THE RESIDUAL STRESS EXPLANATION OF FATIGUE FAILURES IN MECHANICAL PARTS

by

JOHN O. ALMEN
Consulting Engineer
Life Member of Committee E-9
Sierra Madre, California

PAUL H. BLACK, Professor
of Mechanical Engineering
Ohio University
Athens, Ohio

THE RESIDUAL STRESS EXPLANATION OF FATIGUE FAILURES IN MECHANICAL PARTS

ABSTRACT

It has become clearly evident that older theories of fatigue failures are fallacious, incomplete, and misleading. These have been largely replaced by theories based on new concepts which are amply verified by experimental evidence and which provide adequate explanation and prediction of fatigue failures. The purpose of this paper is to contribute to the explanation of these modern and basic concepts in the interest of improved design, particularly for mechanical parts where low weight and low cost are important, and in general to contribute to the science of mechanical design.

1. INTRODUCTION. The fatigue strength of metal specimens is dominated by the fatigue strength of their surfaces. It is known that surfaces are much weaker in repeated loading than sound underlying material. This surface weakness is not yet completely understood, but we may suspect that it is due to low surface elastic limit. However, our immediate problem is to strengthen the surface metal or, somehow, to reduce the stress that is induced at the surface as compared with the stress that is induced in the stronger sub-surface material. A step in this direction is to remove superficial stress concentration due to notches, scratches, corrosion, etc. by fine grinding or by polishing, but this affects only a small part of the surface weakness. How then to proceed? How can an anti-stress be introduced?

If fatigue cracks are propagated by all types of stresses: tension, compression, and shear, as is sometimes believed, the prospects of improving fatigue strength are dim indeed. However, if fatigue cracks are propagated by only one of these types of stress, a benefit can be achieved by superimposing on the surface a stress of the opposite sign and to the depth of the weak surface layer. There are many proofs that show conclusively that fatigue cracks cannot be propagated by repeated compressive stresses or by shear. This statement can be confirmed readily by a few simple experiments. Since as will be shown, fatigue fractures result only from tensile stresses, it is possible to reduce or remove tensile stresses in localized areas by inducing in the critical areas local compressive stresses.

By these means we can and, in effect, have “removed the weak surface layer” and we find that the fatigue strength can thereby be greatly increased even though the thickness of the weak surface layer in metals appears to be less than 0.005 inches.

However, we must recognize that a protective layer of residual compressive stress can easily be lost or reversed, or its magnitude can be greatly reduced depending on the character of the loading. If the applied load stresses the material above its compressive elastic limit, specifically, if the sum of the residual compressive stress and the applied compressive stress exceeds the dynamic elastic limit, the residual compressive stress will be lost, not all in one cycle but a small amount in each load application. Such cumulative reduction in the elastic limit will continue until, in a completely reversed fatigue test, no residual stress remains at the fatigue endurance limit and the sought-for increase in fatigue strength will be lost.

The effect of plastic yielding, however, from applied stresses that are greater than the dynamic elastic limit can be beneficial as well as harmful. If the stress range is such that plastic yielding occurs not in compression but in tension, the resulting residual stress will be compressive. This will increase the fatigue strength as demonstrated by the well-known Goodman effect.

A specimen that is subjected to unidirectional compressive loads, or to reversed loads in which the tensile stress component is well below the endurance limit, will develop residual tensile stress. This occurs when the compressive loads exceed the dynamic elastic limit. Compressive plastic yielding can develop residual tensile stress due to elastic recovery of elastically stressed areas of sufficient magnitude that fatigue cracks may form. Such cracks often grow to considerable depths but the rate of growth becomes very slow or completely stops as the residual tension decreases with depth.

With larger applied compressive loads, the depth of the resulting residual tensile stress increases until complete fatigue fracture occurs.

Unless the magnitude and sign of residual stresses are known and accounted for, we cannot know the magnitude of any stress, including stress range. Residual stresses are always present. We must, therefore, speak of <u>nominal</u> stresses when the effects of residual stresses are unknown.

2. FATIGUE STRENGTH OF METAL SPECIMENS IS DOMINATED BY FATIGUE STRENGTH OF SURFACES BECAUSE SURFACES ARE WEAKER THAN UNDERLYING MATERIAL. Many engineers, and metallurgists, believe that the increased fatigue strength resulting from cold-working operations such as shot peening, superficial rolling, etc. is due to "work hardening". It has been shown[1], * that the hardness of steel as measured by indenters appears to be greater when stressed in compression and softer when stressed in tension. The apparent change in hardness resulting from cold working operation indicates that hardness is not a property of a material or a specimen; rather, hardness is a state of stress and the real result of cold working is residual stress.

