THE UNIVERSITY OF MICHIGAN

WEAR CONSIDERATIONS IN DESIGN

SCORING, SCUFFING, ABRASION, PITTING, FRETTAGE, CAVITATION, CORROSION

CHARLES LIPSON, Ph.D.
Professor, Mechanical Engineering
University of Michigan

Preface

This book is an outgrowth of a series of lectures given during a course "Wear Considerations in Design" and a symposium "Engineering Approach to Surface Damage," conducted at the University of Michigan, Department of Mechanical Engineering. The considerable work and assistance of the following students is acknowledged: A. H. Krafve -Compatability of Metal Pairs; L. Cyr and R. D. Kuieck-Solid Surface Films; A. Jocz-Friction and Wear; R. Sigsbee-Sliding at High Temperature; E. A. Tuncel-Theory of Contact Loading; C. T. Fox-Surface Endurance of Materials; T. R. Forrester and R. J. Tchon-Effect of Inclusions; W. D. Chadwick and D. L. Otto-Anti-friction Bearings; J. M.Leinonen-Gears; D. L. DuMond and L. D. Mitchell-Galvanic Corrosion; D. Laviolette-Cavitation Corrosion; M. F. De Vries, R. D. Brown and B. T. Veldman-Frettage Corrosion; D. La Valley and J. E. Larson-High Temperature Corrosion; H. Totteu-Stress Corrosion Cracking; L. Cyr-Abrasion; M.Allen, G. A. Mitchell and J. N. Karawalla-Wear Resistant Materials; R. M. Cadell and D. Adamski-Residual Stresses. I am also indebted to B. W. Kelly of Caterpillar Tractor Company for his work on Surface Temperature as presented in Chapter 8. Mr. Leonard Cyr, a graduate student at the University was very helpful and effective in assisting me in editing this book.

TABLE OF CONTENTS

CHAPTER 1

CRITERIA FOR SCORING

1. THE MECHANISM OF SCORING AND SEIZURE

The mating surfaces of two machine parts, when considered on an atomic scale, are not smooth. Instead, it will be found that the surfaces consist of hills and valleys. (Figure 1) These peaks, or asperities, will cause the actual mating surfaces to be closer together than a consideration of the mean surface levels would indicate.

In the case of boundary lubrication, the contaminating films often break down, allowing the asperities to contact each other. Therefore, it is necessary to take measures to prevent surface damage when this breakdown occurs.

When two pieces of bare metal come in contact, they tend to adhere to each other. Most physicists reason that the cause for this adhesion is a natural tendency for metals to stick together. This natural tendency arises from the presence of excess electrons, or the lack of sufficient electrons in the atoms on the surface of the metals.

Because the metal parts are moving with respect to each other, the adhesions which are formed must be broken. If the junction formed is weaker than both of the metal parts, the fracture will occur at the interface of the metals, with little transfer of metal from one part to the other. However, a certain amount of plastic deformation, (the amount depending on the strength of the junction,) will occur in the bonding asperities. Because the asperities will tend to come in contact repeatedly, as the

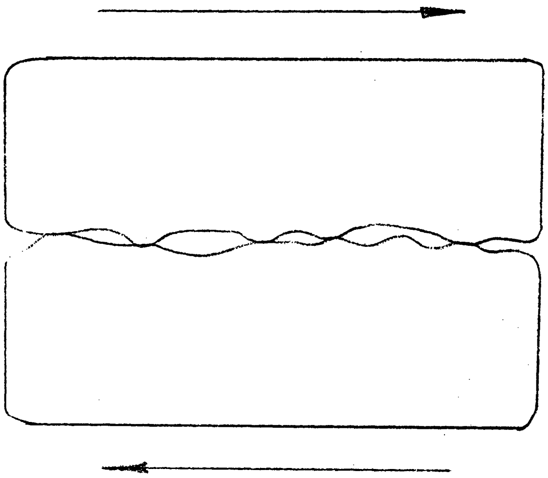


Figure 1

Mendeleeff's Arrangement of Peiodic Chart

Series	Period	Zero Group	Gp. I	Gp. II	Gp. III	Gp. IV	Gp. V	Gp. VI	Gp. VII	Gp. VIII	Gp. VIII	Gp. VIII
1			H									
2	1	He	Li	Be	B	C	N	O	F			
3	2	Ne	Na	Mg	Al	Si	P	S	Ci			
4	3	A	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni
5	3		Cu	Zn	Ga	Ge	As	Se	Br			
6	4	Kr	Rb	Sr	Y	Zr	Cb	Mo	Ma	Ru	Rh	Pd
7			Ag	Cd	In	Sn	Sb	Te	I			
8		Xe	Cs	Ba	La	Ce						
9	5											
10	6					Hf	Ta	W	Re	Os	Ir	Pt
11	6		Au	Hg	Tl	Pb	Bi					
12	7	Rn	87	Ra	89	Th	91	U	93			

Rare Earths not classified

Figure 2

operating cycle is repeated, continuing small amounts of deformation will take place. The result of this will be work-hardening of the asperities, with a consequent decrease in the ductility of the metals. After a time, depending on the amount of deformation at each time of contact, the asperities will have become brittle and tend to break off. The result may be that a loose piece of metal will be found between the parts. Metal transfer will cause scoring, while a loose piece of metal will cause abrasive wear.

If the bond between the metals is stronger than one or both of the metals, the weaker metal will fracture. In this case some of the one metal will adhere to the other. The asperity to which the metal adheres will be larger than before, and will likely contact more asperities than before. Further, in some areas, a condition of like metals in contact will develop. Both of these results will tend to promote even more bonds, and more metal transfer.

The condition of like metals in contact will cause junctions with cohesive rather than adhesive forces. Thus the strength of the bonds will tend to be greater in these areas.

> "As a result, shearing will rarely occur at the interface itself but within the bulk of the metals. Consequently, the surface damage will be very large. It is for this reason that the sliding together of similar metals may cause heavy wear.
>
> In practice, the wear mechanism is very much complicated by the presence of oxide and other contaminant films but the essential features described here are still generally valid."[1]

The transfer of metal from one part to the other will cause one of the parts to be worn away. This wear will occur only in the areas of the asperities, and will tend to occur in a linear pattern. This is the condition labelled scoring. If sufficient bonds develop, the force required to break the bonds will exceed the applied operating force, and the metals will be unable to slide on each other. This is the condition labelled seizure. This is the limiting condition of scoring, and is certainly the most dramatic and positive.

2. CRITERIA FOR THE SELECTION OF SCORE RESISTANT METAL PAIRS

In order to prevent scoring and seizure of metal pairs in sliding contact, with boundary lubrication conditions present, two criteria have been put forth. The sliding pair should: (1) be of mutually insoluble metals; (2) have at least one of the metals from the B sub-group of the periodic table.

It would appear that the pair of metals which would produce the fewest bonds, and the weakest bonds, would be the best pair for sliding applications. With few bonds, the number of possible points of scoring would be small. With weak bonds, the tendency for metal transfer and work hardening of the asperities would be minimized. It was with this in mind that Underwood first suggested that at least one of the metals should be from the B subgroup of the periodic chart.[2] Also, Ernst and Merchant suggested that mutual solubility of the metals is related to friction and wear.[3] These suggestions were formalized into a statement of criteria by Messrs. Roach, Goodzeit, and Hunnicutt, and partially substantiated experimentally by them.[4] The subject was explored further in a report by Goodzeit in 1958.[5] Davies summarized most of the available information in his paper, in a bare outline form.[6]

The reasoning behind these two criteria can be summarized in two basic statements. The mutual solubility of the two metals will determine the number of junctions likely to form. The bonding characteristics of the metals will determine the strength of the bonds formed.

3. CHARACTERISTICS OF THE METALS SUGGESTED

It is held by Bowden that one of the metals in sliding contact will reach its melting temperature due to the pressure and friction temperature at the interface. As the entire normal load plus the pressure due to the contact of the asperities parallel to the plane of sliding, is carried on the relatively small area of the asperities, the local pressures will be quite high. This condition, coupled with the temperature due to friction at the interface, could easily cause the temperature to be sufficiently high to melt one of the metals. This condition would be excellent for diffusiion to occur with soluble metals. As the metal melts, the pressure at the small area will be relieved, and the parts will move somewhat, so cooling and solidification can occur. The solid junction would then exist.[7]

Whether diffusion actually has sufficient time to take place is questionable. It is a time dependent phenomenon, but the actual time of molten contact is necessarily small due to the sliding speeds relative to a very small area of the part. It has been suggested that diffusion may be the effect of a more basic cause such as surface tension or surface energy. Present knowledge is insufficient to ascertain the true nature of the conditions, but good correlation has been found experimentally with the solubility criterion.

The strength of the bonds formed, whether at the surface due to natural adhesion, or within the metal due to diffusion and bonding, will depend on the bonding characteristics of the metals involved. Many metals display a type of bonding known as ionic bonding. In this case, the metal has free electrons in the outer shell of the atom which are free to move within the atom. A bond of this type, due to its mobility, will be ductile and strong. Other metals, specifically those in the B sub-group of the periodic chart, display co-valent bonding. In this case, a pair of outer shell, fixed electrons is shared by the two metals. This bond is immobile, and is consequently weak and brittle.

Thus, if an alloy is formed, regardless of the depth of diffusion, it will be weak and brittle with co-valent bonding, and strong and ductile with ionic bonding. The diffusion will determine how many bonds are formed, and therefore the total strength of the interface bonds. If an intermetallic compound is formed, co-valent bonding will prevail in most combinations of metals. The bonds will be quite weak. It is for this reason that Davies suggested that compound forming pairs constitute a special case of insoluble metals.

The B sub-group, as defined be Mendeleeff and suggested by Underwood, is composed of the elements in the odd numbered rows of the chart. (Figure 2) By this definition, such elements as Beryllium and Carbon are included in the B sub-group. These elements display characteristics quite different from elements like Indium and Germanium, which are also in the B sub-group. Davies used a modified form of Mendeleeff's chart (Figure 3), but included in his B sub-group (all elements to the right of Ni-Pd-Pt) several elements such as Sulfur and Carbon which do not have characteristics similar to the other B sub-group elements.

The arrangement used by Roach, Goodzeit, and Hunnicutt distinctly separates the B sub-group, including only those elements which have generally the same characteristics. (Figure 4) The degree to which these characteristics are present in the elements varies across the group, but the characteristics remain basically the same.

The A sub-group, or alkaline metals, have definite metallic characteristics, such as good ductility, malleability, thermal and electrical conductivity, and predominately close-packed crystal structures. On the other hand, the B sub-group metals show a lessening of these properties, with a steady change from metallic to non-metallic characteristics toward the right hand end of the periodic chart. The crystal structures of the B sub-group metals are more complex than those of the alkaline metals, and are more loosely packed.

A further contrast is found in the type of bonds made by these two classes of metals. As previously mentioned, the B sub-group metals tend to co-valent bonding. The outer shell electron is fixed in position, and the bond is a sharing of these fixed electrons. This is an immobile and brittle type of bond.

H																		He
Li	Be	B												C	N	O	F	Ne
Na	Mg	Al												Si	P	S	Cl	A
K	Ca	Sc		Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y		Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sm	Sb	Te	I	Xe
Cs	Ba	La	*	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac	**															
*	Ce	Pr	Nd	Pm	Sm	Ea	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu				
**	Th	Pa	U	Np	Pu	Am	Cm	Bk										

Figure 3

Davies' Arrangement of Periodic Table

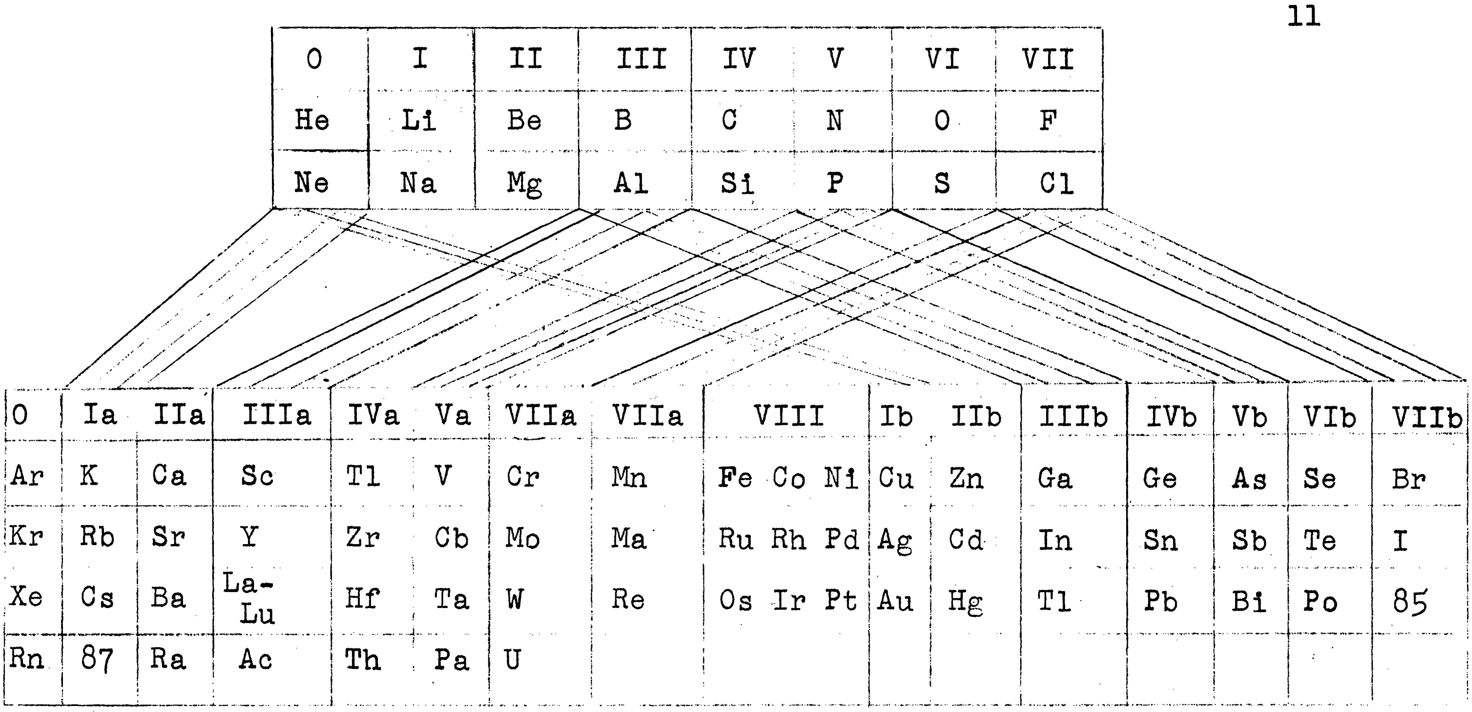
The alkaline metals, as also previously mentioned, have loosely held free electrons in the outer shell which can move about within the atom. This type of bond will be mobile and ductile.

4. EXPERIMENTAL DATA SUBSTANTIATING THE CRITERIA

Experimental data available concerning the suggested criteria show good correlation. Seven percent has been found in a test of 123 different pairs. This error, for engineering applications, is reasonably small. The error could be reduced by limiting the criteria further, but as the pairs in error are rarely found in practice, further limiting seems unnecessary.

At present, little theoretical basis exists to explain some of the phenomena found in bare metal sliding contact. Only limited experimental data is available. Thus, at present, an informed designer may take adavantage of the limited knowledge available to him, but will have to take much of it on faith. The rudiments of the field are in hand, the refinements are yet to be developed.

