The fall of the Silver Bridge into the Ohio River in 1967, with a death toll of 47, was another case of mechanical failure, initiated by stress corrosion cracking.

There have been other infamous cases of mechanical failure, of course. The breakup of Liberty Ships during World War II, the failure of Comet Aircraft in the 1950's, and the recent rash of gas pipeline ruptures, remind us that we live in a mechanically imperfect world. And as we build larger structures, the potential for disaster is magnified. For example, a new class of barge has been designed to carry 42 thousand cubic metres (11 million gallons) of petroleum products. Last winter one of them broke apart and sank in New York harbor despite having met approved construction specifications. Luckily the tanks were loaded with water at the time, or we might still be cleaning up the waterfront.

3. Consequences of Failure

There are, of course, varied consequences of mechanical failure. Death or disability can be one tragic outcome, causing anguish to the victims or their survivors and often leading to lawsuits for the recovery of damages. The assessment of responsibility and liability is a complex problem, one that I shall expand on later in my remarks.

Economic loss is also associated with mechanical failure, and sometimes this can be a double-action effect. For example, when a machine on the assembly line breaks down, there is not only the cost associated with repairing or replacing it, but also the cost of lost productivity.

Another consequence of mechanical failure is that of materials waste. When an item breaks down and we replace it, a finite amount of irreplaceable raw materials is consumed. This aspect of the problem will be reduced as we improve our recycling ability, but we've got a long way to go in this area. There are other consequences associated with the "fail-discard-replace" cycle, such as the energy used and the pollution created in producing an item, and the problem of disposing of it if the materials are not recycled.

A final consequence of mechanical failure is that of frustration. Almost everybody has experienced the flash of anger when the family chariot burns out a bearing on the Interstate, or the local repair shop tells you the central air conditioner will be out of action all summer because compressors are back-ordered at least 3 months. Not only is this frustrating, but it also leads to disillusionment with the system. When things break down, when repairs are slow, difficult, costly, or impossible, some people turn off from technology. And I can't always blame them.

4. Government Concern

I think we all agree that failures do occur, with consequences ranging from major disaster to minor inconvenience. Why is the Federal Government concerned? There are a number of reasons. First, the Government has certain missions that, because of their nature or sheer size, are clearly its special responsibility. For example, providing for the common defense, as spelled out in the Constitution, is a unique function of the Federal Government. Today this means designing, building, and maintaining the most modern and sophisticated machinery that technology is capable of producing. It also requires some assurance that the machinery will perform efficiently and without failure. The Government also is the prime mover in the fields of atomic energy and space exploration.

The requirements in defense, space, and atomic energy have influenced the science of failure avoidance in the areas of fracture control and nondestructive evaluation. The very high strength materials that have been developed for aerospace construction, and the thick section steels developed for the pressure vessels of the AEC, have in common their susceptibility to the presence of cracks. The need for designing against failure despite inherent flaws gave a great impetus to the science of fracture mechanics, and much was learned about the behavior of brittle materials that perhaps might not have been learned, or certainly would have been learned more slowly. With the added impetus of the troubles with the F 111 and C5A, the Air Force now writes fracture control specifications into the requirements for aircraft. In a parallel effort, pressure vessels for use in nuclear power plants are subjected to a rigorous scrutiny for cracks. The zero defects program of NASA owes much to fracture control methods. These three developments added great impetus to an emerging new science, that of fracture control design.

5. Failure Prevention and NDE

Failure prevention need not be based on the statistical probability of failure alone, but on the concept that the life of a given piece of equipment can be accurately predicted if the use of it can be specified, and if the flaws in it can be determined. Both of these are achievable aims, and I view this development as having far reaching consequences for the science of design not only in equipment for the military, for space, and for nuclear energy, but for the private sector as well. Fracture control design and fracture proof manufacture is an emerging discipline that will assume ever more importance.

A necessary partner to the concept of fracture control design is the location of flaws in a structure through the techniques of non-destructive evaluation (NDE). Beginning with early work on radiography at Watertown Arsenal in the early 1920's, and reaching recently to the elaborate requirements for the location of cracks written into the contractor specifications on the B1 bomber program, the Federal Government has had an intense interest in NDE. This interest was intensified in the post World War II period when high-strength, flaw-sensitive materials came into extensive use. Very small flaws indeed must be detected reliably and accurately, if such materials are to be used with the kind of effectiveness required in the military and space programs of the nation.

Until very recently, NDE has been mainly a qualitative tool, very useful for the detection of major flaws in materials. However, with the advent of a drive toward fracture safe design NDE is moving to a higher level of precision. The reason is simply that fracture control design places quantitative requirements on the detection of flaws, and these new requirements are causing rapid developments in the quantification of NDE.