---

*Superscripts pertain to the list of references at the end of this paper.

Since this question is basic to understanding fatigue in metal, it became necessary to substantiate the statement that “work hardening” or, as sometimes called, “strain hardening”, is an illusion. An experiment was devised by which this is accomplished. (See pages 124-127 of Reference 1.) Briefly, the method was to conduct a series of fatigue tests on several groups of identical specimens, all of which were equally cold worked. Although equally cold worked, the induced surface residual stress for each group would range from high magnitude compressive stress into the range of tensile stress.

The resulting SN curves would show whether or not “work hardening” is effective in changing the fatigue strength of the specimens. If all of the specimen groups were equally strong they would show that “work hardening” is effective. However, if the specimen groups should vary in strength, particularly if the strength variation is of the same order or magnitude as the residual stress, the fatigue strength change must be caused by the residual stress and the concept of work or strain hardening must be abandoned.

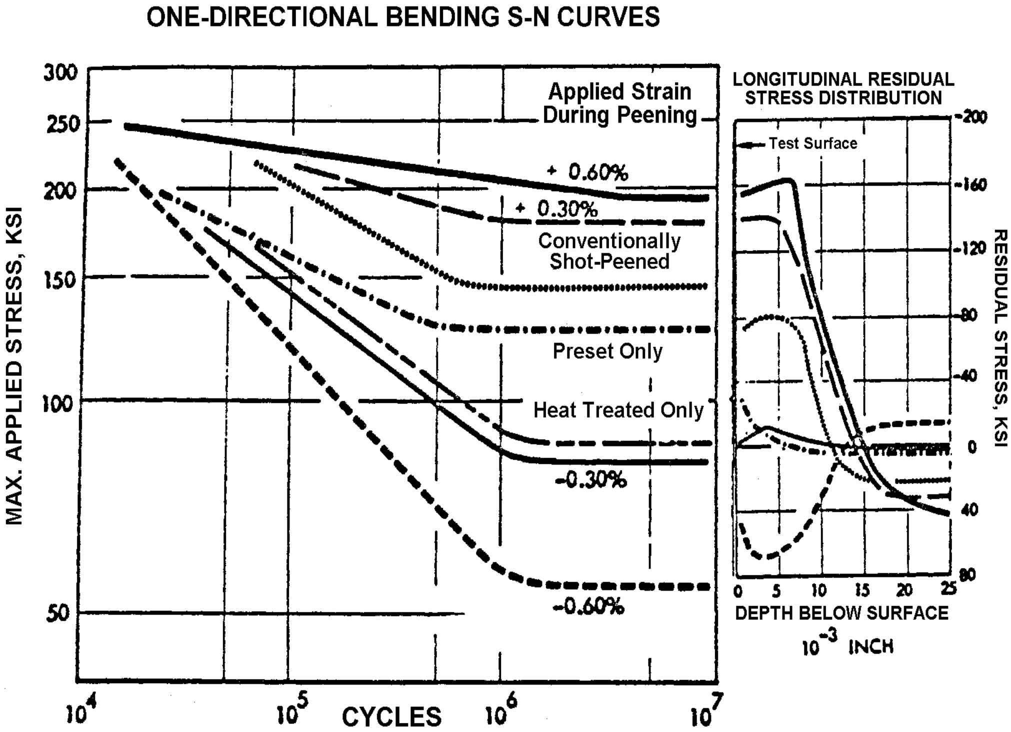
Some years after the proposal to conduct these experiments was suggested, the Society of Automotive Engineers, ISTC Committee, Division 4, agreed to sponsor the experiment whereupon General Motors Research Laboratories offered to conduct the experiments in substantial accordance with the outline stated above.

This investigation was supervised by R. L. Mattson and J. G. Roberts of G. M. Research with results reported to Division 4 in 1958, also by Mattson and Roberts on pages 336-360 of Reference 2, and in Reference 3. In all these publications, interest centers on seven SN curves shown in Figure 1. Approximately, as expected, these show fatigue limits ranging from 55,000 to 194,000 psi for specimens which were identical except for the magnitude of the corresponding residual stresses which are seen to vary from 60,000 psi in tension to 194,000 psi in compression.

Clearly, these tests show no evidence of “strain hardening” but they do show that very thin layers of residual stress can be used to increase or decrease fatigue strength.