The tests carried out and reported by Messrs. Roach, Goodzeit, and Hunnicutt consisted of placing thirty-eight different metal sliders against a rotating disk of steel. In their results, it was found that all of those metals which showed good score resistance with steel were members of the B sub-group and were insoluble in iron or formed a compound with iron. (Figure 5) Of those metals showing fair score resistance with steel, all were insoluble or compound forming and of the B sub-group except Carbon and Copper. Carbon is 1.7% soluble in iron, but because of the form of graphite,



O	I	II	III	IV	V	VI	VII
He	Li	Be	B	C	N	O	F
Ne	Na	Mg	Al	Si	P	S	Cl

O	Ia	IIa	IIIa	IVa	Va	VIIa	VIIa	VIII	Ib	IIb	IIIb	IVb	Vb	VIb	VIIb
Ar	K	Ca	Sc	Tl	V	Cr	Mn	Fe Co Ni	Cu	Zn	Ga	Ge	As	Se	Br
Kr	Rb	Sr	Y	Zr	Cb	Mo	Ma	Ru Rh Pd	Ag	Cd	In	Sn	Sb	Te	I
Xe	Cs	Ba	La-Lu	Hf	Ta	W	Re	Os Ir Pt	Au	Hg	Tl	Pb	Bi	Po	85
Rn	87	Ra	Ac	Th	Pa	U									

Figure 4

Roach, Goodzeit, & Hunnicutt - Arrangement of Periodic Table

GOOD			FAIR			POOR		
Metal	**Group**	**Solub**	**Metal**	**Group**	**Solub**	**Metal**	**Group**	**Solub**
Ge	IVb	Comp?	C	---	1.7%	Mg	A	.026%
Ag	Ib	.0005%	Cu	Ib	4%?	Al	A	.03%
Cd	IIb	.0003%	Se	VIb	Comp	Cu	Ib	4%
In	IIIB	Insol	Cd	IIb	.0003%	An	IIb	.0009-28
Sn	IVb	Comp	Te	VIb	Comp	Ba	A	Insol
Sb	Vb	Comp				W	T	32.5%
Tl	IIIb	Insol						
Pb	IVb	Insol						
Bi	Vb	.0003%						

VERY POOR

Metal	Group	Solub%	Metal	Group	Solub%
Be	A	.05+	Mo	T	34
Si	---	4-5?	Rh	T	100
Ca	A	Insol	Pa	T	100
Ti	T	6.5	Ce	T	Sol
Cr	T	100	Ta	T	7
Fe	T	100	Ir	T	37.8
Co	T	100	Pt	T	100
Ni	T	100	An	Ib	34
Zr	T	---	Th	T	---
Cb	T	12	U	T	---

Figure 5

Score Resistance of Elements Against 1045 Steel the iron is very weak in shear in the hexagonal sheets of the crystals.[3] Further, Carbon is not truly in the B sub-group as defined by these men. Copper showed erratic results ranging from fair to poor. Copper is supposed to be 4% soluble in iron, but some question exists as to the true solubility in certain constituencies. This matter has not yet been satisfactorily resolved. All of those metals showing poor or very poor score resistance were either soluble in iron, or a group other than the B sub-group. Copper was the only exception, with erratic results.

Goodzeit reported a test using aluminum, copper, and silver disks as well as steel, with many more pairs being tested. (Figure 6) It was found that the criteria were substantiated in all cases except the pair of Selenium and Aluminum. As Selenium is a B sub-group metal and forms a compound with aluminum, it would be expected that this pair would show good score resistance. No explanation can be found for this exception. Other apparent anomalies are shown to be pairs of questionable solubility.

Davies reported the most extensive experimental data available, comprising 123 different pairs of metals. (Figure 7) He found that 114 of these pairs substantiated the criteria, while only nine did not. However, of these nine exceptions, some question can be raised as to the true nature of several of them. The Copper-Steel is called soluble, and shows fair score resistance. This is the same problem discussed previously; the solubility of copper in iron is not known under all conditions. The Aluminum-Carbon pair, called an insoluble pair with one metal from the B sub-group, shows poor score resistance. This again brings up

Miscible Pairs With Poor Seizure Resistance

Steel Disk	Al Disk	Cu Disk	Ag Disk
Be	Be	Be	Be
Al	Mg	Mg	Mg
Si	Al	Al	Si?
Ti	Si	Si	Zr?
Cr	Ca	Ca	Cd
Mn	Ti	Ti	In
Fe	Cr		Th?
Co	Fe	Co	Au
Ni	Co	Ni	
Zn	Ni	Cu	
Zr	Cu	Zn	
Cb	Zr	Zr	
Mo	Cb?	Cb?	
Rh	Mo?	Mo?	
Pd	Rh	Rh	
Ce	Ag	Ag	
Ta	Sn	Cd	
W	Ce	In	
Ir	Ta	Sn	
Pt	W	Ce	
Au	Pt	Ta	
Th	Au	W	
U	Th	Pt	
	U	Au	
		Th	
		U	

Miscible Pairs With Fair Seizure Resistance

Steel Disk	Al Disk	Cu Disk
Cu?	Zn	Sb

Immiscible Pairs With Good or Fair Resistance To Seizure

Steel Disk	Al Disk	Cu Disk	Ag Disk
Ag	Cd	Cr(Fair)	Ti?(Fair)
Cd	In	Tl	Cr(Fair)
In	Tl	Pb	Fe(Fair)
Tl	Pb(Fair)	Bi(Fair)	Co(Fair)
Pb	Bi(Fair)		Cb(Fair)

Immiscible Pairs with Poor Resistance to Seizure

Steel Disk	Al Disk	Cu Disk	Ag Disk
Li	C	C	C
Mg		Fe?	Mo?
Ca			Ni
Ba			

Pairs That Form Compounds With Fair or Good Resistance to Seizure

Steel Disk	Al Disk	Cu Disk
C	Sb	Se
Ge?	Te	Te
Se		
Sn		
Sb(Good)		
Te		

Exception:

Se(poor) on Al

Figure 6

Soluble Pairs With Poor Score Resistance

Al Disk	Steel Disk	Cu Disk	Ag Disk
Be	Be	Be	Be
Mg		Mg	Mg
Al	Al	Al	
Si	Si	Si	Si
Ca		Ca	
Ti	Ti	Ti	
Cr	Cr		
	Mn		
Fe	Fe		
Co	Co	Co	
Ni	Ni	Ni	
Cu		Cu	
	Zn	Zn	
Zr	Zr	Zr	Zr
Nb	Nb	Nb	
Mo	Mo	Mo	
Rh	Rh	Rh	
	Pd		
Ag		Ag	
		Cd	Cd
		In	In
Sn		Sn	
Ce	Ce	Ce	
Ta	Ta	Ta	
W	W	W	
	Ir		
Pt	Pt	Pt	
Au	Au	Au	Au
Th	Th	Th	Th
U	U	U	

Soluble Pairs With Fair or Good Score Resistance

Al Disk	Steel Disk	Cu Disk
	Cu(F)	
Zn(F)		
		Sb(F)

These Pairs Do Not Substantiate The Stated Criteria

Insoluble Pairs, Neither From the B Sub-group, With Poor Score Resistance

Steel Disk
Li
Mg
Ca
Ba

Insoluble Pairs, One From the B Sub-group, With Fair or Good Score Resistance; (F) = fair

Al Disk	Steel Disk	Cu Disk	Ag Disk
	C(F)		
			Ti(F)
		Cr(F)	Cr(F)
			Fe(F)
			Co(F)
		Ge(F)	
	Se(F)	Se(F)	
			Ib(F)
	Ag		
Cd	Cd		
In	In		
	Sn(F)		
	Sb(F)	Sb	
Te(F)	Te(F)	Te(F)	
Tl	Tl	Tl	
Pb(F)	Pb	Pb	
Bi(F)	Bi	Bi(F)	

Insoluble Pairs, One From the B Sub-group, With Poor Score Resistance

Al Disk	Cu Disk	Ag Disk
C	C	C
		Ni
Se		
		Mo

These Pairs Do Not Substantiate The Stated Criteria

Figure 7

the question of the definition of the B sub-group. If Davies' definition is accepted, Carbon is in the B sub-group, and this pair is an exception to the rule. If, however, the periodic table used by Roach's group is accepted, then Carbon is not in the B sub-group, and this pair substantiated the criteria. Again, the weakness in shear of graphite may mean that Carbon is a special type of metal which cannot give good score resistance.

The pairs of Carbon-Copper, Carbon-Silver, Nickel-Silver, and Molybdenum-Silver are also called pairs which fall within the criteria, but all exhibit poor score resistance. The carbon pairs may be explained by the previous discussion. In all of these cases, the B metal is one near the left end of the group, where the characteristics are much like those of alkaline metals. This seems to suggest that these metals may retain sufficient alkaline metal characteristics that they will form strong bonds, and thus be poor in score resistance.

If Davies' definition of the B sub-group is accepted, only 7% error is found with the criteria. If the further questions above are considered, the error appears to be even less. For most engineering applications, the pairs following the criteria should provide sufficient choice.

If a zero error is desired, the criteria could be limited further, perhaps to include only the right side of the B sub-group. This appears unnecessary, as 7% error is not too great for engineering work.

BIBLIOGRAPHY

1. Bowden, Moore, and Tabor, "Formation and Shearing of Junctions", Journal of Applied Physics, #14, 1943, p. 80.
2. Underwood, A. F., "Some General Aspects of Rubbing Surfaces," Proceedings of the Special Summer Conference on Friction and Surface Finish, Massachusetts Institute of Technology, 1940 pp. 5-12.
3. Ernst and Merchant, op. cit. pp. 76-101.
4. Roach, Goodzeit, and Hunnicutt, Scoring Characteristics of Thirty-eight Different Elemental Metals on High-speed Contact Sliding Contact With Steel, American Society of Mechanical Engineers, Paper No. 54-A-61, 1954.
5. Goodzeit, C. L., The Seizure of Metal Pairs During Boundary Lubrication, American Society of Mechanical Engineers, Paper No. 58-MD-9, 1958.
6. Davies, R., Compatability of Metal Pairs, Engineering Approach to Surface Damage, University of Michigan, 1958.
7. Bowden and Tabor, Friction and Lubrication, Aberdeen University Press, London, 1956.
8. Metals Handbook, American Society for Metals, 1939, pp. 72-77.
9. Burton, Applied Metallurgyfor Engineers, McGraw-Hill, New York, 1956, Chapter 1.
10. Savage, R. H., "Vapor Lubrication of Graphite in Relation to Carbon Brush Wear", Symposium on Wear, Massachusetts Institute of Technology, 1948, ASM, 1950, p. 90.

CHAPTER 2

THEORIES OF FRICTION AND WEAR

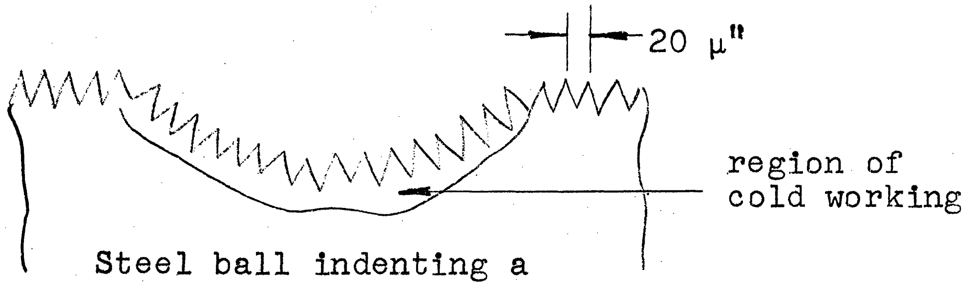
1. RELATIVE MOTION

In order to have either friction or wear, relative motion must first exist between two bodies. For the sake of continutiy, this section of the chapter will be concerned only with two metallic bodies. Figure 1 was produced by indenting a steel ball on a sheet of copper, machined with a sharp tool. The asperities and the region adjoining the dent were slightly cold-worked. In contrast, Figure 2 resulted from the sliding of the steel ball perpendicular to the section under the same applied load as in the first case. We can notice considerable wear of the asperities and an extended region of cold-working in the interior of the copper sheet.

2. THEORIES OF FRICTION AND WEAR

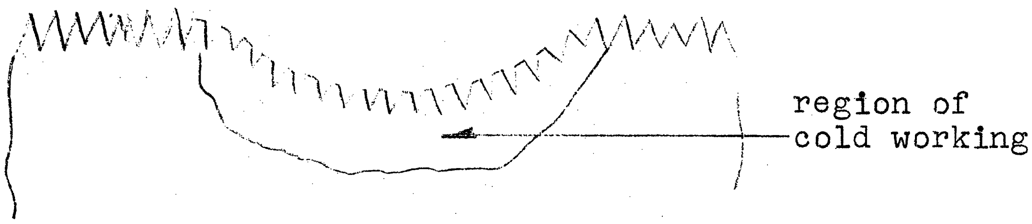
Adhesion Theory

Two theories have been presented to explain the phenomenon of friction and wear. The traditional one is commonly referred to as the "adhesion theory". In essence, it says that contact is made over a small fraction of the apparent area at relatively high spots in the surface, referred to as asperities. The actual stress is considerably higher than predicted by the traditional calculation of $\frac{\text{load}}{\text{apparent area}}$ by the ratio of apparent to actual areas. Therefore, the asperities are under an elastic deformation. This plastic deformation in the neighborhood of contact (Figure 3) work-hardens the metal, increasing its hardness, and



Steel ball indenting a copper sheet

Figure 1



Steel ball sliding on Copper sheet with no lubrication

Figure 2

Regions of Elastic and Plastic Flow

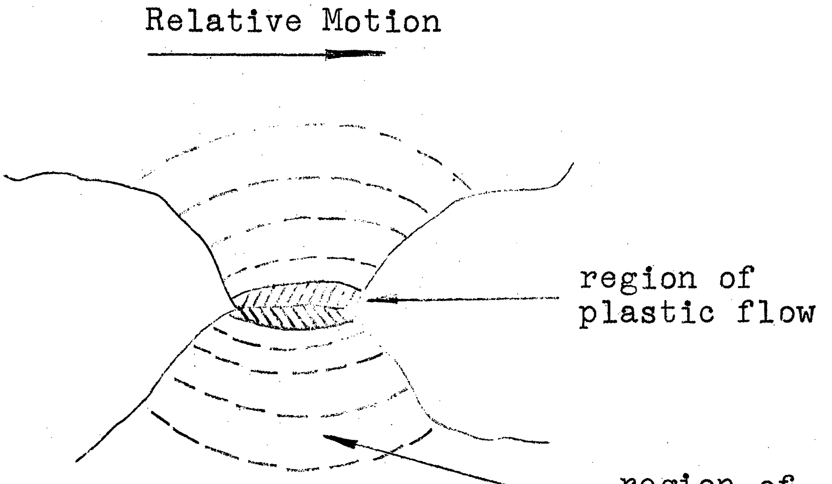


Figure 3

causes local high pressures and temperatures. Combined with the wiping or cleaning action of the metal by relative motion, conditions exist condusive to welding the metals. The tangental force necessary to shear the metals accounts for about eighty percent of the friction force, while the rest is attributed to the force necessary to push, or plow, an asperity of the harder metal through the softer one.[6] On the other hand, wear is determined by the force necessary to separate the weld. When the load is removed from the hard metal, the relieving of the elastic deformation will cause a relative motion sufficient to break the weld and prevent wear.

Interlocking Theory

Recently, I-Ming Feng proposed a new theory. He stated that in the course of motion, some asperities will only be able to cause elastic deformation; with no resulting wear. Larger asperities will mechanically interlock, mutually roughening their surfaces and strain-hardening the adjacent metal. (Figure 4) Further motion will result in the shearing of the weaker of the two metals some distance from the contact region at the plane where the product of the decreasing pressure and the increasing area will yield the smallest force. This force again results in the shearing frictional force.