Despite the advances that have been made, NDE is not yet a precision technique. Consider ultrasonic testing, one of the most popular NDE approaches. No standard is available against which to make meaningful calibrations; phase and frequency data that could greatly increase the information output are ignored; and automation to increase efficiency and reduce operator variability is not used widely. Similar problems are common to other NDE techniques, and a great deal of fundamental work lies ahead if NDE is to become a truly useful, quantitative tool.

6. Health and Welfare

The Government is also concerned with the health and welfare of the people. Congress has created special agencies whose primary concern is public safety, and mechanical failure prevention is a major concern of these agencies. One of the newest is the Consumer Product Safety Commission. Its interests have been ably and thoroughly covered by the previous speaker, my good friend Dr. Kushner, so I will not say anything more about that one.

The Federal Highway Administration, responsible for the correction of highway conditions that contribute to accidents, has had a major program on the detection of cracks in highway bridges. Some of the techniques being investigated for this purpose are derived from those developed for the defense, space, and atomic energy efforts I previously discussed.

The Coast Guard, in carrying out its responsibilities for merchant marine safety, has recently required fracture mechanics-based design for tankers carrying liquified natural gas.

The Federal Aviation Administration has the responsibility of enforcing rules, regulations, and minimum standards relating to the manufacture, operation and maintenance of aircraft. The interest of this agency in failure prevention is well known to all of you, an interest attested to by the fact that they are cosponsors of this Symposium.

This concern for public safety is such that, when the private sector cannot do the job alone, it is appropriate for the Federal Government to join in.

The Government is also concerned with the economic health of the nation, and has legitimate interest in mechanical failures that may cause undue economic dislocation. And, finally, the Government is the world's largest consumer, and like us as individuals, wants its purchases to be safe, economical, and long lasting. In 1972 Uncle Sam purchased over $1.6 billion worth of supplies and equipment for civilian agencies, and $58 billion for the military. Through this huge purchasing activity the Government can do much to influence the reduction of mechanical failures, an area under investigation by the NBS Experimental Technology Incentives Program.

7. Public Policy

In all of its concerns, be they military, health and welfare, or space, Government policy is concerned with mechanical failures. First a few definitions. To define public policy to everyone's satisfaction reminds me of the old story of the blind men and the elephant, in that the answer is different depending on the way in which you approach the subject. Policies provide the basic framework of principles and rules that are used as guides in decision making. According to some experts, policy has two distinct aspects. One aspect considers what is being done, and the other considers how it is achieved. There are many other definitions, but let me talk about how public policy comes into being, and how it is implemented.

The President, of course, helps formulate public policy. President Nixon has done so in his Buyer's Bill of Rights, in which he states that the consumer has a right to safety in the products he buys. This credo obviously has implications for mechanical failure prevention, as there is often a direct tie between failure and injury.

Congress formulates policy through legislative enactments, and many Acts of Congress have impact on mechanical failure. For example, the Natural Gas Pipeline Safety Act is directly concerned with protecting the public from this class of mechanical failure.

Neither the President nor the Congress can carry out policy on a day-to-day basis, and a number of Federal agencies have been established to do so. I mentioned some of them earlier. These agencies, in interpreting and executing national policy, often establish rules, requirements, and procedures that are in effect mini-policies of their own. For example, the FAA has established stringent rules concerning air safety, and every time I fly I'm glad they have.

Now none of this policy setting is done in a vacuum, nor is it a one way street. The President and Congress reflect the will of the people, and respond to the needs of the day. The creation of the Consumer Product Safety Commission was a response to a new wave of consumerism that has grown over the past 10 years. The standards setting efforts of the private sector also have policy impact, for such standards are often used in the requirements of various Government agencies.

8. Product Liability

The courts, too, play a major role in policy setting. They not only interpret formal legislation, but they also provide a forum for the challenge of specific rules set by Federal agencies. And the courts play a major role in their interpretations of common law—by common law I mean that arising out of precedents and not out of legislation—and here, too, they are having broad impact. First a few statistics. The public is turning more and more to the courts in cases of product liability, and in terms of this conference that often means mechanical failure. In 1960 there were 50,000 such suits in this country; in 1970 there were 500,000.

Not only is the absolute number of product liability suits growing astronomically, but there has been a drastic change in court climate regarding liability. A hundred years ago courts held that the manufacturer had no liability except where a contractual relationship existed. Over the years there was a shift to the situation in which the plantiff had to prove negligence on the part of the manufacturer to the situation today in which the plaintiff need only prove that the article was unsafe in some way which caused him injury or damage to win his case.

This shift has potential implications for public policy. Perhaps the notion of no-fault insurance for manufacturers should be considered. The no-fault concept is being applied in the area of automobile insurance, and extension in some form to the area of product liability is worth debating. Certainly it would go a long way towards reducing court backlogs, as a million liability cases per year are foreseen within the next decade.