Results of the paper by Mattson and Roberts include the following as indicated by the curves in Figure 1:

(a) Beginning with curve No. 5 from the top and labeled “Heat Treated Only”, for the specimen without any presetting or peening, it is seen that its endurance limit is 85,000 psi. The surface residual stress for this specimen is small; about 5000 psi in tension.



This family of SN curves and corresponding residual stress patterns show the basis for many modern theories on fatigue in metals. Among these are: (1) strain hardening is an illusion; (2) fatigue strength increases directly as surface residual compressive stress; (3) surface residual tensile stress reduces fatigue strength; (4) In fatigue the surface is only 1/2 as strong as subsurface metal.

Figure 1.

(b) The residual stresses in all of the specimens to a depth of 0.025 in. are shown in the diagram to the right. It is noted that the relative endurance limits for specimens 1, 2, 3, and 4 are increased by the peening and presetting operations in almost direct proportion to the residual compressive surface stress. The relative endurance limit for specimen 7 is <u>decreased</u> by applied strain during peening which leaves the surface residually stressed in tension.

(c) The thicknesses of the weak surface layer do not exceed 0.007 in. as shown in the diagram at the right.

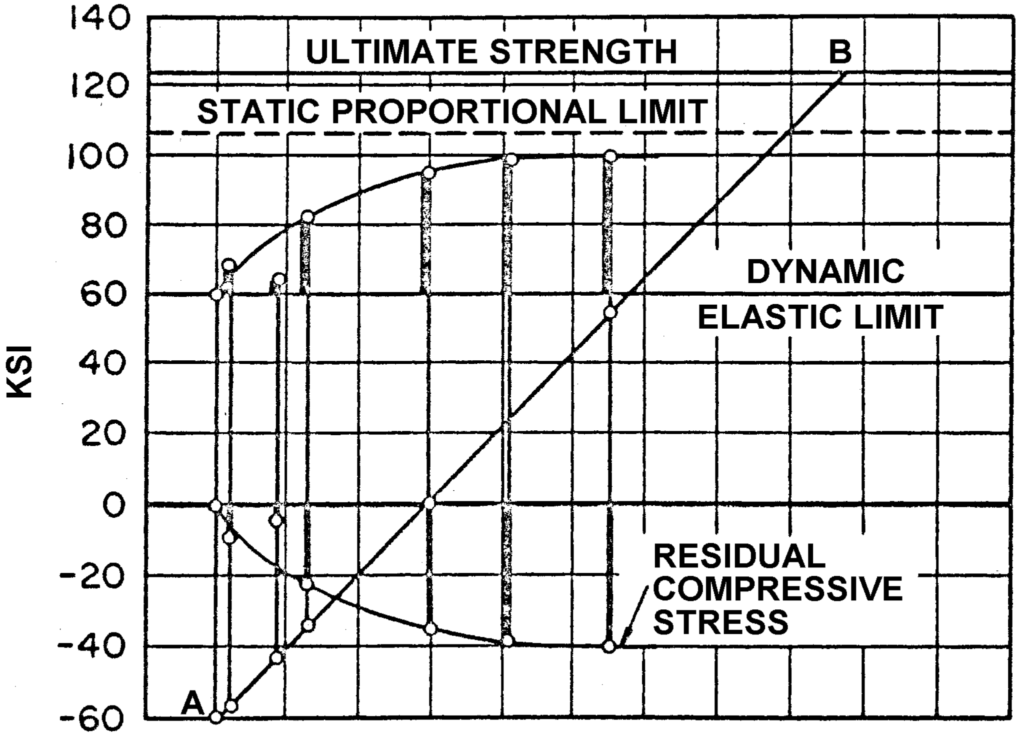
(d) These demonstrated effects of residual stresses on endurance limits are possibly only because of the fact that fatigue cracks grow only from tensile stresses. They cannot propagate in compressively stressed metal.

3. ELASTIC LIMIT IN REPEATED LOADING. The first measurements of elastic limits under repeated loads were made by J. Bauschinger and were reported in 1886.[4] Additional experiments were conducted by Leonard Bairstow II and reported in 1909.[5] The conclusion was

that if a specimen of mild steel was stretched beyond its elastic limit in tension, the elastic limit was increased in tension but was decreased in compression. This characteristic, that stretching a metal in one direction improves its mechanical properties in that direction but lowers them in the opposite direction, becomes important in members subjected to reversal of stress.