A secondary adhesion force determines wear. That is, the wiping motion that cleans the two surfaces will bring the metals in molecular contact where the adhesion force ($F = K\frac{m_1 \, m_2}{d^2}$) may

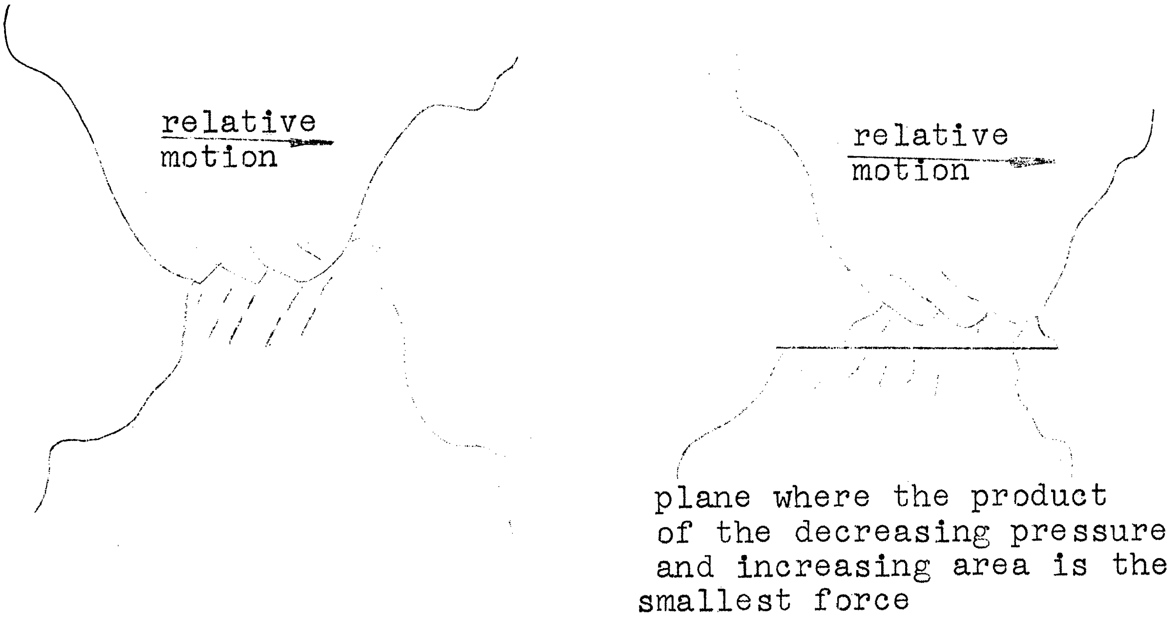


Figure 4

Mechanism of Friction and Wear

Static Friction

	Hard Steel	Mild Steel	Pt	Ni	Cu	Brass	Al	Glass	Sn	Pb
Hard Steel	.39									
Mild Steel	.41	.41								
Platinum	.40	.43	.45							
Nickel	.43	.43	.30	.39						
Copper	.55	.53	.50	.56	.60					
Brass	.54	.51	.56	.50	.62	.63				
Aluminum	.65	.61	.80	.75	.70	.71	.94			
Glass	.61	.72	.57	.78	.68	..87	.85	.94		
Tin	.79	.77	.86	.90	.88	.75	.91	.94	1.11	
Lead	1.96	1.93	2.07	2.15	1.95	2.11	2.00	2.40	2.20	3.30

Figure 5

Clean dry surfaces at low velocities, essentially static friction unite the particle with the other metal in the presence of instantaneously high temperatures and pressures. Whether or not the particles unite is dependent on the adhesive force of each metal. If this force is not strong enough, the particle will be discharged as debris on the surface.[8]

This subtle difference between the two theories explains the presence of wear particles on the surface. The attempt to explain this in the first theory by relating hardness to debris is not a complete explanation of observed phenomenon.

3. FRICTION

At this point let us examine the factors influencing friction and wear separately. Friction is essentially dependent on the following seven factors:

1. Materials
2. Contamination
3. Humidity
4. Surface finish
5. Area
6. Temperature
7. Fragments

First, each material and metal pair will exhibit a characteristic value for the coefficient of static and kinetic friction. Experimental results fortify our common sense that the coefficient of static friction is greater than that of kinetic friction. Figures 5 and 6 contain a listing of various frictional coefficients for different metals.

	Static			Sliding		
	Dry	Lubrication		Dry	Lubrication	
Hard steel on self	.78	.11 .23 .15 .0075 .0052	Oleic acid Mineral oil Castor oil Palmitic acid Stearic acid	.42	.12 .108 .081 .058 .029 .080	Oleic acid Mineral oil Castor oil Graphite Stearic acid Grease
Mild steel on self	.74			.57	.09	Oleic acid
Hard steel on graphite	.21	.09	Oleic acid		.19	Rape oil
Hard steel on Babbitt ASTM #1	.70	.23 .15 .08	Mineral oil Castor oil Lard oil	.33	.16 .06 .11	Mineral oil Castor oil Lard oil
Hard steel on Babbitt ASTM #8	.42	.17 .11 .09	Mineral oil Castor oil Lard oil	.35	.14 .065 .07 .08	Mineral oil Castor oil Lard oil Stearic acid
Hard steel on Babbitt ASTM #10		.25 .12 .10	Light mineral Castor oil Lard oil		.13 .06 .055	Light mineral Castor oil Lard oil
Mild steel on CdAg					.097	Med. Mineral
Mild steel on P. bronze				.34	.373	Med. Mineral
Mild steel on CuPb					.145	Med. Mineral
Mild steel on C. I.		.83	Castor oil	.23	.133	Med. Mineral
Mild steel on Pb	.95	.5	Med. Mineral	.95	.3	Med. Mineral
Ni on mild steel				.64	.178	Triolein
Al on mild steel	.61			.47		
Magnesium on Mg	.6	.08	Paraffin oil	.42		
Cu on mild steel	.53			.36	.18	Oleic acid
Ni on Ni				.53	.12	Octyl alcohol

Figure 6

In order to see the part the contaminants play, or more specifically lubricants, let us examine the following expression:

$$= \frac{S_m}{P} + (1 - \quad) \frac{S_f}{P}$$

where:

= percentage of area in metal to metal contact

S_m = shear strength of metal to metal junction

S_f = shear strength of film

P = flow pressure of softer metal = $\frac{\text{load}}{\text{area}}$

This expression states the intuitive; the presence of luricants reduces the coefficient of friction by a factor related to the amount of metal to metal contact.

Humidity plays an important part primarily when graphite is present. The absence of water vapor as a film on graphite has the effect of changing the lubricant into an abrasive with a considerable increase in the coefficient of friction.

The coefficient of friction decreases with surface smoothness increases up to a certain point. It is entirely possible to get a surface so smooth that the oil won't "wet" or stick. Discontinuities will exist in the film surface allowing metal contact and an increase in friction.

At first glance it would appear that friction is independent of area, since an increase in the actual area of contact results in a direct decrease in contact pressure. However, as mentioned earlier, the total coefficient of friction is composed of eighty percent shearing and twenty percent ploughing. A change in area does not affect the "shearing" component, but does affect the ploughing component. The contact area (1) is very small (often less than 1/10,000 of the apparent area of a flat surface) and independent of the over-all size of the surfaces, (2) is directly proportional to the applied load, and (3) is influenced little by the shaped and roughness of the surfaces.[1] Tests have proven that while contact area is dependent on load, the coefficient of friction, not the friction force, is independent of loads producing stresses up to 255,000 psi.

Temperatures produced primarily by velocity, may reach the lower melting point of the two metals involved. Below the melting point, there is no change in the coefficient of friction with temperature, while above it, the two extremes may occur. The coefficient of friction may increase if the lubricant breaks down, or may decrease if an oxide, sulfur, or phosphorous film is formed.

The presence of wear fragments from previous traverses on the same path increase the coefficient of friction as described by the formula:

$$f = 2af_{cu-cu} + (1 - 2a)f_{cu-st}$$

where; a = fraction of whole sliding area taken up by Cu fragments - (approximately 2 %)

The term "2a" was derived by multiplying the average value of "a/2" of the surface contact by "4" to correct for real contact areas.[16] After "n" revolutions of the same path;

$$f = 2 \quad 2n + 1 \quad a \, f_{cu-cu} + \quad 1 - 2(2n + 1)a \quad f_{cu-st}$$

4. WEAR

We are somwhat surprised to learn that a worn out truck weighs five pounds less than a new one, however, when we get an inkling of the conditions present, we can appreciate the problems involved. A one hundred horsepower engine operating for 5,000 hours on a one percent sulphur fuel would produce enough SO_2 to manufacture 1 1/2 tons of H_2SO_4.[19] no wonder the piston rings are the fastest wearing item in an engine.

Yet, when wear seems to be completely objectionable, it does have an engineering significance in the process of "break-in". The mechanism of "break-in" can be explained in terms of initial sharp protuberances subjected to high stresses leading to immediate high wear. However, wearing away of high spots brings new surfaces into contact so that the stress may drop to the low-wear region. If the original surfaces are too far out of conformity, the average stress and wear will remain high. If the original area is too smooth, leading to a rapid increase in wear, the stress may fall to the low region too soon and insufficient mating will cause poor sealing. Therefore, some parts like piston rings should initially be quite rough.[1]

When we speak of wear particles, we mean chunks of metal from 50-100A^o in diameter and not atoms. One wear particle is produced for every metal junction, ("weld" or "adhesion"). An increase in stress will increase the number of junctions, not the particle size, and thus the number of particles.

Factors Influencing Wear

Wear is influenced by the following factors:

1. Adhesion
2. Corrosion or chemical attack
3. Loose abrasive particles
4. Ploughing
5. Corrosion, surface fatigue, etc.
6. Pressure
7. Temperature
8. Hardness

All but the first may be controlled to minimize wear.

There are four conditions of adhesion. First, if the junction is weaker than either metal, (Sn on Sn) shearing will occur at the junction with little metal removal, although the coefficient of friction may be high. If the junction is stronger than one of the metals (Pb on steel), shear will occur in the weaker material leaving fragments on the harder. Wear may be great with repeated cycles leading to lapping. If the junction is stronger than both metals (Cu on steel), shear will occur predominantly in the weaker, with some in the stronger. The weaker metal will undergo a high amount of wear. Finally, if similar metals are involved (Cu on Cu) deformation and subsequent cold-working will increase the junction shear stress so that shearing occurs in the bulk of either metal. Wear of both metals will be high.[6]

As illustrated earlier, corrosive forces may present severe wear conditions. Chemical forces are accelerated by increased temperatures, increasing concentration of reactants (velocity accelerates both), and an atmosphere predominantly of oxygen and/or slightly acidic water. These latter conditions may be avoided by keeping the water in a gaseous form (above the dew point), and using sulfur and phosphorus as protective films.

The presence of wear particles from previous traverses will substantially increase wear in a self-accelerating nature due to the abrasive effect.

Large asperities, by their very nature, increase the chances for welding or adhesion and thus increase wear.

Erosion and surface fatigue are caused by gases and fluids flowing over a metal surface. The eroding of surface films leads to accelerated chemical wear.

If the pressure or stress is below a value equal to 1/3 BHN (approximate yield point), as shown in Figure 7, the wear will be independent of the stress and can be calculated by the following relationship:

$$V = k\ LAP_m$$

V = volume of metal removed
k = constant
L = distance of travel
P_m = flow pressure

This shows that wear is proportional to the true area of contact below the yield point. An increase in stress above the yield point results in a reduced shear strength of the asperities and a greatly increased volume of wear. Along with loose abrasive wear particles, the wear proceeds in a self accelerating manner. When the asperities wear to the point where the stress in the base metal equals the yield stress, gross "welding" occurs.

As an example of temperature as a controlling influence on wear, the one or two molecules on a phonograph needle in the region of the wax record have been measured at a temperature of $2000^{\circ}F$.[18] In addition, an increase in temperature will generally cause more oxidation to occur with a resulting increase in wear.

The problem of selecting wear resistant materials has been governed by experience or use of Brinell hardness numbers. However, these factors are not an accurate measuring stick and have their shortcomings in predicting the wearability of materials. In 1952, T. L. Oberle introduced the Modell factor as a measure of the depth of penetration that a metal can tolerate before reaching its elastic limit. Materials of high Modell behave like a spring, absorbing energy and preventing loads from building up to high values. On the other hand, low Modell materials are most suited for use in bearings and make a fine combination with a high Modell shaft.[15]

$$\text{Modell} = \frac{\text{BHN} \cdot 10^6}{\text{Young's Modulus}}$$

The Modell is a combination of three physical properties of the metal: BHN, stress, and strain. There has been limited discussion in literature questioning whether the BHN is weighted properly, ie. should it be BHN^2? For lack of any concrete evidence to the contrary, it stands as originally defined. See Figure 8 for sample values.

At first glance, the only controllable variable appears to be the BHN, since scores of manufacturing processes are available to meet the designer's needs: however, within pres-

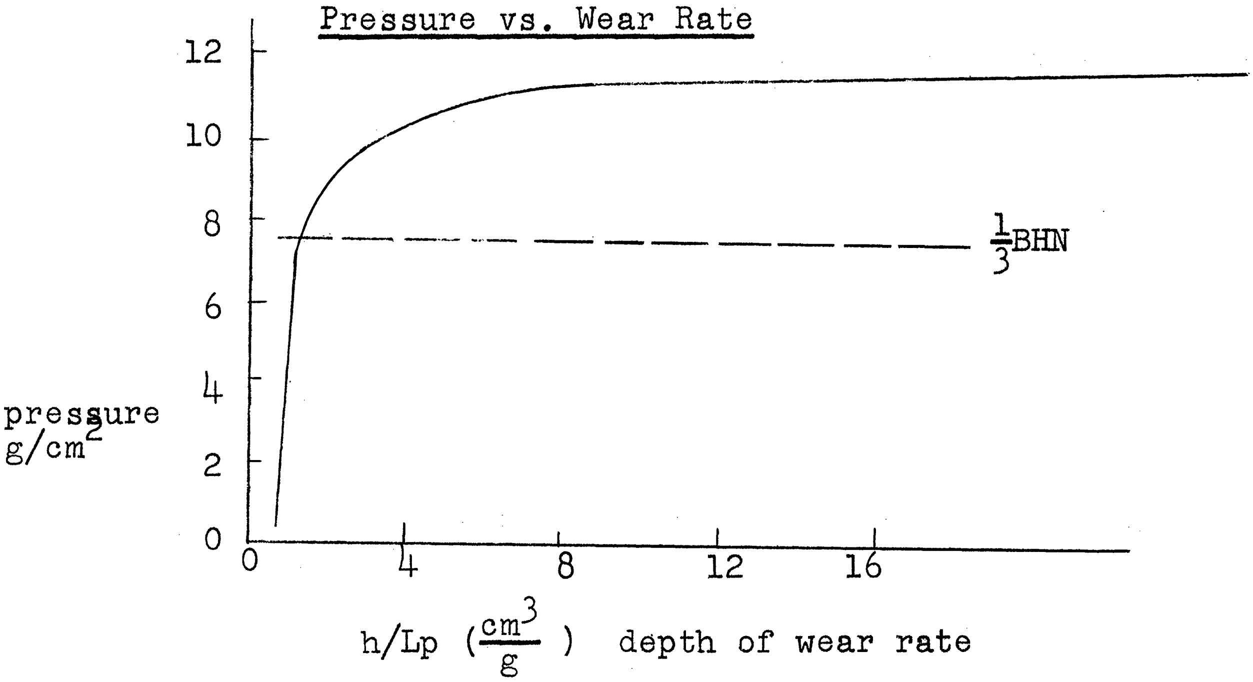


Figure 7

Materials	Young's Modulus	BHN	Modell
Alundum Al_2O_3	14,000,000	2,000	143
Chrome plate	12,000,000	1,000	83
Gray iorn-hard	15,000,000	500	33
Tungsten Carbinde	81,000,000	1,800	22
Steel-hard	29,000,000	600	21
Titanium	17,000,000	300	17
Aluminum alloy-hard	10,000,000	120	11
Gray iron-as cast	15,000,000	150	10
Structural steel	30,000,000	150	5
Malleable iron	25,000,000	125	5
Wrought iron	29,000,000	100	3.5
Chromium Metal-as cast	36,000,000	125	3.5
Copper	16,000,000	40	2.5
Silver	11,000,000	25	2.5
Aluminum	10,000,000	20	2
Lead	2,000,000	4	2
Tin	6,000,000	4	.7

Figure 8
Modell Chart

cribed limits, the modulus may be altered to the designer's wishes. Basically, Young's Modulus depends on the melting point and atomic volume of each material. Fortunately, heat-treatment, interruptions, and anisotropy are metallurgical tools available to provide some variations.*

Most engineers acknowledge the presence of graphite in gray cast iron as the reason for its excellent wearing properties. If graphite were the sole explanation, why does gray iron wear twice as well as malleable iron of the same carbon content? Basically, the answer lies in the difference of their respective Young's Modulus,(due to a difference in their structure), as shown in Figure 24, which accounts for the factor of two in the Modell number.