At a time when sophisticated new engineering applications and shortages of traditional materials put a premium on innovations in materials and on imagination in design concepts, this legal climate may lead to an overly cautious approach by industry. For example, the avoidance of new materials, overdesign, and delays in marketing. This is an appropriate area for new policy —for strong Governmental support of materials characterization research and of research and development of NDE test methods. Advances in these technical areas would benefit the public, the manufacturer and would lessen the load on the courts. There is a pretty good track record in such areas as aerospace, and the time is right for an active extension of NDE to consumer products. This is a big job, an expensive job, and one in which the Government could well be a willing partner. This policy area should also be explored.

9. Other National Concerns

There are other national concerns that have both materials failure and policy implications. Let me cite just a few.

In energy generation, the desire for higher efficiency has led to development of new high temperature turbine blade materials. These are new alloys with special shock sensitivity properties, or new high temperature ceramics, both subject to fracture from mechanical and thermal shock. Reliable assessment techniques are needed for such materials.

In building, steel-fiber reinforced concrete can be used without further reinforcement, but new tests are needed to assess this material's long term strength and durability. Asphalt roofing is subject to mechanical failure as a result of thermal cycling, and reliable tests are needed for proposed replacements.

The use of recycled materials will bring about new testing requirements. For example, the stress-rupture resistance of high strength nickel base superalloys is seriously impaired by parts per million traces of selenium, a common element in scrapped electronics. New, fast, reliable analytical techniques will be needed in this area.

There are mechanical failure problems in the development of synthetic body implants, in the use of plastics as load bearing materials, and in the use of polymer-impregnated concrete. Mechanical failure has a double impact on our foreign trade. First, our productivity must be high if we are to compete successfully in foreign markets. Productivity is in turn linked to new products, new materials, and a minimum of breakdowns in the manufacturing process. Second, the products we sell abroad must also have a reputation for reliability, and that means few mechanical failures.

10. Policy Needs

I said earlier that policy is a many faceted thing. It embodies what we as a nation hope to achieve as well as the courses of action by which we get the job done. If you look hard enough, you'll find that there is national policy on materials and mechanical failure, but it is a diffuse, uncoordinated, stop-gap policy that lacks a real focus. We as a nation need a coherent materials policy. Public safety, new materials, productivity, materials shortages, pollution are all part of a problem that affects all Americans. There are obvious mechanical failure components in each of these categories, and failure prevention should be an integral part of policy formulation.

I'll be specific about the needs. The Government should foster the development and use of reliable NDE techniques so that safe, long-lasting products can be made available at reasonable cost. There will be other benefits. With longer lasting products there will be reduced needs for raw materials, less energy used in the production and transportation of goods, and less pollution at both the production and disposal ends of the cycle.

We need both policy and technology that leads to increased recycling and reuse. We need a better understanding of the relationship between flaws and failure. We need research into the characterization of materials, so that they can be better matched to service requirements. And, finally, there should be a *clear* definition of liability when the rare but unavoidable failures do occur.

Of course, such policies must be developed in terms of the latest and best information. You and I, both as concerned citizens and technical specialists, can and should contribute to the creation of such policies. This is a large assignment, but a necessary one if we are to make the best use of our national resources—materials, energy, technology and people.

11. Discussion

H. Piehler, Carnegie-Mellon University: I would like to react somewhat to your posture toward product liability and its role in this entire mechanical failure milieu. There seems to be a general feeling here that a governmental agency is all that is required in order to alleviate these ills and that the liability suits provide no other function than to goad manufacturers and create backlogs in the courts. Having been involved in a study of the whole litigation process at Carnegie-Mellon with people from the Duquesne Law School for the last 2 years, let me just comment that this vehicle tries to incorporate all of the societal judgements into situations where, in fact, consumers have been injured and have no other form of recourse. If I were given the option of doing away with this system and replacing it with a no-fault system, I would definitely opt for the present system but with refinements from the technical community. One goal of this group ought to be injecting realistic criteria into the present legal system so that strict liability can work the way it really should.

R. W. Roberts: There is merit in what you say. By bringing up a concern such as no-fault you not only examine what is good and bad about that but you also examine what is good and bad about the current system. The major thrust that I was trying to make with respect to what was happening in the courts is that perhaps we can avoid many of these suits if we take the tack that Commissioner Kushner was pushing of trying to build better reliability into products. If we take some of the techniques that have been developed in the aerospace, nuclear, and defense industries and apply these to the consumer industry, we probably will be able to do that. You know there haven't been any reactors that have had major failures. There are very few aircraft that fail. But what's the product failure rate in some of the consumer product lines? I would submit that they are a good deal higher than you would find in the nuclear industry. With respect to cost, and that's really the key, we're going to have to work fairly hard to build NDE techniques into any manufacturing line, but I think unless we try, we're never going to succeed. It's my impression that some of the larger manufacturing corporations in this country want to do that, too, because the profit margin isn't very large on certain items. The first service call that occurs under warranty can wipe out the margin on many, many products, not just the one that's being serviced. There is a real incentive to the corporation to be concerned with product reliability.