Expressed in different words: residual stresses will always develop from plastic deformation, unless the plastic flow is uniform throughout the loaded member, an almost impossible occurrence. Wherever plastic flow occurs from tensile stressing, residual compressive stresses are developed, and when compressive plastic flow occurs, residual tensile stresses appear. Therefore, in unidirectional bending, torsion or pull-pull fatigue tests, limited plastic yielding increases the fatigue strength of specimens not otherwise prestressed. However, plastic flow in reversed fatigue loading will decrease the fatigue strength of the part. See Ref. 6, pages 12 to 17.

4. THE GOODMAN EFFECT. This is a manifestation of residual compressive stress that develops in the weak surface of a specimen when the applied tensile stress exceeds the



THE GOODMAN EFFECT IS A RESULT OF A RESIDUAL COMPRESSIVE STRESS THAT DEVELOPS IN THE WEAK SURFACE LAYER OF THE SPECIMEN WHEN THE APPLIED TENSILE STRESS EXCEEDS THE DYNAMIC ELASTIC LIMIT.

Figure 2.

elastic limit of the metal. The Goodman diagram, and its several modifications, shows that the fatigue endurance limits that are obtained from a series of tests of identical specimens under various stress ranges. As shown in Figure 2, these begin with completely reversed stresses, i.e., the specimens are alternately stressed in tension and compression of equal magnitudes. In successive tests, the stress range is altered by reducing the applied compressive stress in several steps until, in the fifth step in the figure, the applied stress is unidirectional tension. In the two remaining tests, the low limit of the stress is fixed by static tensile stresses of approximately 20,000 and 50,000 psi. In simple terms, the low stress limits of all seven tests lie on the diagonal line A-B. The upper limits are fixed by the fatigue endurance limit. It is seen that the fatigue limit increases with each reduction of the applied compressive stress and with each increase of the static tensile stress.

The stress at the fatigue endurance limit is slightly greater than the dynamic elastic limit. We may, therefore, tentatively accept the endurance limit as measured by completely reversed loads, i.e., 60,000 psi, as the dynamic elastic limit of the steel that was used in these tests. Note that the endurance limits for all tests, except the first, exceed the dynamic elastic limit. In explanation, let us examine the completely reversed test. In each half cycle, the specimen is stressed above the dynamic elastic limit by a minute amount. The minute plastic yielding that occurs develops residual stress of opposite sign to the stress causing the plastic yielding. The following half cycle also exceeds the dynamic elastic limit, and corresponding plastic yielding occurs but of opposite sign to the first. The net result is that each half cycle cancels the residual stresses developed by the other half. See Reference 6, pp. 13-17.

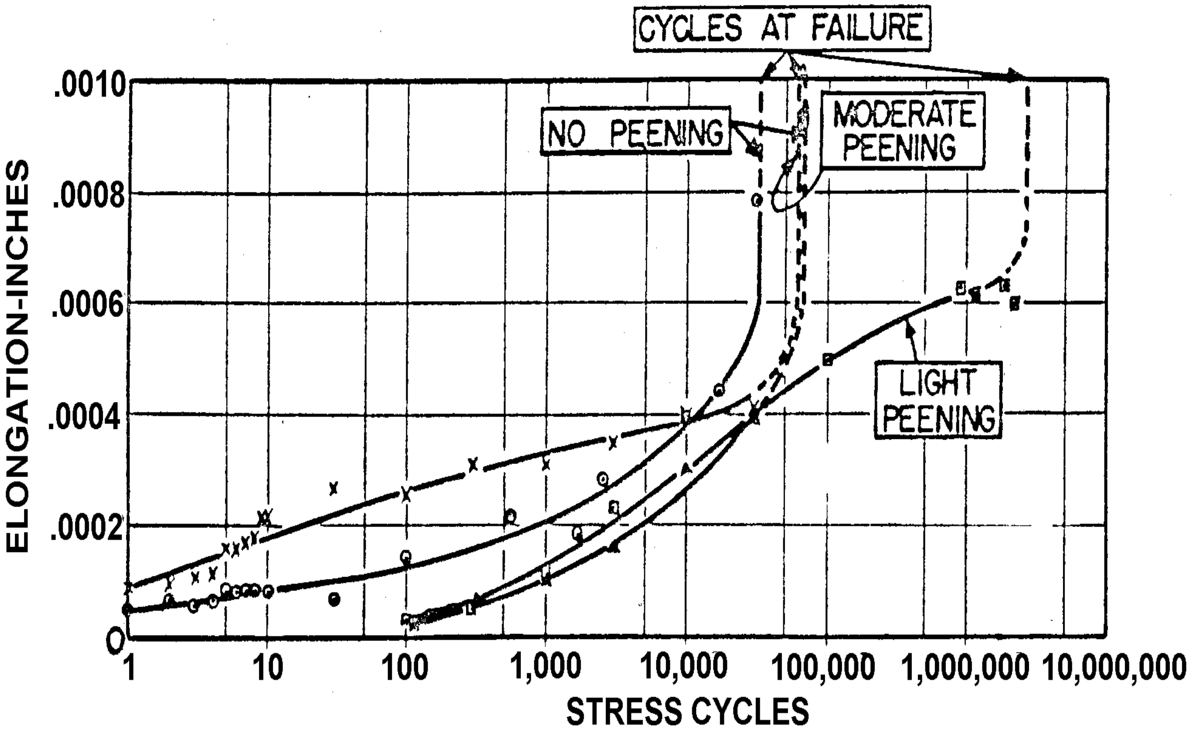
From the above discussion, we learn again that processing operations as shot peening, cannot be beneficial in specimens or parts that are subjected to completely reversed loading because any residual stress initially present will be quickly removed by plastic yielding. The sum of the applied compressive stress and the residual stress from shot peening exceeds the compressive dynamic limit by an amount approximately equal to the residual stress. See Reference 6, p. 80.