As the proceeding discussion has indicated, factors which tend to reduce friction <u>may</u> increase wear. Fortunately for the designer, applications seldom require a minimizing of both friction and wear. Wear is of utmost importance when designing gear teeth, while, on the other hand, a small variance in friction is of vital interest in the manufacturing process of "metal drawing".

---

*Interruptions-Cracks, gas cavities, or void (i.e. graphite in gray iron) that result in a discontinuous metal.

Anisotrpic-Variation of properties with crystal direction. A single crystal of iron has a modulus of $19.2 \cdot 10^6$ psi in direction parallel to the three major axis, but along the space diagonal, "E" equals $41.24 \cdot 10^6$ psi.

5. CONTAMINATION

The subject of lubrication falls under a general topic of contamination, which is divided into two main categories: abrasives and lubricants. As we can conclude from the previous discussion, the essential characteristics of lubricants are to:

1. Reduce amount of weld at the surface of the asperities.
2. Decrease the shear force; provide a lower shear strength than the softer metal.
3. Carry away heat.
4. Seal against undesirable contaminents.
5. Act as a carrier for rust preventatives, etc.[18]

There are three types of lubrication that may exist: hydrodynamic, boundary, or a combination of both. Hydrodynamic lubrication exists when the minimum distance between the closest asperities is around 1/40,000 in.[19] Under this condition, the laws of hydrodynamics hold and the governing factors becomes the viscosity of the lubricant. For the sake of approximation, the coefficient of friction is in the range of .001 to .01. However, more important from the engineering standpoint, is the fact that the only uncontrollable wear factor is the adhesion force. Proper lubricant selection reduces the other factors (corrosion, abrasive particles, ploughing, erosion, temperature, and localized pressure) to negligible values.

Essentially, the problem of lubrication reduces to that of satisfying the boundary layer conditons. Although in boundary lubrication, the parts are separated by a film at least one mole in thickness, complete protection does not exist. The type of lubricant plays an important part, since a poor one may only reduce the metal pick-up several hundred times, while a good fatty acid will reduce the pick-up 20,000 fold.

Critical conditions for boundary lubrication are present at slow speeds, heavy loads, and starting or restarting. Worst is the condition of restarting after the engine has been running "at heat". The coolant stops flowing, allowing the large reservoir of heat stored in nearby masses, capable of raising the temperature 500°C, to flood the bearing and cause thermal expansion. When the engine is restarted and before coolant flow is re-established, the engine may "freeze".

As indicated earlier, film thickness is important from the viewpoint of lasting protection. A film one molecule thick will greatly reduce friction but is subject to wear itself. Experiments have shown that only 900 traverses were required to raise the coefficient of friction from .11 to .24 for a one layer film while after 4,000 traverses of a seven molecule thick film the coefficient remained essentially constant. As a rule of thumb, a film of five molecules is considered sufficient for self repair.

Types of Lubricants

There are six major types of lubricants from which the engineer may select:

1. Solid
2. Oxides
3. Lacquers
4. Metallic soaps
5. Liquids and greases
6. Air

A good solid lubricant must be able to strongly adhere to the surface of metal, have a high melting point, low hardness (associated with low shear and thus low friction), laminar structure, limited solubility, limited chemical reactivity, and freedom from abrasive contaminents.[6] Of the five possible materials available MoS_2, graphite, WS_2, PbI_2, and Ag_2SO_4, only the first two fulfill the physical requirements, in addition to thoseof cost and availability.

Molybdenum disulfide, MoS_2, is generally considered the better of the two materials due to its stable, low coefficient of friction. On the other hand, graphite is dependent upon an absorbed layer of water or gas, rather than on its inherent properties. Low humidity of high temperatures that tend to decrease the availability of liquid water will convert graphite into an abrasive.

Solid lubricants will provide a low-friction film suitable in such metal processes as drawing. These solid lubricants (paraffin wax, beeswax, palm oil, and hand soap) are limited to slow speed operations because of their "friction drag". It should be noted at high pressures, liquid lubricants may be solidified while, on the other hand, high temperatures will liquify solids and thus reduce their effectiveness. Exceptions to the latter statement are fatty acids that become metallic soaps and remain effective at temperatures up to their new higher melting point.

Plating films of solid metals can produce extremely low friction coefficients, providing the underlying metals do not contact each other through the film. Thus films of FeS and In are capable of reducing the coefficient by as much as 50%.

An oxide film is around 100A° thick and, in the absence of plastic flow, will remain intact to prevent metal contact. Pressures, leading to plastic flow and relative motion, may lead to a film breakdown and intermetallic contact. In metals capable of supporting fair loads, breakdown does not occur until appreciable deformation has taken place.[1]

In machines where cyclic contact occurs, imbedability becomes an important factor. It has been shown that oxide particles may penetrate to a considerable depth within the disrupted surface zone, depending to some extent on sliding velocity and lubricant. Although metal hardness is generally considered an advantage from the wear standpoint, the imbedability is adversely affected. The oxide layer on hard steel steel is continuously formed and sloughed off, resulting in high wear, while in softer steels, the oxide becomes imbedded with foreign contaminants and forms a protective wear-reducing film.

This phenomena, along with the Modell number, accounts for the pairing of hardened journals and soft metal bearings.

The formation of lacquer films on hot surfaces such as bearings is accelerated at high temperatures by the processes of oxidation, polymerization, and decompositon of the lubricant.

This lacquer provides a safety factor and increases the load carrying capacity of the bearings by being able to employ a lubricant of poor ability or "film strength," ie. silicones. It is important to realize that the formation of lacquer on the wrong parts, ie. not on the journals may result in a fifteen fold increase in wear.[5] Fortunately the high temperature areas generally occur around bearings.

When fatty acids break down at their melting point, they may react chemically with the metal surface to form a soap that is capable of lubrication for another 70°C., or so, above the original melting point. The formation of this metallic soap is dependent on metal involved,(preferably an active one), and the presence of oxygen and water as catalysts. Fatty acids attack most metals indirectly via the oxide film. Experiments show that if metals are cut under a pool of lubricant so as to expose a clean metal surface free from oxide, they are not lubricated by fatty acid solutions at all. Under practical conditions, where metal surfaces slide on one another in the presence of a lubricant, the abrasion of the surfaces through the film may expose portions of metal free from surface oxide. These areas may not be lubricated by fatty acids. When the fatty acid attacks the oxide on the metal, the soap film formed is several molecular layers thick and produces good lubrication.[6,12]

The category of liquids and greases is most familar to everyone, but is is also the most difficult to understand. In essence, a comprehensive discussion is beyond the scope of this section. Organic compounds may react with oxygen, sulfur, chlorine, phosphorus, lead, and combinations thereof to form an array of lubricants with varying uses and properties.

Most lubricating greases are plastic solids and differ from Newtonian liquids in flow characteristics in two respects: (1) The curve relating rate of shear to shear stress passes through the origin in the case of liquids, i.e. an infinitesimal shear stress results in flow. In contrast, a plastic solid requires a definite stress before flow begins. (2) The rate of shear vs. shear stress curve is linear for Newtonian liquids, but not so for plastic solids. The viscosity of greases cannot be described by a single number at a given temperature and pressure, but is also a function of rate of shear.[19]

Finally, air has become the most recent lubricant and is used in gyroscopes and guidance systems of today's missiles.

BIBLIOGRAPHY

1. American Society of Lubrication Engineers, *Fundamentals of Friction and Lubrication in Engineering,* 1952.
2. Baumeister, *Mechanical Engineering Handbook,* 6th ed. McGraw-Hill, 1958, p. 3-40.
3. Bisson, Johnson, Swikert, and Godfrey, "Friction, Wear, and Surface Damage of Metals as Affected by Solid Surface," *NACA TN 3444*, May 1955.
4. Bisson and Johnson, "NACA Friction Studies of Lubrication at High Sliding Velocities," *Lubrication Engineering*, Feb. 1950, p. 16-20.
5. Brophy, Militz, and Zisman, "Dimethyl-Silicone-Polymer Fluids and their Performance Characteristics in Unilaterally Loaded Journal Bearings," *Trans. ASME*, Vol. II, No. 5, p. 325.
6. Burwell ed., *Mechanical Wear*, American Society for Metals, 1950.
7. Daniels and West, "The Influence of Moisture on the Friction and Surface Damage of Clean Metals, *Lubrication Engineering*, July-August, 1955, p. 261.
8. Feng, I-Meng, "Metal Transfer and Wear," *Journal of Applied Physics*, 1952, Vol. 23, p. 1011-1019.
9. Johnson, Swikert, and Bission, "Effects of Sliding Velocity and Temperature on Wear and Friction of Several Materials," *Lubrication Engineering*,May-June, 1955, p. 164.
10. Johnson, Swikert, and Bisson, "Friction at High Sliding Velocities," *NACA TN 1442*,October 1947.
11. Landair, *Wear-A Discussion of the Mechanism of Wear Phenomena and Influencing Factors,* The Nitralloy Corp., 1944.
12. Larsen and Perry, "Chemical Aspects of Wear and Friction," *Wear Considerations*, ASM, p. 73-94.
13. Lipson, *Engineering Approach to Surface Damage*,The University of Michigan, 1958, p. 29-53.
14. Marks ed., Mechanical Engineer's Handbook, 4th ed., McGraw-Hill, 1941, p. 234.
15. Oberle, "Hardness, Elastic Modulus, Wear of Metals," *SAE Quarterly Transactions,* July 1952, Vol. 6, No. 3, p. 511.
16. Peterson and Johnson, "Factors Influencing Friction and Wear with Solid Lubricants," *Lubrication Engineering,* No. 5, p.325.
17. Twiss, Lewis, and Teague, "Electron Microscope Study of Lubrication and Wear," *Lubrication Engineering*,March-April 1956, p. 102-109.

18. Zuidema, "Bearing Corrosion," The Oil and Gas Journal, Feb. 16, 1946, p. 100-103: Feb. 23, 1946, p. 151-155, 158: and March 3, 1946, p. 66-72.

19. Zuidema, Performance of Lubrication Oils, American Chemical Society, Runhold Publishing Corporation, 1952.

CHAPTER 3

EFFECT OF SURFACE FILMS

1. NATURALLY FORMED FILMS

The Adhesion Theory of Friction

Briefly, the fundamental concept is this: In the boundary or thin film lubrication region, the contaminant is not sufficient to fully separate the two sliding surfaces, and contact takes place at the surface asperities. (A contaminant is defined as any material other than those comprising the sliding surfaces) Plastic flow occurs at these highly stressed contact points, and these asperities, which have been cleaned by this action of any lubricant or contaminant, cold weld together. These welded junctions must shear before relative motion can take place.

Ploughing is the displacement of soft metal from the path of a harder metal, and although this action contributes to friction, it is usually small. Bowden and Tabor[1] assumed this term to be negligible, and concluded that the coefficient of friction can be expressed by:

$$\mu = \frac{s}{p}$$

where

s = shear strength of the softer of the two contracting metals

p = the pressure to cause plastic flow of the softer metal, or the yield strength in compression.

It should be emphasized that this expression was derived on the basis of the friction force between two clean metal surfaces; a hard steel rider and a softer test material.

Reduction of the Friction Coefficient

The best to reduce the coefficient of friction and still maintain compressive strength for loading is to use a low shear strength film (thus reducing the numerator of the preceeding equation) on a high strength backing. Thus the load will be supported through the film by the hard base material, while shear occurs within the soft thin film.

An experiment was performed by Bowden and Tabor using lead and indium films on backings of different materials.[2] The results of the test on indium are shown in Figure 1. It can be seen that the friction is almost independent of the underlying metal. The friction is determined on the surface film or on the bulk indium. Thus the friction coefficient is dependent on the shear strength of the surface film. These low shear strength films can be of the following types: oxides, chemical reaction films(chlorides, sulfides, etc.) metals, fluid lubricants, etc.

Another important function of the contaminating film besides reducing the shear strength, is to act as an antiflux to reduce welding or adhesion of the asperities. This will decrease both the friction force which is necessary to break these welds, and the wear which results from the breaking off and metal transfer of these asperities. The better the bond between film and surface, the greater the resistance to rupture of the film and,consequently, the greater the protection from welding.