LIST of REGISTRANTS

Mechanical Failures Prevention Group

20th Meeting

Mechanical Failure—Definition of the Problem

National Bureau of Standards, Gaithersburg. Maryland

May 8-10, 1974

Frank Almater
Consumer Product Safety Commission
Bethesda, Maryland 20016

E. Ambler
Deputy Director
National Bureau of Standards
Washington, DC 20234

Clifford Astill
National Science Foundation
Washington, DC 20550

Alfred J. Babecki
NASA—Goddard
Code 755
Greenbelt, Maryland 20771

Bruce C. Baird
The Boeing Company
P. O. Box 58747
Houston, Texas 77058

William F. Bangs
NASA–Goddard Code 321
Greenbelt, Maryland 20771

Bruce P. Bardes
Bimba Mfg. Co.
101 E. Main Street
Monee, Illinois 60449

William J. Baxter
General Motors Research
12 Mile & Round Roads
Warren, Michigan 48013

Alan Beerbower
Esso Research & Engineering Co.
P. O. Box 51
Linden, New Jersey 07036

Dennis C. Bely
U.S. Army Ballistic Research Labs
131 Law Street
Aberdeen, Maryland 21001

Russell E. Benner
Lehigh University
Dept. of M.E. & Mech.
Bethlehem, PA 18015

J. A. Bennett
National Bureau of Standards
Metallurgy Division
Washington, DC 20234

Isadore Binder
U.S. Army TACOM
Warren, Michigan 48090

Raymond L. Bisplinghoff
National Science Foundation
1800 G. Street, NW.
Washington, DC 20550

T. Blair
Travelers Insurance Company
1401 Elm Street
Dallas, Texas 75202

Alfred P. Brackney
Monsanto Polymers & Petrochemicals Co.
730 Worcester Street
Indian Orchard, MA 01051

David C. Broom
Alcan International, Limited
P. O. Box 8400
Kingston, Ontario
Canada K7L 4Z4

H. C. Burnett
National Bureau of Standards
Metallurgy Division
Washington, DC 20234

Herbert S. Cheng
Northwestern University
Technological Institute
Evanston, Illinois 60201

John F. Clayton
Gulf Oil of Canada
Sheraton Park
Ontario, Canada

Maurice G. Cockcroft
U.K. National Gas Turbine Establishment
British Embassy
3100 Massachusetts Ave., NW.
Washington, DC 20008

L. A. Colucciello
National Transportation
Safety Board
Washington, DC 20591

W. D. Compton
Director, Ford Scientific Res. Lab.
Ford Motor Company
Dearborn, Michigan 48021

Herbert Corten
University of Illinois
321 Talbot Laboratory
Urbana, Illinois 61801

Raymond A. Coulombe
Naval Ship Engineering Center
Prince Georges Center
Hyattsville, Maryland 20782

James J. Coyle
Villanova University
Villanova, Pennsylvania 19085

Arnold R. Daddario
Pratt & Whitney Airc. Div. UAC
400 Main Street
East Hartford, Connecticut 06108

Paulo Sergio daSilva
Instituto de Pesqisas Techno Pogicas
Sao Paulo Cx Postal F141, Brazil

Thomas L. Daugherty
Naval Ship R&D Center
Annapolis, Maryland 21402

Ann Davis
Naval Ship Res. and Dev. Center
Annapolis, Maryland 21401

Evan J. Deemer, Code ASCR 3370
Naval Air Systems Command
NAS Norfolk
Norfolk, Virginia 23155

John V. Deller
Naval Ship Engineering Center
(Sec. 6148)
3700 East West Highway
Hyattsville, Maryland 20782

Edward J. Devine
NASA Goddard Space Flight Center
Greenbelt, Maryland 20771

Madan Dhawan
Naval Ship Engineering Center
Hyattsville, Maryland 20782

V. Dobroselskiy
USSR Embassy
1125 16th Steet, NW.
Washington, DC 20036

Thomas J. Dolan
Engineering Consultant
510 S. Highland Avenue
Champaign, Illinois 61820

Stanley W. Doroff
Office of Naval Research
Arlington, Virginia 22217

John J. Dougherty
Boeing Vertol Compay
P. O. Box 16858
Philadelphia, Pennsylvania 19142

James F. Dray
Naval Ship Engineering Center
(Code 6148)
3700 East West Highway
Hyattsville, Maryland 20782

D. C. Drucker
College of Engineering
University of Illinois
Urbana, Illinois 61801

Robert J. Dudash
American Telephone & Telegraph Co.
P. O. Box 665
Herndon, Virginia 22070