Note now that the stress range in the second group shown in the diagram is slight “biased”. The magnitude of the applied compressive stress has been slightly reduced. Therefore, any tensile plastic yielding and the resulting residual compressive stress that occurs in the first half cycle are not erased in the second half cycle. The accumulation of minute compressive stresses continues through many cycles until the total is equal to the reduction of applied compressive stress at which time cancellation prevents further build-up. The fatigue endurance limit, of course, is increased by an amount equal to the accumulated residual compressive stress as is indicated by the heavy vertical line below the zero stress line.

In the succeeding groups of specimens (3, 4, and 5), the applied compressive stress is further reduced and the corresponding residual compressive stresses are increased. The reduction in residual stresses occurs in groups 6 and 7, the only difference being that the static tensile loads are applied to change the stress ranges. In each case, the accumulated residual compressive stress is shown in heavy vertical lines below the zero-stress line. Their magnitudes are, of course, the inverse of the Goodman curve at the upper part of the figure. The residual stress, shown by vertical heavy lines, for any group, can appear in this location, only when all other stresses in the group equal zero. When the maximum load in any group is applied, the residual stress is above the dynamic elastic limit. This means that the weak surface metal is not required to carry more than 60,000 psi in tension; however, the stronger underlying metal is stressed up to 100,000 psi.

To summarize, we find that:

(1) The Goodman effect is the result of residual compressive stress that develops from tensile plastic yielding of the surface metal when the applied tensile stress exceeds the dynamic elastic limit.



YIELD OF PULL-PULL FATIGUE SPECIMENS

Figure 3.

(2) Plastic yielding is continuous, but at decreased rates, throughout the fatigue tests. See Figure 3.

(3) The elastic limit under repeated loads, herein called the dynamic elastic limit, occurs at much lower stress than the elastic limit that is measured by static loading.

(4) The fatigue endurance limit is approximately but slightly greater than the dynamic elastic limit.

(5) The source of the residual stresses that causes the Goodman effect is also the source of (a) changes in stress range; (b) reduction (sometimes called "relaxation") of magnitude of residual stress; (c) increase in magnitude of residual stress; (d) reversal of the residual stress from compression to tension, or from tension to compression. Among the conditions when the latter, i.e., (d), refers to fatigue failures resulting from compressive loads. This is often mistakenly construed to indicate that fatigue fractures progress in regions of compressive stress. However, fatigue fractures are always tensile fractures.

REFERENCES

1. Almen, J. O.: Fatigue Weakness of Surfaces, Prod. Eng., vol. 21, no. 11, pp. 118-140, 1950.

2. Rasweiler, G. M., and W. L. Grube, ed.: Internal Stresses and Fatigue in Metals, Elsevier Publishing Co., New York, N.Y., 1959.

3. S.A.E. Journal, August, 1962.

4. Bauschinger, Johann: Mittheilungen aus dem Mechanisch-technischen Laboratorium der k. Technischen Hochschule in Munchen, Heft XIII, 1886.

5. Bairstow, Leonard, II: The Elastic Limits of Iron and Steel under Cyclical Variations of Stress, Phil, Trans, Roy, Soc, London, Ser, A, vol. 210, 1911.