Results and Discussion of Friction Tests

Clean Steel on Clean Steel

As previously mentioned, it was important to determine the friction coefficient at high sliding velocities. Probably the first important data to be collected concerning this subject is

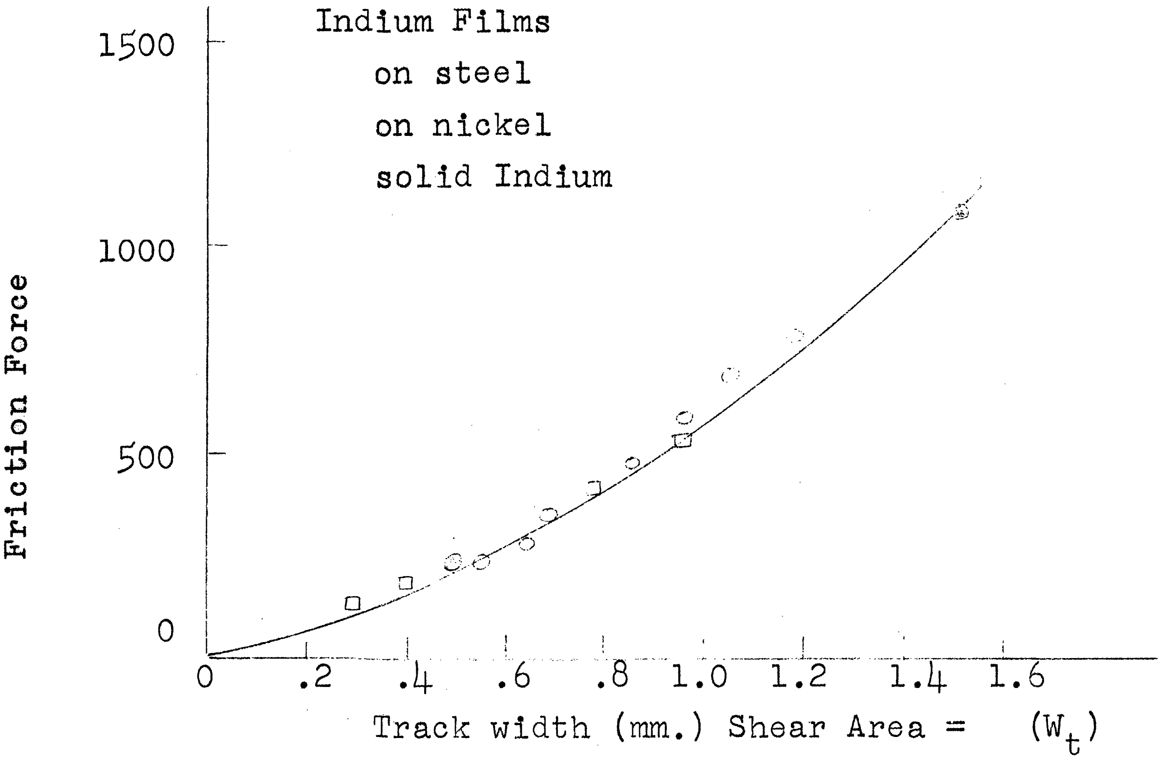


Figure 1
Friction of Indium Film on Various Metals

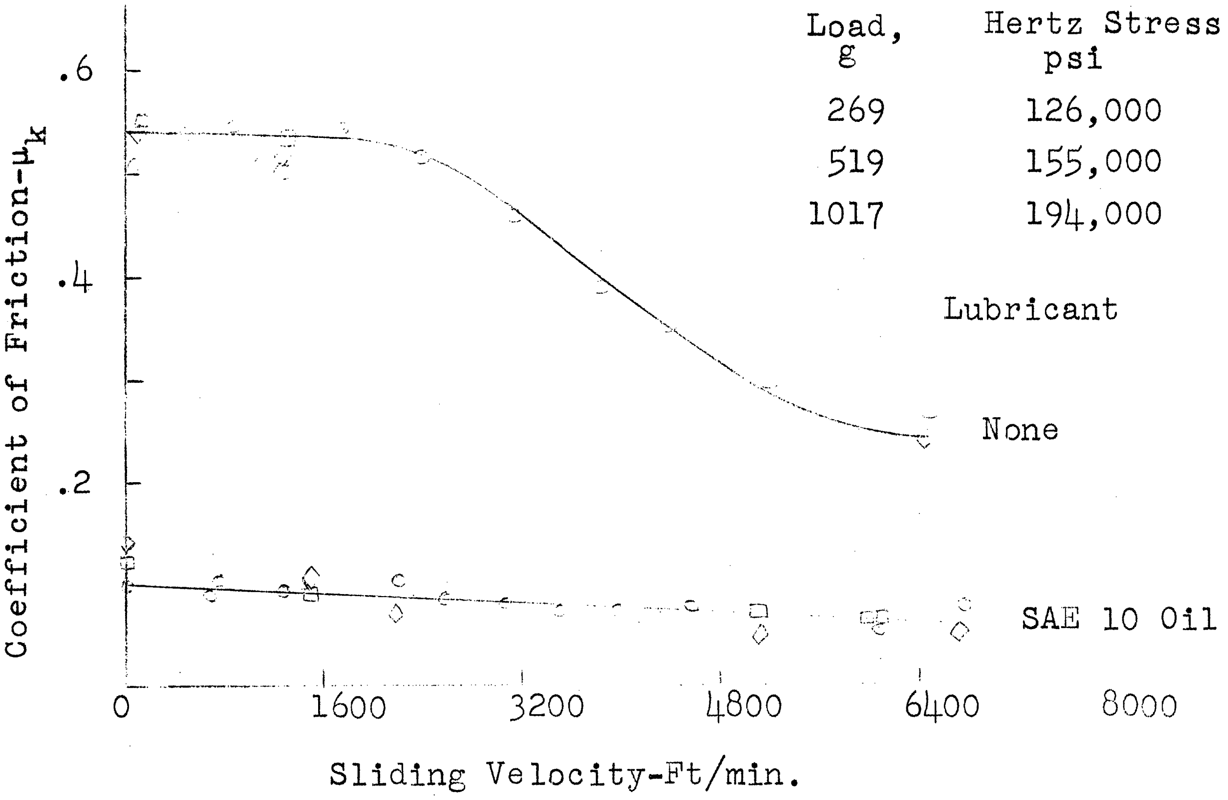


Figure 2
Friction at High Sliding Velocities of Steel on Steel

shown in Figure 2 where μ_k is plotted against sliding velocity for both a dry and a lubricated steel surface.[3] The friction coefficient decreases by over one half at the higher sliding velocities. In as much as film thickness growth is limited by wear and rupture, the friction coefficient would be expected would be expected to reach a stable value as the film thickness stabilizes. Indeed, preliminary investigations[4] indicate that μ_k may become constant for some range of sliding velocities above the range covered in this test (6600 ft/min.). The downward trend of the fricton coefficient for the dry steel specimens at the higher velocites is not considered a violation of one of the classical friction laws (friction is independent of velocity); rather it is considered the result of other variables entering into the mechanism. The downward trend is considered primarily a result of the formation of beneficial iron oxide films of appreciable thickness. The formation of oxide films would be accelerated at higher sliding velocities because of the higher rate of heat generation at the sliding surfaces resulting from the greater release of frictional energy. The increase in temperature also reduces the already low shear strength of the surface film

To partially confirm this theory, these same tests were run in an oxygen free area blanketed by cetane where, of course, no oxides could form.[4] The coefficient of friction increased almost three times to a value greater than one at the higher velocities. These high values clearly illustrate that the coefficient is independent of the load, and dependent on the high amount of cold welding that takes place at the asperities. It will be shown that this cold welding results in tearing out and metal transfer and is the major factor of wear.

Comparison of Iron Oxides

Under sliding conditions of variable severity for specimens of steel on steel, surface oxides are believed to form in the order FeO to Fe_3O_4 to Fe_2O_3. On most run-in surfaces, oxides have been identified. Figure 3 shows (qualitatively only) the estimated amounts of Fe_3O_4 and Fe_2O_3 present after various degrees of run-in.[3] (Run-in is usually defined as the process through which surfaces acquire, by mutual interation during sliding contact, increased capacity to carry load and to withstand wear.) As shown, the amount of Fe_3O_4 increases to a maximum for the "fully run-in" condition; after further operation at higher load, less Fe_3O_4 is present. On the other hand, Fe_2O_3 shows a slight increase with degree of run-in to the "fully run-in" point, after which a marked increase in Fe_2O_3 is observed with failure of the surfaces. (The surface films were identified by X-ray defraction). It has been shown that, at the minimum wear rate (which is usually associated with attainment of a well run-in condition), the amount of Fe_3O_4 is large. At the maximum wear rate (implying incipient surface failure) the amount of Fe_3O_4 is lower and the amount of Fe_2O_3 is larger than at the minimum wear rate.

In agreement with the implications of Figure 3, Finch (mentioned in Reference 3) states that iron oxide layers are detected on most run-in surfaces. He indicates that oxides play a part in the mechanism of friction by affecting the ratio of welded to unwelded area. He also indicates that injurious oxides can be formed by excessive loading. This point is confirmed in the data of Figure 3, which show that the oxide film Fe_2O_3 appears on the surface under "surface failure" conditions.

It was also shown in another investigation,[6] that the coefficient of friction of surfaces coated (preformed films) with Fe_3O_4 is appreciably less (about 30-40%) than that of surfaces coated with Fe_2O_3.

Thus Fe_2O_3 is an abrasive, detrimental oxide that should be avoided. However, it should be realized that control of these oxides is not an easy task and, as shown, all three may be present at one time.

Miscellaneous Films

Further study into the subject of the effect of high velocity on friction and wear were conducted on metals that are important as bearing cage materials. These cages have such a configuration in the bearing that they do not retain grease, and thus they must rely on other types of lubrication for low friction and wear.

Figure 4 is plotted on the same co-ordinates as Figure 2. It can be seen that Monel and Nichrome V, which are nickel alloys, exhibit a decreasing coefficient; while the coefficient

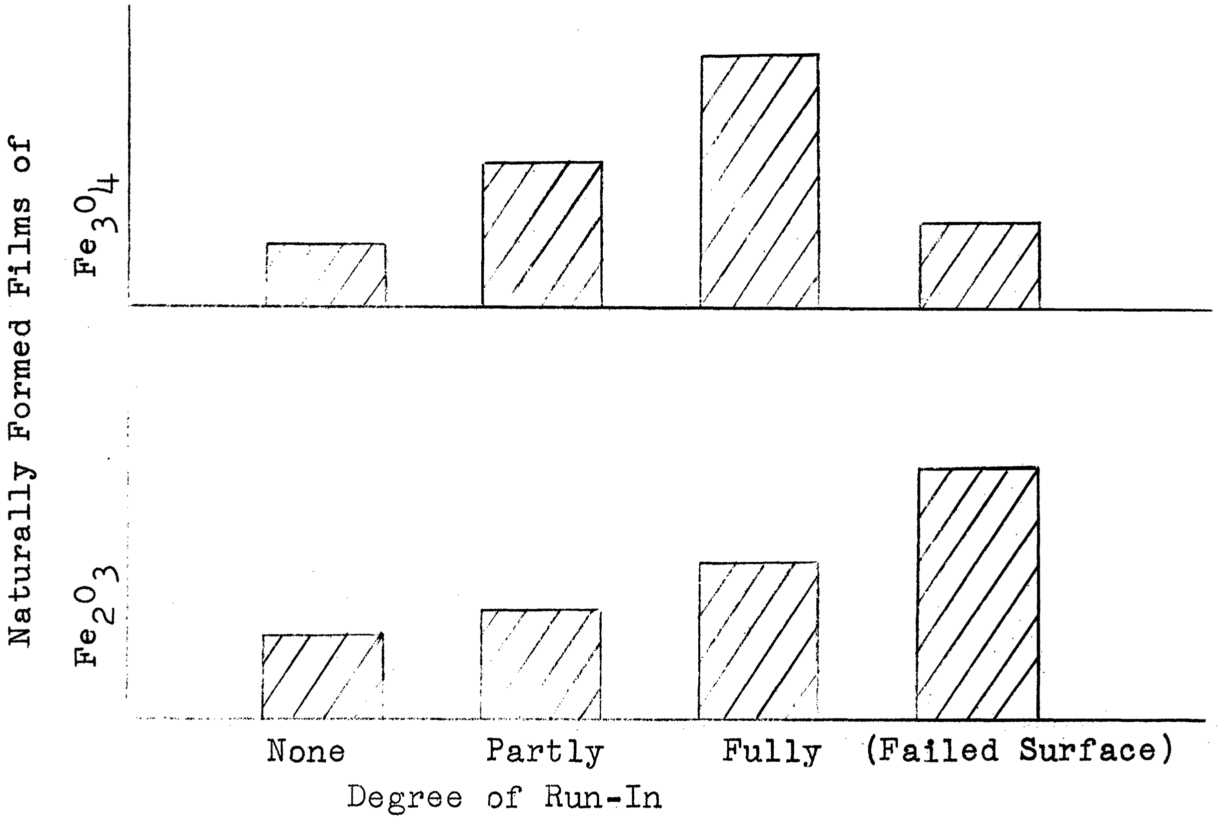


Figure 3

Effect of run-in on formation of natural films of Fe_2O_4 and Fe_2O_3. Data are qualitative and are estimated from X-ray diffraction results

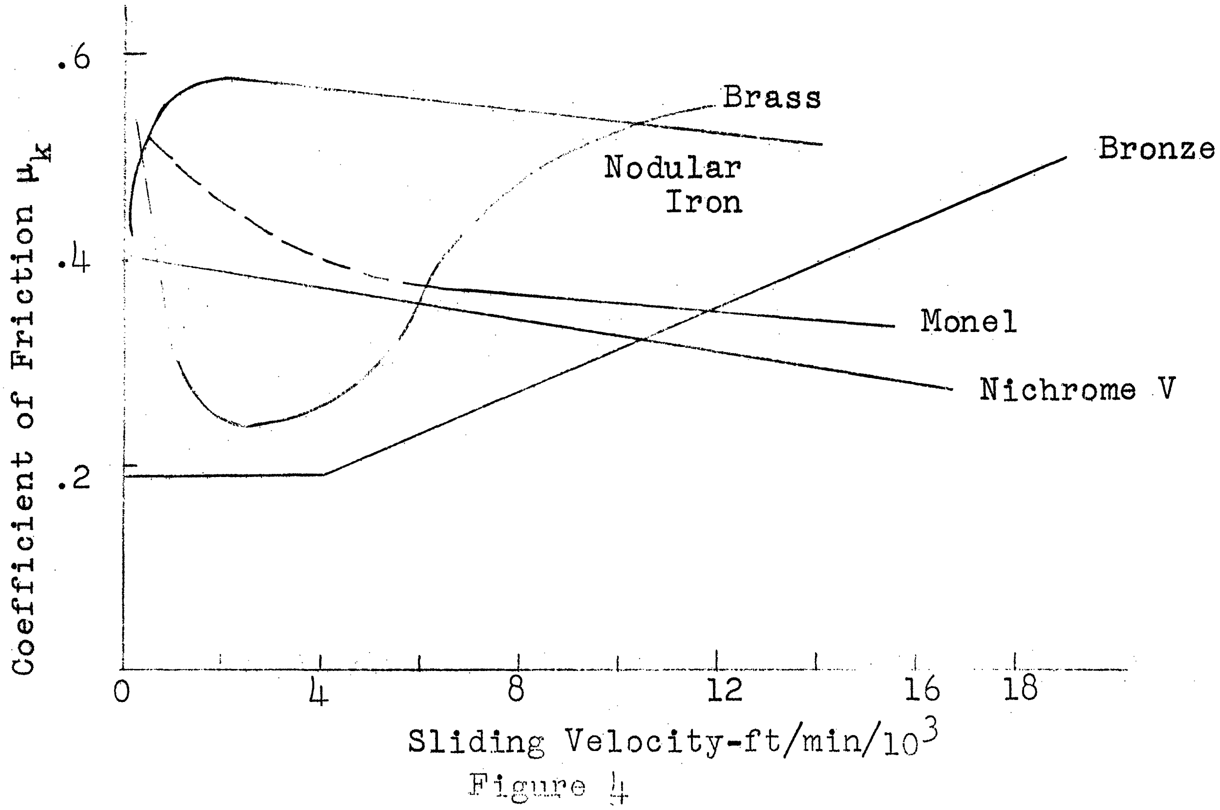


Figure 4

Effect of sliding velocity on kinetic friction of several materials sliding on hardened SAE 52100 steel without lubrication. Load 100 grams of brass and bronze increases. The coefficient of nodular iron, while high, shows the trend to decrease at the higher sliding velocities.

The explanation for these trends will be explained individually in the following paragraphs where a correlation with wear properties will be attempted, but the major reason for low friction coefficients is the ability of the materials to form a surface film.

Wear of Bearing Cage Materials

Figure 5 shows the wear volume for various materials as a function of time. Figure 6[7] has been included which shows the same materials for a load approximately five times as great. It can easily be seen that the increased load increases the wear rates considerably, and in most cases by an amount approximately proportional to the increased load.

Brass (65% Cu, 35% Zn)

In both graphs it can be observed that brass exhibits a very high wear rate. Brass has also been shown (Figure 4) to have an increasing friction coefficient at the higher sliding velocities. The reason for these trends is simple. Copper and zinc do not form appreciable oxides that would give a low shear strength or low asperity welding tendencies.