George Earle
Boeing Vertol Company
P. O. Box 16858
Philadelphia, PA 19142

M. Keith Ellingsworth
Office of Naval Research
800 N. Quincy Street
Arlington, Virginia 22217

Dean F. Elliott
NASA, GSFC
6703 Lamont Drive
Lanham, Maryland 20801

John Erthal
Naval Air Development Center
Warminster, Pennsylvania 18974

James Fei
Naval Ship Engineering Center
Prince George's Center, Code 6107B1
Hyattsville, Maryland 20782

Richard S. Fein
Texaco Inc.
P. O. Box 509
Beacon, New York 12508

Harry Feingold
U.S. Naval Ship R&D Center
11801 Prestwick Road
Potomac, Maryland 20854

Jacob Feld
Consulting Engineer
114 East 32nd Street
New York, New York 11016

Hal Ferris
Hughes Helicopters
Centinella & Teale
Culver City, California 90230

Paul Fleming
National Bureau of Standards
Metallurgy Division
Washington, DC 20234

William R. Fontaine
Naval Ship Res. and Dev. Center
Bethesda, Maryland 20034

J. L. Frarey
Shaker Research Corp.
Northway 10
Executive Park
Ballston Lake, New York 12019

A. C. Gangadharan
Foster Wheeler Corp.
110 S. Orange Avenue
Livingston, New Jersey 07039

C. Gerald Gardner
Southwest Research Institute
P. O. Drawer 28510
San Antonio, Texas 78284

Frank S. Gardner
Office of Naval Research
495 Summer Street
Boston, Massachusetts 02210

Luciano R. Gatto
Metcut Research Association, Inc.
3980 Rosslyn Drive
Cincinnati, Ohio 45209

Wilmer Gaylord
U.S. Postal Service
Research and Eng. Dept.
11711 Parklawn Drive
Rockville, Maryland 20852

B. K. Genetti
General Electric 14–3
2901 East Lake Road
Erie, Pennsylvania 16501

Peter S. Given
SKF Industries, Inc.
1100 First Avenue
King of Prussia, PA 19006

Kay M. Glaesner
St. John's Evangelical Lutheran Church
Wittenberg Avenue at Columbia Street
Springfield, Ohio 45502

Nathan Glassman
Naval Ship R&D Center
Annapolis, Maryland 21402

Irving W. Glater
Pratt and Whitney Aircraft. Div. UAC
Box 611
Middletown, Connecticut 06457

William R. Godwin
Westinghouse Electric Corp.
Box 153 MS 7380
Baltimore, Maryland 21203

W. G. Grenier, Code 755
NASA–Goddard Space Flight Center
Greenbelt, Maryland 20771

Edmund J. Gutowski
Naval Ship Engineering Center
3700 East West Highway
Hyattsville, Maryland 20782

Barry Hammond
Naval Ship R&D Center
Annapolis, Maryland 21401

R. Emerson Harris
NTSB
6513 Byrnes Drive
McLean, Virginia 22101

Keith L. Hawthorne
Association of American Railroads
1920 "L" Street, N.W.
Washington, DC 20036

Donald E. Hayes
Research Center
ARMCO Steel Corporation
Middletown, Ohio 45043

Henry Hegner
IIT Research Institute
10 West 35 Street
Chicago, Illinois 60616

Rolf M. Hellerung
Burns and Roe, Inc.
P. O. Box 663
Paramus, New Jersey 07652

John P. Hirth
Ohio State University
Dept. of Metallurgical Eng.
Columbus, Ohio 43210

Stanley Hlady
Commonwealth Edison Company
Chicago, Illinois 60690

Charles P. Hoffman
NTSB–FOB–10A
800 Independence Avenue, S.W.
Washington, DC 20234

J. D. Hoffman
Institute for Materials Research
National Bureau of Standards
Washington, D.C. 20234

Rudy Hohenberg
AVCO Lycoming Division
550 South Main Street
Stratford, Connecticut 06497

E. Horowitz
Institute for Materials Research
National Bureau of Standards
Washington, DC 20234

Paul L. Howard, Jr.
SKF Industries, Inc.
1100 First Avenue
King of Prussia, Pennsylvania 19406

Jack Hubacher
Value Engineering Laboratory
2550 Huntington Avenue
Alexandria, Virginia 22303

Patrick W. Humphrey
NASA/GSFC
Code 754
Greenbelt, Maryland 20771

Redin G. Jaime (INEN)
Instituto Ecuatoriano de Normalizacion
Ponton 136
Quito, Ecuador