6. Almen, J. O., and P. H. Black: Residual Stresses and Fatigue in Metals, McGraw-Hill Book Co., New York, N.Y., 1963.

FATIGUE FAILURES ARE TENSILE FAILUES

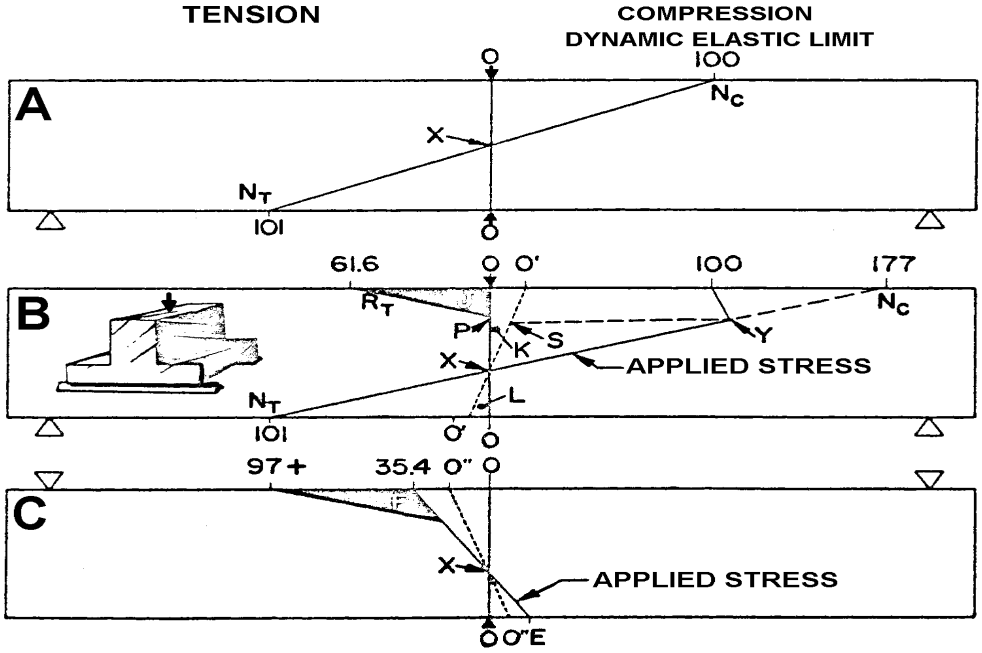
The most conclusive way to prove that fatigue failures resulting from repeated compressive loads are actually tensile fatigue failures is to measure the residual stress in the specimen after a few thousand load applications but before cracks appear. This will not only show that the stresses causing failure are residual tension but will also show the approximate magnitude of the residual stress. However, a close approach to the results obtained by this complicated residual stress-measuring method can be had without leaving your desk. All that is necessary is to construct a stress diagram based upon the methods and stresses that have been used in an actual compressive fatigue test.

Attached diagrams B and C are based on a paper presented before the 1946 summer meeting of this committee by R. S. Jensen and H. F. Moore, both from the University of Illinois. (See ASTM Proceedings, Vol. 46, pp. 799-813. Also see Almen and Black book, page 165, fig. 13-7.) The 1946 Jensen-Moore paper is preferred to other similar papers because it was given wide publicity, being published at least four times, and because it contains enough detailed information regarding the specimen dimensions, manner of loading and applied nominal stresses to permit the construction of accurate diagrams.

A preliminary, completely reversed, fatigue test of the steel used in all subsequent tests showed that its fatigue limit was 33,400 psi. Since the fatigue limit under completely reversed stresses is also (very nearly) the dynamic elastic limit of the steel being tested, 33,400 will hereafter be represented by 100 stress units and all other stresses will be indicated in percent of 100.

The inverted T-shaped specimens, a sketch of which is inserted in diagram B, were dimensioned so that their neutral planes were 2/3rds of their vertical thickness below the top surfaces. The purpose of this design was to divide the stresses from applied bending loads between the top surface and the bottom surface in the ratio 177 compression to 57 percent of 177 in tension, this being the relative stress magnitude that had been measured in rails in actual service. The applied bending loads were partially reversed so that the upward load was 20 percent of the downward load. The three attached diagrams show what actually happened under the above applied stresses.