Bronze

The bronze used for the test illustrated in Figure 5 and 6 was a bearing bronze that contained 7.5% lead. At the low loads

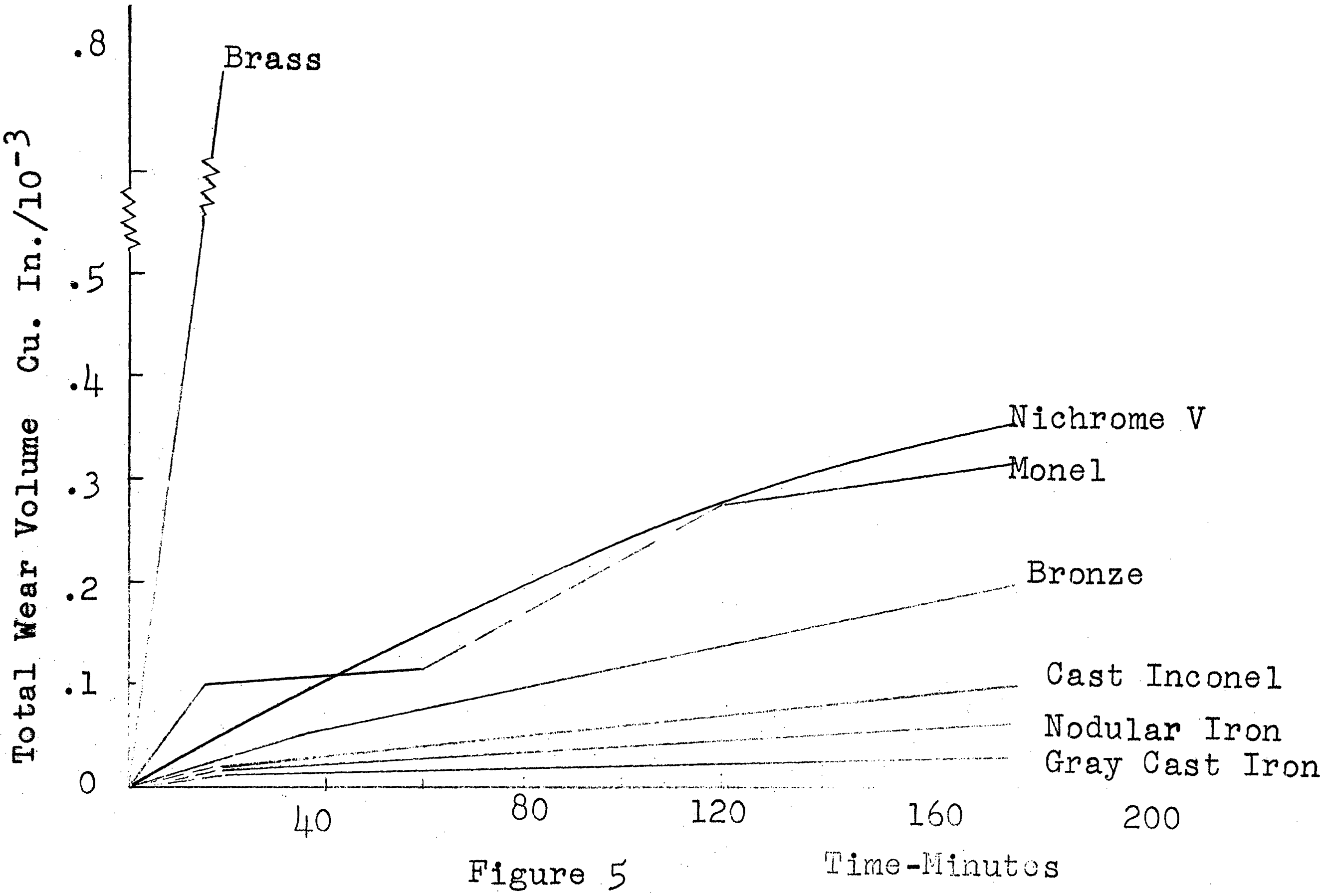


Figure 5

Wear of several materials sliding against hardened SAE 52100 steel
Sliding velocity, 5000 ft/min. Load, 50 Grams

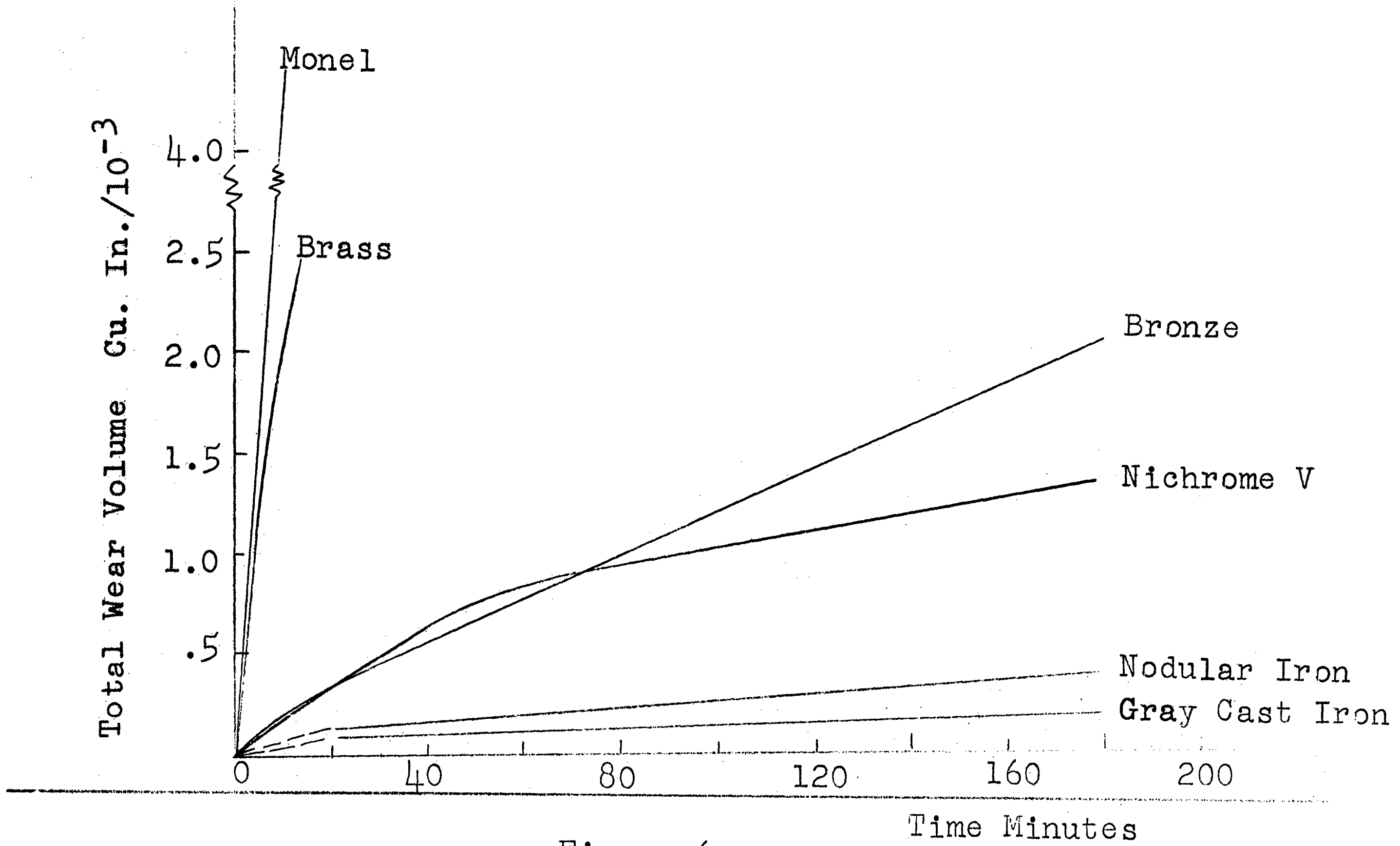


Figure 6

Wear of several materials sliding against hardened SAE 52100 steel
Sliding velocity, 5000 ft/min. Load, 269 Grams

of 50 grams, bronze shows a low wear rate in comparison to brass and even in comparison to the nickel alloys. This property is attributed to the lead in the alloy structure being able to form a surface film. However, at the higher load, the metal showed a greater increase in wear rate, but this value was still low relative to brass. Most important, as far as its use as a bearing cage material is concerned, Figure 9 illustrates the extremely low wear rate exhibited at the higher sliding velocities where lead provides an effective low surface shear strength film. To be noted also (see Figure 4) is the fact that the friction coefficient increases with velocity, which would certainly not be an indication of its wear properties.

As a further illustration of the value of the lead as a surface film, Figure 7 is submitted.[8] Here we are investigating wear as a function of temperature. (note the change in the ordinant) Presumably, in this case, the effectiveness of the lead (1%) increased with the higher surface temperatures that accompany increased sliding velocities; that is, the shear strength of the lead surface film decreased. Also it could be said that the increased plasticity of the lead would allow it to be more effective in forming a film for protecting the wearing surface. It would be expected that at some extreme condition the film would approach its melting point and wear would increase as molten film material was displaced. This might also be an explanation for the large increase in wear at the increased load, where higher surface temperatures resulting from the high load would cause failure of the film.

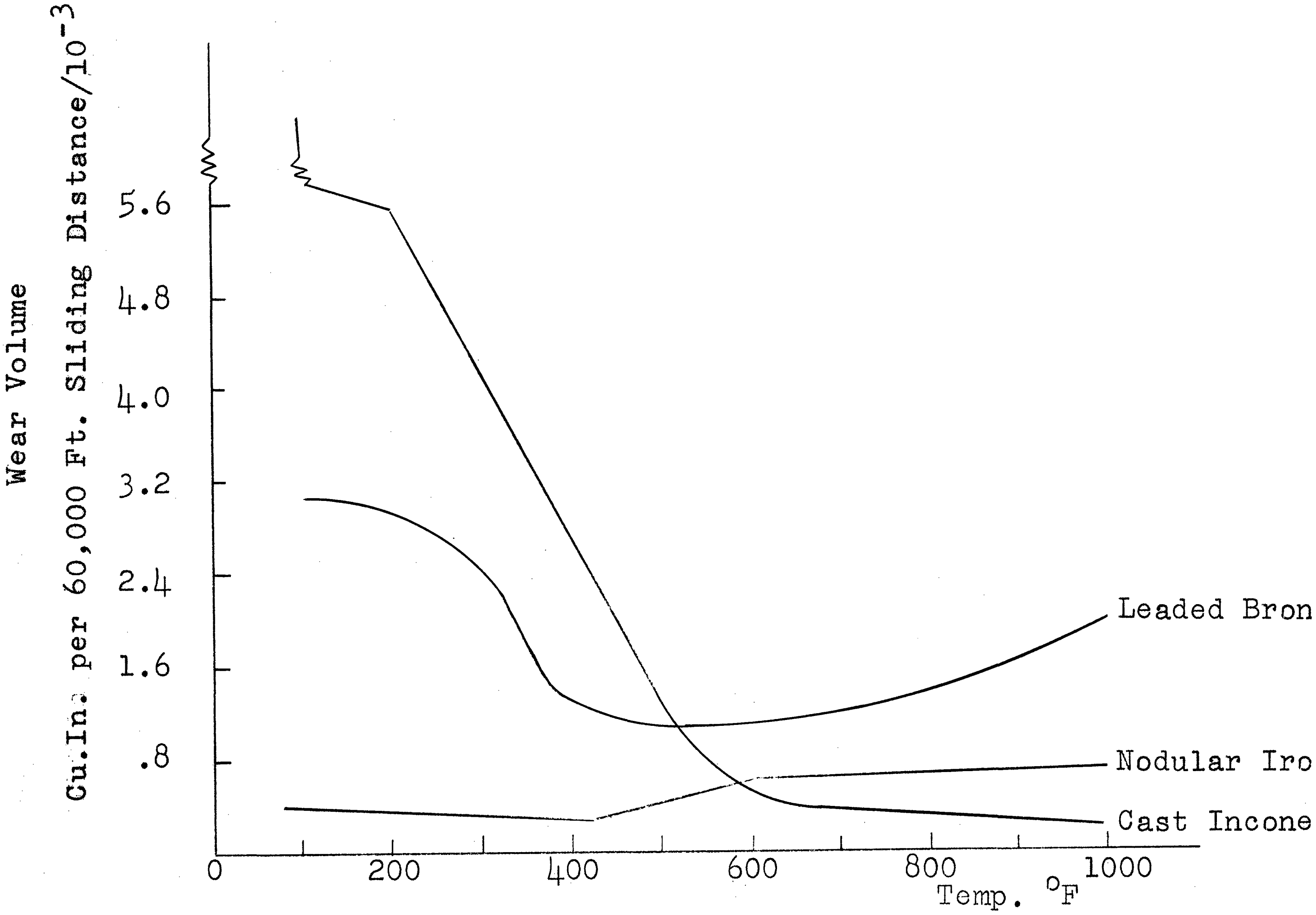


Figure 7

Effect of temperature on wear for a unit distance of sliding with various rider materials against M-10 steel. Load 1200 grams; sliding velocity, 120 ft/min.

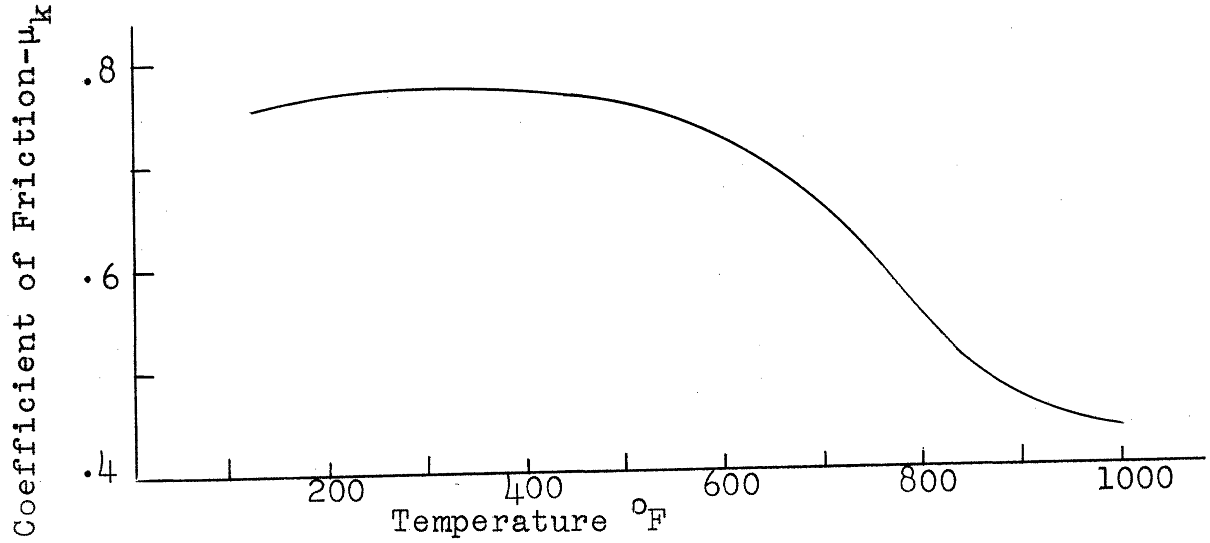


Figure 8

Effect of temperature on friction for cast Inconel (rider) sliding against M-10 tool steel (disk) unlubricated. Sliding velocity, 120 ft/min.; load, 1200 grams

Cast Irons

On the basis of wear and resistance to welding only, casting irons were the most promising materials investigated, and they have been included here because of their common use in machinery. They showed the least wear and the least tendency towards surface welding of any of the materials when dry.[7] The same observations were made for boundary lubricated conditions where it was established that the cast irons had the highest load capacities of all the materials. These properties are attributed to surface films of graphitic carbon from the metal structure being formed. Nodular iron has been illustrated in Figures 5, 6, 7, and 9 and in all instances, (that is with increased load, temperature, and sliding velocity), it exhibits a low wear rate. It should also be noticed from Figure 4 that nodular iron has a very high coefficient of friction. This again shows that it cannot always be concluded that low wear rates and low friction coefficients go together.