Om Johari
IIT Research Institute
Chicago, Illinois 60616

Carl R. Johnson
NASA–Goddard Space Flight Center
Greenbelt, Maryland 20771

Charles W. Johnson
U.S. Government Printing Office
Washington, DC 20401

R. B. Johnson
National Bureau of Standards
Institute for Materials Research
Washington, DC 20234

Frank C. Jones
K West
9371 Kramer Avenue
Westminster, California 92683

H. Peter Jost
Angel Lodge Laboratories
K. S. Paul Products, Ltd.
Nobel Road
London N18 3DB, England

John M. Karhnak, Jr.
U.S. Army MERDC
Attn: STSFB–HM
Fort Belvoir, Virginia 22060

James Kerley
GSFC—NASA
Glenn Dale Road
Greenbelt, Maryland 20771

Elmer E. Klaus
The Pennsylvania State University
133 Chemical Engineering Building
University Park, Pennsylvania 16802

Joseph D. Kochevar
Chrysler Corporation Defense Engr.
P. O. Box 1316
Detroit, Michigan 48231

Edgar F. Koenig
General Electric
Knolls Atomic Power Lab
Schenectady, New York 12300

William Krewson
Naval Research Laboratory
Code 5360
Washington, DC 20375

J. Kruger
National Bureau of Standards
Metallurgy Division
Washington, DC 20234

Lawrence M. Kushner
Commissioner
Consumer Product Safety Commission
Washington, DC 20016

Donald Laaksonen
U.S. Coast Guard
Research and Development (G–DST–2/62)
400 7th Street, S.W.
Washington, DC 20024

Howard W. Leavenworth, Jr.
U.S. Dept. of the Interior
Bureau of Mines
Interior Building
Washington, DC 20240

R. Lenich
Caterpillar Tractor Co.
100 N.E. Adams Street
Peoria, Illinois 61602

Robert Leon
Vibration Specialty Corporation
100 Geiger Road
Philadelphia, Pennsylvania 19115

Lawrence Leonard
Franklin Institute Res. Labs
20th & Race
Philadelphia, Pennsylvania 19103

Samuel J. Leonardi
Mobil Research & Development
Paulsboro, New Jersey 08066

Frederick F. Ling
Mechanics Division
Rensselaer Polytechnic Institute
Troy, New York 12181

W. E. Littmann
Manager, Metallurgy Research
The TIMKEN Company
Canton, Ohio 44706

Seng Hin Low
Singapore Inst. of Standards
and Industrial Research
18 Hai Sing Road
Singapore 19

Joseph C. Lucas
Sun Oil Company
12 Carriage Lane
Marlton, New Jersey 08053

Jesse B. Lumsden
Ohio State University
Dept. of Metallurgical Eng.
116 W. 19th Avenue
Columbus, Ohio 43210

R. Maringer
Battelle Columbus Labs
505 King Avenue
Columbus, Ohio 43201

R. F. Martin
National Bureau of Standards
Institute for Materials Research
Washington, DC 20234

W. A. Martin
Ontario Hydro Research Division
800 Kipling Avenue
Toronto M8Z 5S4
Ontario, Canada

David K. Matlock
Colorado School of Mines
Dept. of Metallurgical Engineering
Golden, Colorado 80401

A. J. McEvily
Metallurgy Department, U–136
University of Connecticut
Storrs, Connecticut 06105

Michael McKeowan
Atomic Energy Commission
Washington, DC 20545

William J. McNair
Federal Aviation Administration
800 Independence Avenue, SW.
Washington, DC 20591

George McPherson, Jr.
University of Missouri–Rolla
E.E. Department
Rolla, Missouri 65401

Lt. Richard S. Miller, USN
Office of Naval Research
800 N. Quincy Street
Arlington, Virginia 22217

Ray Misialek
Naval Ship Engr. Center
(Code 6762)
Philadelphia Div., Naval Base
Philadelphia, Pennsylvania 19112

Jan K. Miska
Naval Ship Engineering Center
NAVSEC 6158E
Prince George's Center
Hyattsville, Maryland 20782

Richard A. Mollicone
Air Force Materials Laboratory
Wright Patterson AFB
Dayton, Ohio 45433

D. Monson
Travelers Insurance Company
1401 Elm Street
Dallas, Texas 75202

H. E. Morgan
National Bureau of Standards
Admin. Bldg. A1007
Washington, DC 20234

Peter A. Morris, Director
Office of Operations Evaluation
Atomic Energy Commission
Washington, DC 20545

Angeles J. Mortera
Philippines Bureau of Standards
6th Floor Manufacturer's Building
Plaza Station Cruz
Manila, Philippines

R. Mowatt-Larssen
Office of Safety
Federal Railroad Admin., DOT
2100 2nd Street, SW.
Washington, DC 20590

D. Muster
University of Houston
Houston, Texas 77004

Jack R. Nicholas, Jr., Sec 6187
Naval Ship Engineering Center
Prince George's Center
Hyattsville, Marylard 20782