DIAGRAM SHOWING THAT FAILURE DUE TO NOMINAL COMPRESSIVE STRESS IS ACTUALLY TENSILE STRESS



The introductory diagram A shows a simple, rectangular (not a T section) beam that is supported near its ends. A downward-acting load is applied on its top surface at 0. The resulting stress in the beam is indicated by the diagonal line N_tN_c. The stress at any depth in the beam is indicated by the diagonal line $N_tN_{c.}$ The stress at any depth in the bean is measured by the horizontal distance from the vertical line 0-0 to the diagonal line N_tN_c. The stress on the left side is tension. The stress on the right is compression. When the load is removed the diagonal line N_tN_c rotates on the intersection point X until it coincides with the vertical line 0-0 at zero load. In rotating, the points N_t and N_c do not follow arcs of circles. Instead, they follow the top and bottom surfaces.

Diagram B illustrates the effects of downward-acting loads on the Jensen-Moore T-shaped specimens. The nominal stresses from the applied loads are indicated by the distance N_t0 on the bottom surface, which represents 101 units of surface compressive stress.

Since the dynamic elastic limit is 100 units, 77 units of top surface compressive stress are lost by plastic yielding, but they reappear on the top surface as the residual tensile stress R_t0 when the downward-acting load is removed. The portion Y N_c of the nominal compressive stress reappears as the line R_tS.

As in the case of diagram A, the removal of the downward-acting load in diagram B, the stress line N_tC_C rotates on the intersection point X but, unlike A, the rotation is stopped at the line 0'0' before it reaches the vertical line 0-0. This is a result of the plastic compressive yielding. The new position 0'0' is a measure of the permanent set of the beam due to plastic yielding. The line cannot rotate further because it has reached the point of balance (estimated) of the clockwise moments which are measured by the triangles K and L and the counter-clockwise moment measured by the triangle J about the point X. In the now unloaded beam, we have tensile residual stress of 61.6 units (20,574 psi) on the top surface, which is indicated by the distance R_tO.

Diagram C differs from diagram B in that it shows the stress line E-F from the upward-acting load in the second half-cycle. This stress has the effect of straightening the beam until the line 0'-0' of diagram B becomes the line 0"-0" in diagram C. The new tensile stress, indicated by the surface distance FO, is equal to 35.4 (20% of 177) stress units. Since the upward-acting load coincides in time with the greatest residual tensile stress on the same surface from the downward load of the first half cycle, the total tensile stress is now 97 units or 32,400 psi. Additional residual tensile stress, of unknown magnitude, was no doubt developed from adjustments of the crank throw. I quote from the report:

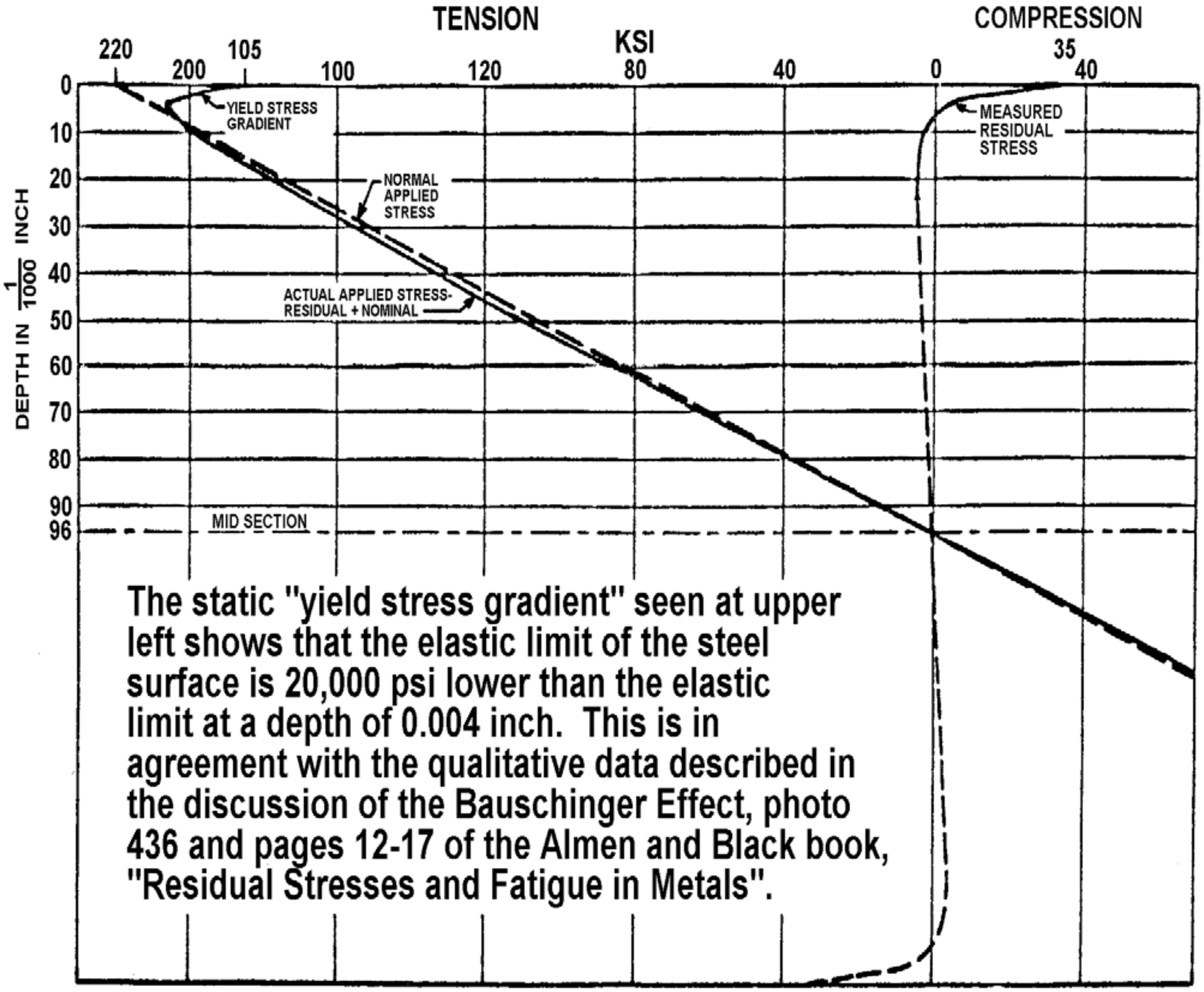
> "During all tests the constant ration of tensile stress to minimum compressive stress was maintained at –0.20 by stopping the machine at intervals, taking readings of deflections of the specimen and, if necessary, adjusting the throw of the crank."