Finally, with respect to design, it should be mentioned that nodular iron has physical properties that are superior to those of gray iron with respect to strength, ductility, modulus of elasticity and impact strength.[9]

Nickel Oxide

It was observed that good performance (low wear and prevention of surface damage) of nickel alloys was obtained when a surface film of NiO was present.

Figure 4 through 9 illustrate data for nickel alloys (monel, Inconel, and Nichrome V). Inconel and Nichrome V are similar alloys containing about 75% and 80% nickel respectively. and Monel contains about 65% nickel.

Figure 4 shows a decreasing friction coefficient as the sliding velocity increases.

Of particular importance, is the effect of temperature on cast Inconel as shown in Figure 7 and 8. It can be seen that as the temperature increases the coefficient of friction[3] and the wear volume[8] show large decreases in value; the latter decreasing to less than one tenth of its value at room temperature. With respect to wear at high velocities, Figure 9 shows that the wear increases very little and has an extremely low value.

Included is data for a load of 1200 grams for Inconel,[8] where the trend of wear volume is downward or the reverse of the experienced at the lighter load. There was, however, found to be visible metal transfer to the disk surface with the heavier load condition at all velocities, and thus the wear volume is still higher than for the lighter load.

Monel

This metal has been included in this report to illustrate the importance of the tenacity of the surface film. At low loads, (50 grams in Figure 5) Monel, because of the nickel oxide formed, exhibits a low wear rate. However, at the higher load of 269 grams (Figure 6), it exhibits a wear rate exceeding even brass. The reason for this is simply that the film, under greater pressure, experiences plastic flow and is not able to cling to the surface and wears away.

Effect of Combined High Sliding Velocity and Evelated Temperatures

Limited information was available on this topic probably because very little investigation has been done on combined effects so far. It is difficult from the one graph illustrated[8] (Figure 10)to draw any set of conclusions particulary because the variable of load is changed. However, it can be seen that for a material forming an abundant oxide, the trend of wear to decrease with increased temperatures prevails at both low (reference curve from Figure 9) and high (Figure 10) velocities. Also to be noted is that high velocities lower the wear rate.

2. PREFORMED FILMS

Oxides of Iron

There are three oxides of iron; ferrous oxide FeO, ferroso-ferric oxide Fe_3O_4 and ferric oxide Fe_2O_3. The effectiveness of these oxides in reducing friction between sliding surfaces is shown in Figure 11. The oxides of iron were chosen because it was felt that iron is one of the most important engineering materials and therefore is of considreable interest to the engineer.

When Fe_3O_4 is formed, it forms in a small grained randomly oriented film. From this, then, it is evident that the crystals of this film can very easily adapt themelves to the irregularities

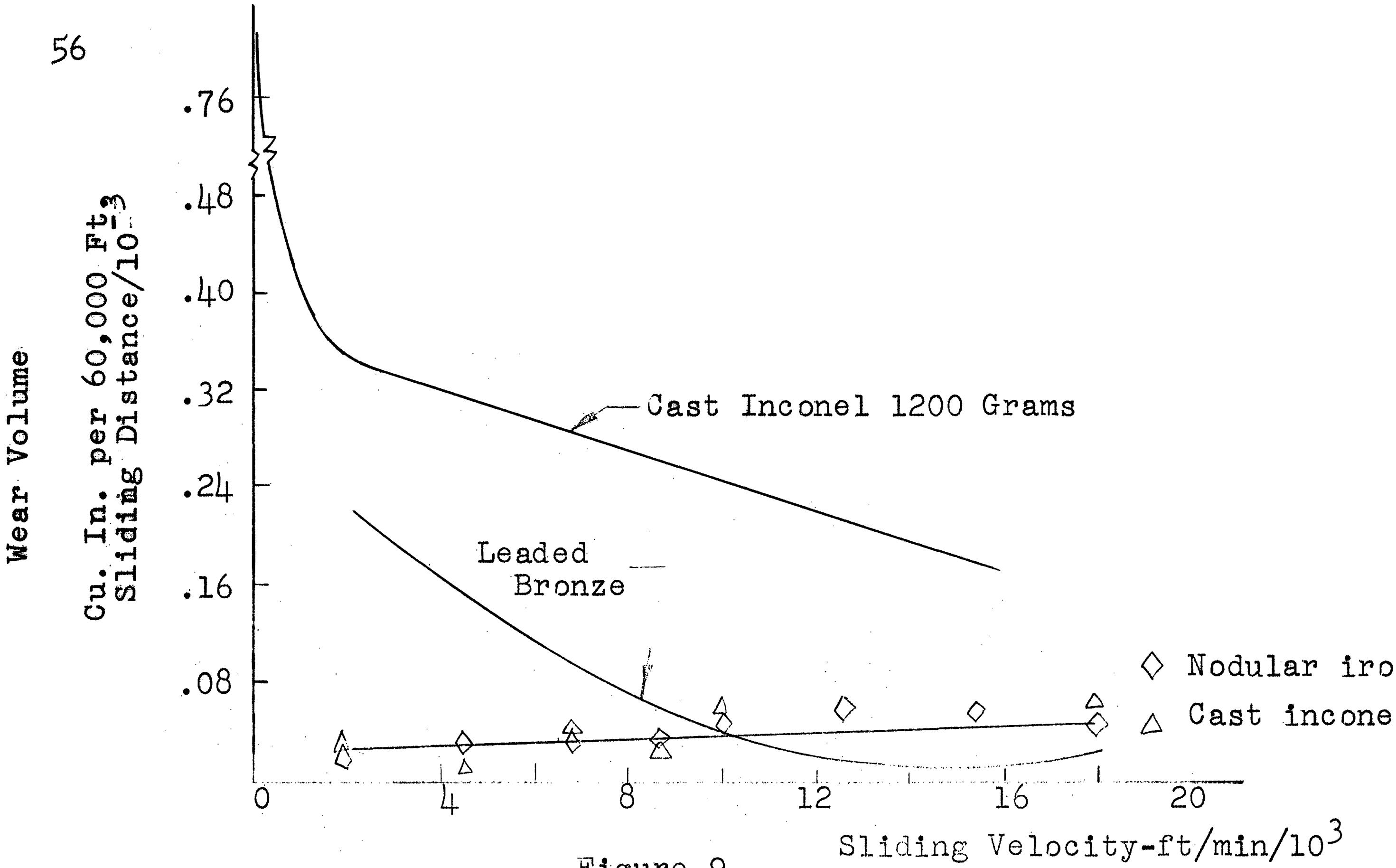


Figure 9

Effect of sliding velocity on wear for a unit distance of sliding with materials against SAE 52100. Load 269 grams; room temperature

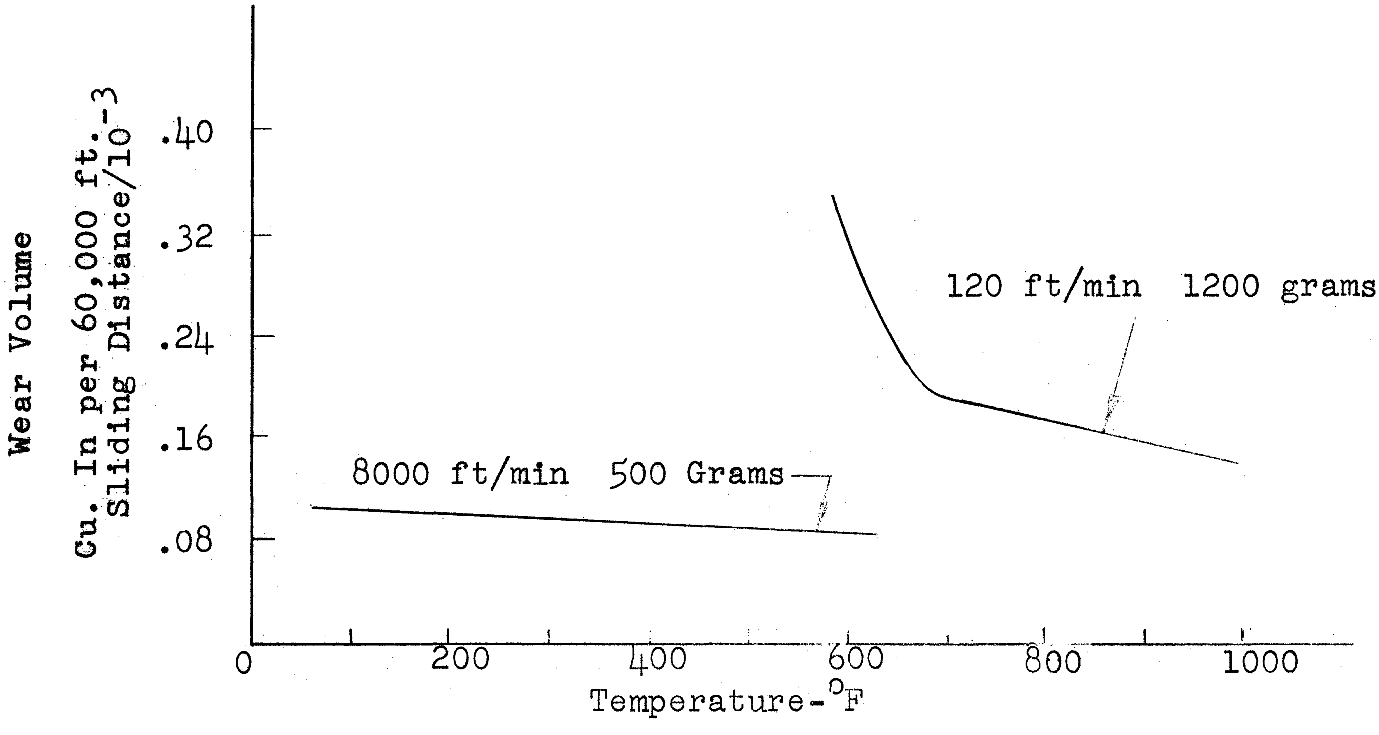


Figure 10

Effect of temperature on wear for a unit distance of sliding for cast Inconel against M-10 steel of the surface. This property of the material indicates a higher load carrying capacity which results from the film tenacity. The Fe_3O_4 also discourages plowing and ripping and thereby helps to prevent metal to metal contact and keep friction values low.

By contrast, consider Fe_2O_3. When this oxide is formed, hard, coarse grained crystals are formed in a preferred orientation. Because of the size of these crystals, they cannot readily adapt themselves to the irregularities of the surface. Because of this and their preferred orientation, this film cannot support a high load and shears quite easily, allowing metal to metal contact and therefore increasing friction and wear.

FeO was indicated as a possible effective lubricant. However, to form this oxide, very high temperatures are required which are above the annealing points of many steels. Therefore, in a practical application, the cost and economics of producing this film would not be beneficial.

Films Formed by Extreme Pressure Lubricants

Most of the extreme pressure additives used to increase load capactities of lubricants include active chlorine or sulfur. We should then investigate these two materials as to their effect on friction and wear between sliding surfaces.

Ferrous Sulfide

As evident from Figure 12, FeS is quite stable. However, even though it is stable, at 4200 feet per minute the coefficient