C. Noonan
Naval Ship R&D Center
Bethesda, Maryland 20034

Chester R. Oberg
U.S. Atomic Energy Commission
631 Park Avenue
King of Prussia, Pennsylvania 19406

E. Passaglia
National Bureau of Standards
Metallurgy Division
Washington, DC 20234

H. W. Paxton
Carnegie-Mellon University
Office of the Director for Research
Pittsburgh, Pennsylvania 15213

C. J. Pentlicki
Comsat Labs
Box 115
Clarkesburg, Maryland 20734

Robert Perkins
Naval Ship R&D Center
Bethesda, Maryland 20034

David Perlmutter
Comsat Corporation Labs.
P. O. Box 115
Clarkesburg, Maryland 20734

Marshall Peterson
Wear Sciences Inc.
Scotia, New York 12302

E. E. Pfaffenberger
FMC Corp. Bearing Division
P. O. Box 85
Indianapolis, Indiana 46206

Henry Piehler
Carnegie-Mellon University
Department of Metallurgy and Material Science
Pittsburgh, Pennsylvania 15213

Robert E. Pierce, Jr., Sec 6107B4A
Naval Ship Engineering Center
3700 East West Highway
Hyattsville, Maryland 20782

Bernard L. Poppert
Naval Air Systems Command, AIR34DE
Washington, DC 20361

George W. Post
Reliability Laboratory
Naval Underwater Systems Center
Newport, Rhode Island 02840

S. Blair Poteate, Jr.
US AAMRDL
Ft. Eustis, Virginia 23604

Russell E. Quackenbush
Endevco
Rancho Viejo Road
San Juan Capistrano, California 92675

George Quirk
Monsanto Co.
730 Worchester Street
Indian Orchard, Massachusetts 01051

Curt Reimann
National Bureau of Standards
Office of Standard Reference Data
Washington, DC 20234

Gregg A. Reynolds
Fruit Growers Express Co.
16 Roth Street
Alexandria, Virginia 22314

Richard W. Roberts
Director
National Bureau of Standards
Washington, DC 20234

John Ryan
General Electric Co.
570 Lexington Avenue
New York, New York 10022

Souren Sadjian
Frankford Arsenal
Philadelphia, Pennsylvania 19137

David M. Sanborn
Georgia Institute of Technology
School of Mechanical Engineering
Atlanta, Georgia 30332

Wing Commander W. E. Sansum
Royal Australian Embassy
1601 Massachusetts Avenue
Washington, DC 20036

William T. Sawyer
Holly Harbor
Oxford, Maryland 21654

Thomas G. Schleiter
Atomic Energy Commission
Washington, DC 20545

R. F. Scott, Jr.
Travelers Insurance Company
1401 Elm Street
Dallas, Texas 75202

Peter Senholzi, Code AR–5
Naval Air Engineering Center
Philadelphia, Pennsylvania 19112

Cdr. Richard W. Shafer 6107.02
Naval Ship Engineering Center
Prince George's Center
Hyattsville, Maryland 20782

George R. Shepherd
Mechanical Technology, Inc.
7308 Nevis Road
Bethesda, Maryldan 20034

John J. Sherlock
MALC Div.–Pure Carbon Co.
441 Hall Avenue
St. Mary's, Pennsylvania 15857

T. R. Shives
National Bureau of Standards
Metallurgy Division
Washington, DC 20234

J. A. Smik
Superior Tube Company
Box 191
Norristown, Pennsylvania 19404

Charles O. Smith
College of Engineering
University of Detroit
4001 W. McNichols
Detroit, Michigan 48221

Kempton A. Smith
General Electric Co.
1 River Road
Schenectady, New York 12301

R. W. Staehle
Ohio State University
116 West 19th Avenue
Columbus, Ohio 43210

Charles L. Staugaitis
Goddard Space Flight Center
Glen Dale Road
Greenbelt, Maryland 20771

Robert H. Steele
Atomic Energy Commission
5520 Collander Drive
Springfield, Virginia 22151

Jesse E. Stern
Goddard Space Flight Center
NASA—Code 325
Greenbelt, Maryland 20771

William C Stitzel
Research Center
ARMCO Steel Corporation
703 Curtis Street
Middletown, Ohio 45043

James J. Stork
Chrysler Corporation Defense Engr.
P. O. Box 1316
Detroit, Michigan 48231

Patricia Strandell
Naval Ship Res. and Dev. Center
Annapolis, Maryland 21402

J. R. Strang
AFML/MBT
Wright-Patterson AFB
Dayton, Ohio 45433

Thomas DeW. Styles
National Transportation Safety Board
800 Independence Avenue, S. W.
Washington, DC 20591

John Sudey, Jr.
Goddard Space Flight Center, Code 721
Greenbelt, Maryland 20771

David M. Thomson
Columbia Gas Systems Service Corp.
1600 Dublin Road
Columbus, Ohio 43220