From the above, we learn that compression fatigue failures are tensile failures. Whenever pseudo compression fatigue failures occur, we must find the source of the tensile stress that was the real culprit. (See Almen and Black book, page 177, Art. 14-6).

Often residual tensile stresses from compressive loads alone are great enough to develop fatigue cracks. Several such cracks are shown in the Almen-Black book. Fig 8-2 on page 93 shows a coil spring in which more than twenty fatigue cracks appear, none of which penetrated more than 1/3rd of the diameter of the rod from which the spring was coiled and none prevented near normal functioning of the spring. The reason for the limited depth of fatigue cracks is seen in diagram B. If the residual tensile stress shown in this diagram had been great enough to initiate a crack or cracks, none could have penetrated to the depth of the residual tensile stress, i.e. the depth O-P. The crack would have stopped when the remaining tensile stress became too low to continue crack growth. In the 1946 Jensen-Moore experiment the crack grew to greater depth because there remained enough tensile stress from the upward-acting load in the second half-cycle to continue crack propagation at a slow rate. Again I quote from report:

> "Not only is the progress of the crack slow in a region of high compression but especially in the case of a T-shaped specimen it may, if not entirely, cease to spread when it reaches the upper edge of the flange . . ." and "Several of the specimens tested developed more than one crack . . ."

These observations are typical of cracks that propagate in fields of diminishing residual tensile stress.



The static "yield stress gradient" seen at upper left shows that the elastic limit of the steel surface is 20,000 psi lower than the elastic limit at a depth of 0.004 inch. This is in agreement with the qualitative data described in the discussion of the Bauschinger Effect, photo 436 and pages 12-17 of the Almen and Black book, "Residual Stresses and Fatigue in Metals".

The diagram on the opposite side of this paper was constructed from data that appears in the book “INTERNAL STRESSES AND FATIGUE IN METALS” edited by Gerald Rassweiler and William Grube, published by ELSEVIER PUBLISHING CO., distributed by D. VAN NOSTRAND COMPANY, INC., pages 337-357.

The specimen used was one of the group from which the S-N curve, labeled “Preset Only”, seen in Figs, 7, 8, and 10, was constructed. These figures also show the residual stress that resulted from the presetting operation, which was accomplished by bending the specimen between two curved blocks to a nominal surface stress of 220 ksi.

The elastic limit curve is the algebraic sum of the nominal stress (dashed line) and the residual stress. See Almen and Black book, pages 12 to 17.