Robb Thomson
Federal Energy Office
Eenergy Research & Development
Room 145—EOB
Washington, DC 20500

A. R. Timmins
Goddard Space Flight Center
Greenbelt, Maryland 20771

James W. Titus
Naval Research Lab
Code 5360
Washington, DC 20375

Henry G. Tobin
IIT Research Inst.
10 W. 35 Street
Chicago, Illinois 60616

J. Torgesen
National Bureau of Standards
Institute for Materials Research
Washington, DC 20234

Alan F. Tregenza
Alcan Canada Products Ltd.
Box 2000
Kingston, Ontario
Canada K7L 4Z5

Emil Troc
Curtiss Wright Corp.
Woodridge Facility
1 Rotary Drive
Woodridge, New Jersey 07075

Gilbert M. Ugiansky
National Bureau of Standards
Metallurgy Division
Washington, DC 20234

F. J. Uhlig
GMSRD C
Mfg. Dev. Bldg. B
GM Technical Center
Warren, Michigan 48090

George Uriano
National Bureau of Standards
Institute for Materials Research
Washington, DC 20234

Frans M. VanGelder
Naval Air Prop Test Center
Box 176
Trenton, New Jersey 08628

William C. Vesser
TEKTRONIX, Inc.
P. O. Box 500
Beaverton, Oregon 97005

C. E. Vest
NASA Code 755
Greenbelt, Maryland 20771

Henry H. Wakeland, Director
Bureau of Surface Trans. Board
Dept. of Transportation
Washington, DC 20591

David Wallace
Black & Decker Manufacturing Co.
3616 Forest Garden Ave
Baltimore, Maryland 21206

John M. Ward
Naval Air Systems Command
Washington, DC 20361

Vernon C. Westcott
Trans-Sonics, Inc.
P. O. Box 326
Lexington, Massachusetts 02173

R. M. Whittier
Endevco
Rancho Viejo Road
San Juan Capistrano, Calif. 92675

R. W. Widner
The TIMKEN Co.
Canton, Ohio 44706

Richard A. Wilde
Eaton Corp.
26201 Northwestern Highway
Southfield, Michigan 48076

M. Wildmann
Ampex Corporation
401 Broadway MS 3–22
Redwood City, California 94063

W. A. Willard
National Bureau of Standards
Metallurgy Division
Washington, DC 20234

Ward O. Winer
Georgia Institute of Technology
School of Mechanical Engineering
Atlanta, Georgia 30332

Victoria C. Wittenbreder
U.S.A. Maintenance Management Center
Lexington, Kentucky 40507

Alan K. Wolff
Massachusetts Materials Research
55 Millbrook Street
Worcester, Massachusetts 01606

☆ U.S. GOVERNMENT PRINTING OFFICE: 1975 O—588-797

NBS-114A (REV. 7-73)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBS SP-423	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE Mechanical Failure: Definition of the Problem MFPG 20th Meeting	5. Publication Date March 1976 6. Performing Organization Code
7. AUTHOR(S) T. R. Shives and W. A. Willard, Editors	8. Performing Organ. Report No.
9. PERFORMING ORGANIZATION NAME AND ADDRESS NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, D.C. 20234	10. Project/Task/Work Unit No. 11. Contract/Grant No.
12. Sponsoring Organization Name and Complete Address (Street, City, State, ZIP) Office of Naval Research, Arlington, VA 22217; Consumer Product Safety Commission, Wash., DC 20207; Frankford Arsenal, U.S. Army, Phil, PA 19137; Naval Air Dev. Center, Wash., DC 20362; Am Soc. Mech. Eng., NY, NY 10017; FAA, Wash., DC 20591; NASA, Greenbelt, MD 20771	13. Type of Report & Period Covered Final 14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 76-600003

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

These Proceedings consist of 20 submitted papers and discussionas from the 20th Meeting of the Mechanical Failures Prevention Group held at the National Bureau of Standards, May 8-10, 1974. The central theme of the proceedings is the definition of the problem of mechanical failure, with emphasis on modes of failure, consequences of mechanical failure, and implications of mechanical failure.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)*

Brittle fracture; economics of failure prevention; fatigue failure; mechanical failure; mechanical reliability; stress corrosion cracking.

18. AVAILABILITY	19. SECURITY CLASS (THIS REPORT)	21. NO. OF PAGES
[X] Unlimited	UNCLASSIFIED	228
[ ] For Official Distribution. Do Not Release to NTIS	20. SECURITY CLASS (THIS PAGE)	22. Price
[X] Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13.10:423	UNCLASSIFIED	$ 6.10
[ ] Order From National Technical Information Service (NTIS) Springfield, Virginia 22151		

USCOMM-DC 29042-P74