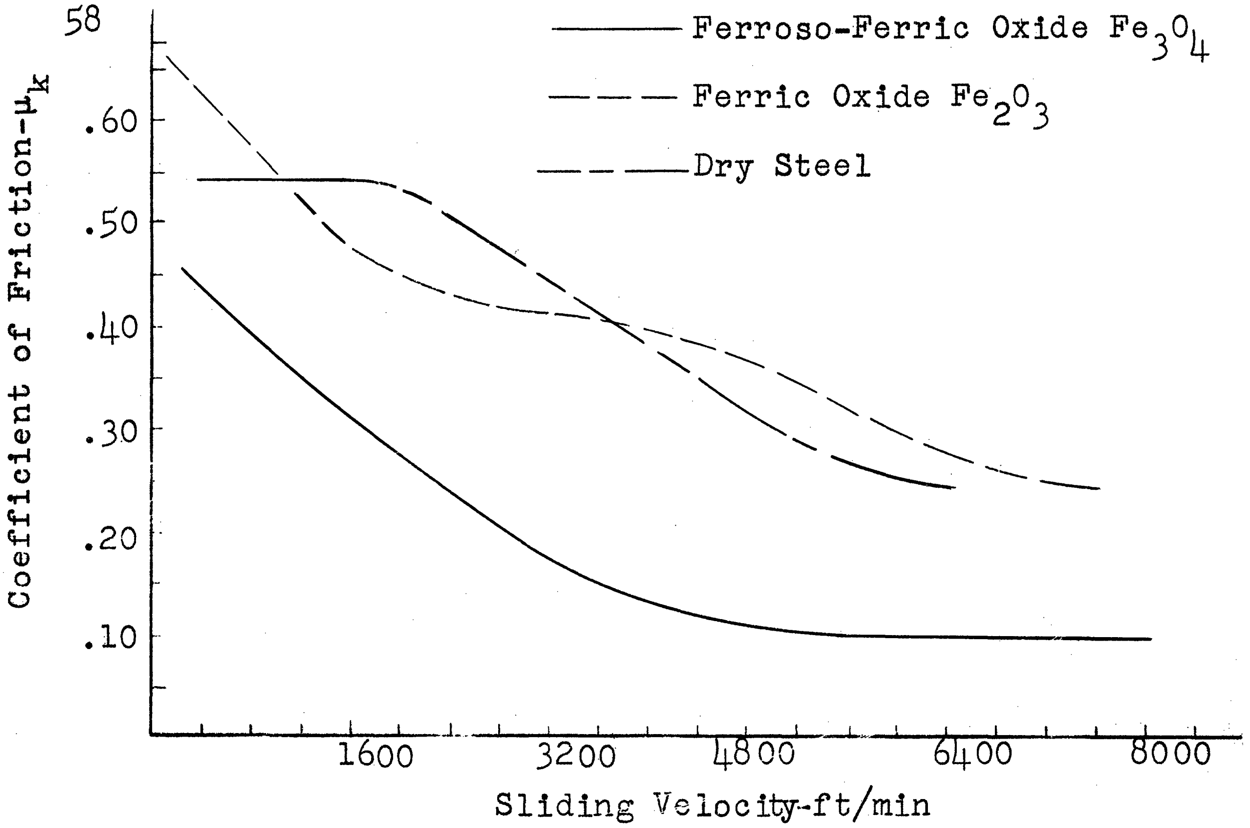


Figure 11

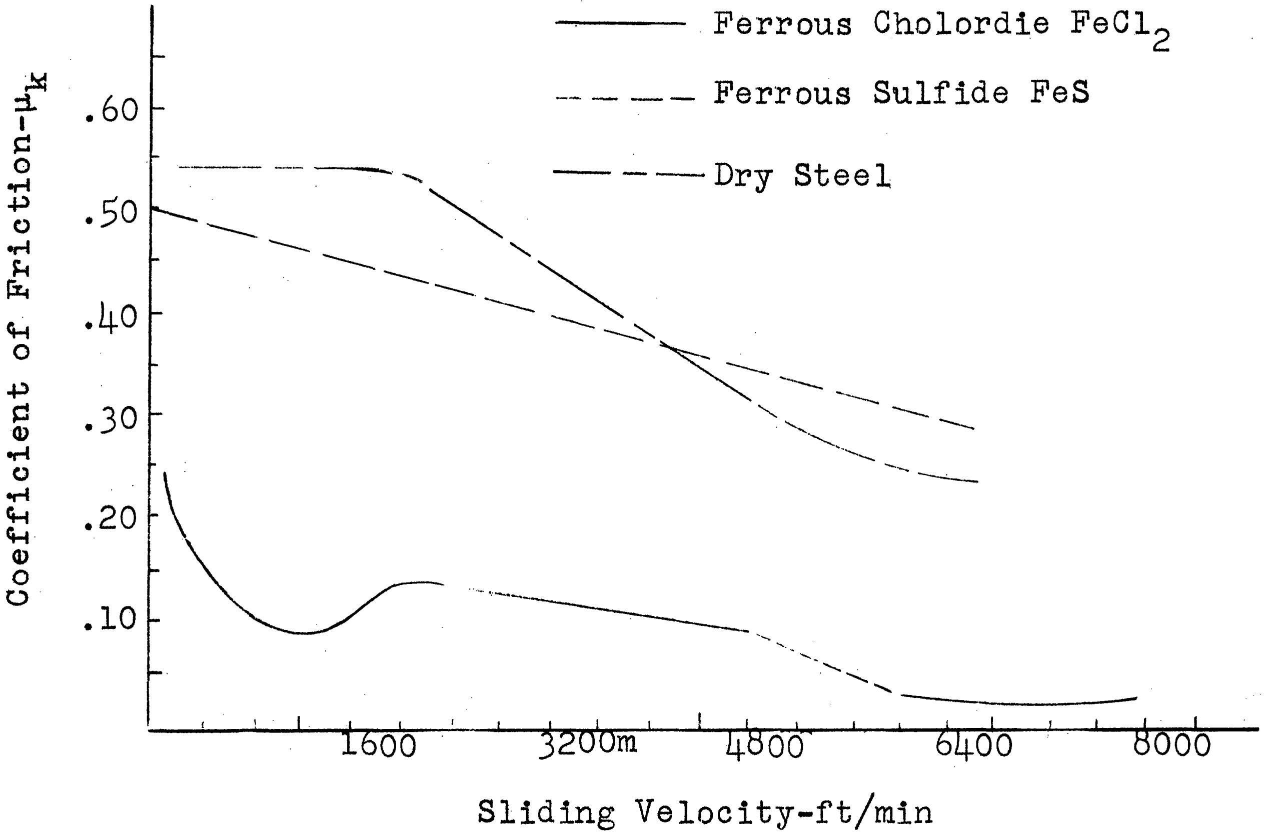


Figure 12

of friction for FeS is higher than that for dry steel. We can then say that the FeS is not a good lubricant at high sliding velocities.

From Figure 12 it can be seen that $FeCl_2$ is a good lubricant. No wear data was available for this lubricant. Due to the moisture absorbing properties of the chloride, hydrochloric acid was formed, which corroded the rider and thus the wear areas were poorly defined.

Ferrous Chloride

The curve for $FeCl_2$ from Figure 12 shows a critical range from 1600 feet-per-minute to 2000 feet-per-minute. At a sliding velocity of from 4800 feet-per-minute to 6000 feet-per-minute there is a sharp decrease in the coefficient of friction which is evident from the discontinuity of data. There are two theories as to what the action of $FeCl_2$ is:

1. "Under existing conditions at the contact points, solid FeCl may readily shear and
2. temperatures at the contact points may become high enought to melt the ferrous chloride... Overcoming the friction force will then involve shearing only a liquid film of $FeCl_2$."

These two theories are the most logical explanation of the two critical areas of the $FeCl_2$ curve. The $FeCl_2$ undergoes a phase transformation in the sliding velocity range from 1600 to 2000 feet-per-minute which effects its shear strength. In the range from 4800 to 6000 feet-per-minute, the chloride undergoes another phase transformation which results in the formation of the fluid film. This change is probably due to the higher velocities and therefore higher temperatures.

This shows then, under certain sliding conditions, the two theories on the action of the extreme-pressure additives may be substantiated. With shock or extreme loading, the liquid form of the film is probably obtained.

Supplemental or Dry-Lubrication Films

There are many lubricants in this catagory, however, probably the two best examples of such films are molybdenum disulfide and graphite. The uses of these lubricants are increasing continuously.

Graphite

From Figure 13 we can see that for light loads and increasing velocities, the friction increasing slightly. For higher loads and increasing velocities, we see an almost constant coefficient of friction up to about 2400 feet-per-minute. As we know, graphite depends on an adsorbed film of mositure for its lubricating properties. At high velocities and therefore higher temperatures the adsorbed moisture desorbs leaving a hard brittle carbon film. This film breaks up, imbedding itself in one of the sliding surfaces thereby causing abrasion. Due to its breaking up, it provides metal to metal contact. A graphite photomicrograph shows that the type of surface damage is the same for metal to metal contact. This, then is proof of our statement that the mositure desorbs leaving hard brittle carbon which the breaks and provides metal to metal contact.

Figure 13
Oriented Graphite

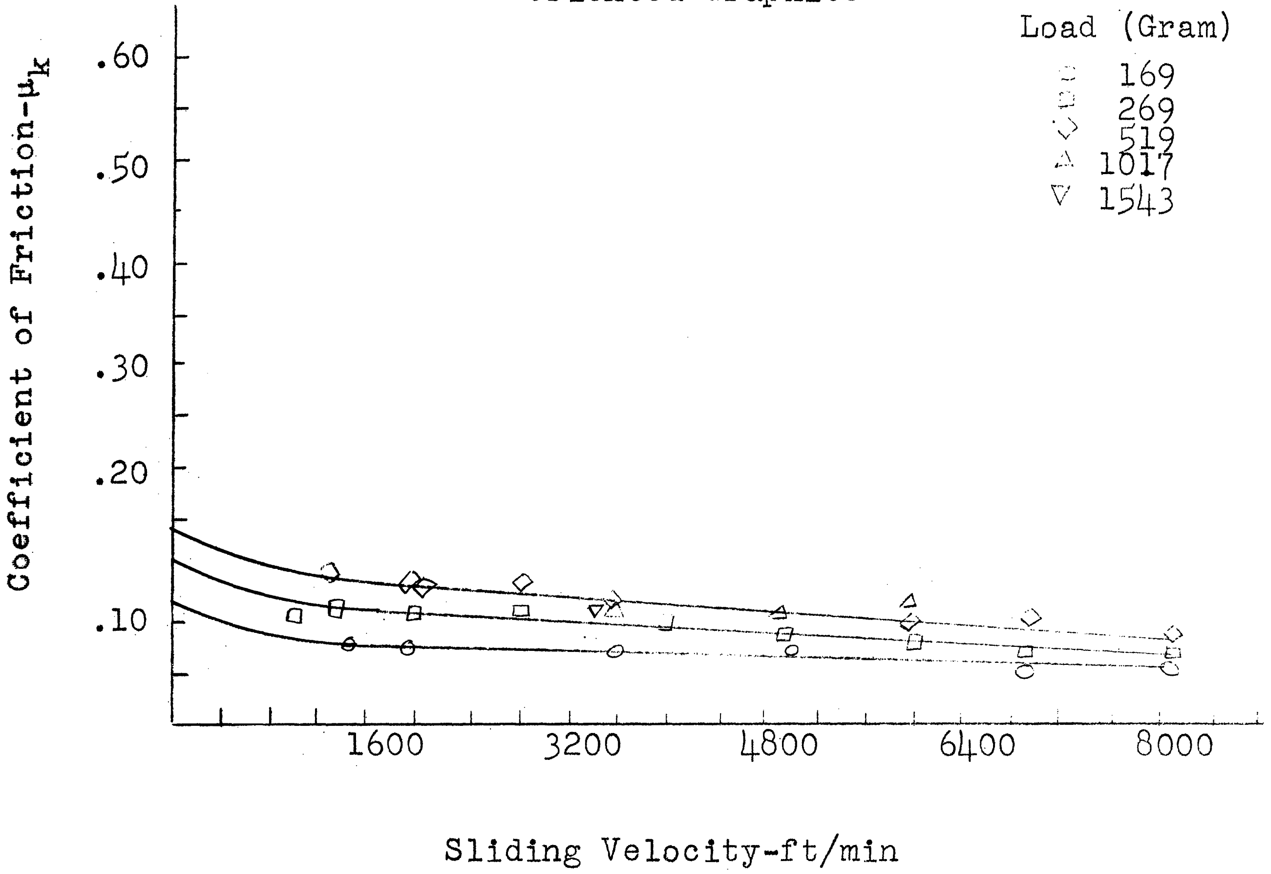


Figure 14
Molybdenum Disulfide

Molybdenum Disulfide

Molybdenum disulfide is very stable at high temperatures. It starts to oxidize at 750°F. The oxide formed is molybdenum tri-oxide MoO_3 which is detrimental to low friction. These two statements seem to contradict each other but upon examination of the mechanisms involved, we can readily see how both statements are true. As stated before, when MoS_2 oxidizes it forms MoO_3, however, not all of the MoS_2 oxidizes. A thin layer remains on the surface of the metal, between the MoO_3 and the metal as seen in Figure 16. This thin film, then, is what prevents metal to metal contact and keeps the coefficient of friction low.

From Figure 14, we note that with increasing sliding velocities, the coefficient of friction for MoS_2 decreases. This shows that MoS_2 is a very good lubricant for high sliding velocities. MoS_2 forms in a layer type film, that is, a layer of molybdenum atoms between layers of sulfur atoms. The sulfur-sulfur bond is very weak and therefore causes these layers to slip easily. The sulfur-metal bond is strong because of a very tenacious resin which binds the two together.

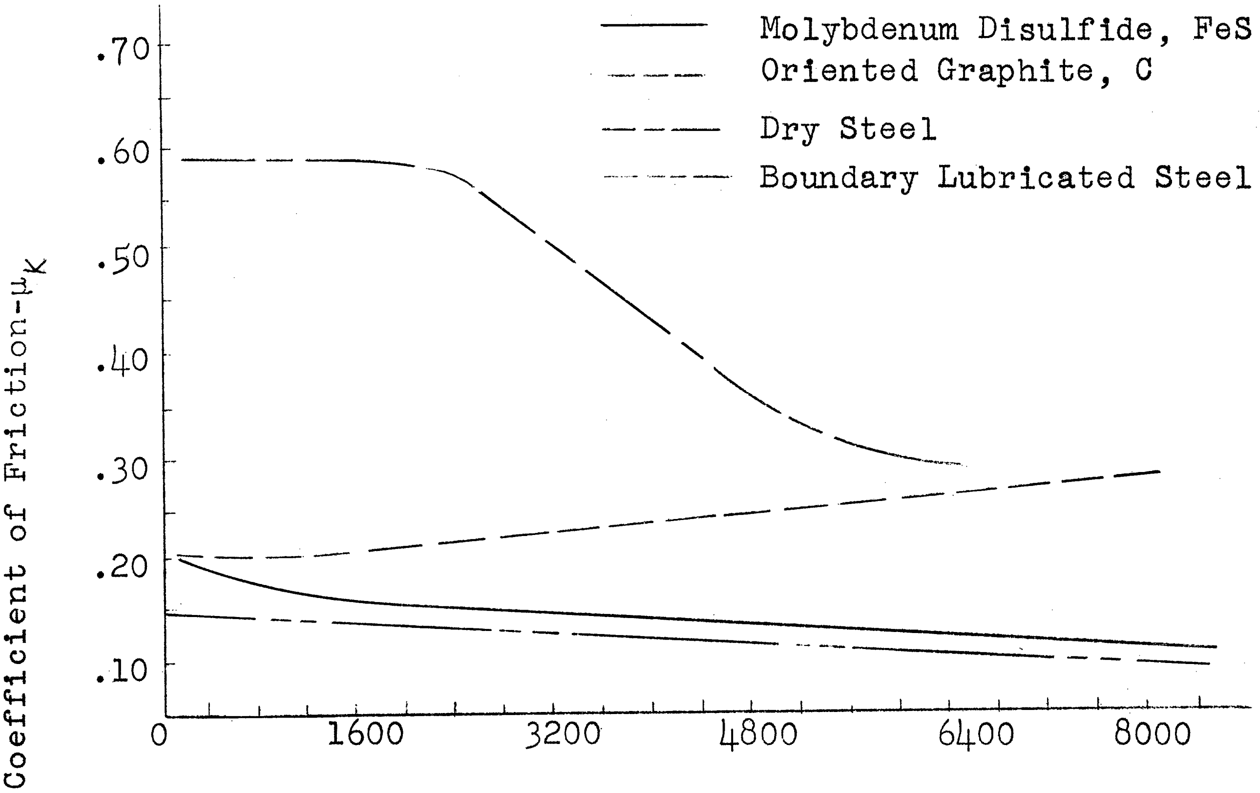


Figure 15

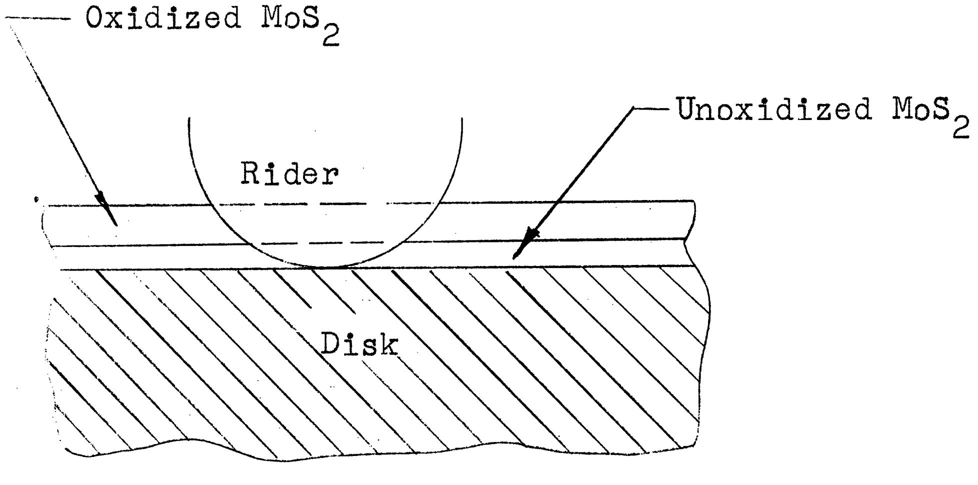


Figure 16

BIBLIOGRAPHY

1. Bowden, F. P., and Tabor, D., "The Lubrication by Thin Metallic Films and the Action of Bearing Metals," Journal of Applied Physics, Vol. 14, No. 3, March 1943, pp. 141-151.
2. Bowden, F. P., and Tabor, D., The Friction and Lubrication of Solids,Clarendon Press (Oxford), 1950.
3. Bisson, E. E., Johnson, R. L., Swikert, M., and Godfrey, Douglas, "Friction, Wear, and Surface Damage of Metals as Affected by Solid Surface Films" NACA TN 3444, May 1955.
4. Johnson, Robert L., Swikert, Max A., and Bisson, Edmond E., "Friction at High Sliding Velocities of Surfaces Lubricated with Sulfur as an Additive," NACA TN 1720, 1948.
5. Johnson, Robert L., Swikert, Max A., and Bisson, Edmond E., "Friction at High Sliding Velocities," NACA TN 1442, 1947.
6. Johnson, Robert L., Godfrey, Douglas, and Bisson, Edmond E., "Friction of Solid Films on Steel at High Sliding Velocities," NACA TN 1578, 1948.
7. Johnson, Robert L., Swikert, Max A., and Bisson, Edmond E., "Investigation of Wear and Friction Properties Under Sliding Conditions of Materials Suitable for Cages of Rolling-Contact Bearings," NACA TN 2384, 1951.
8. Johnson, Robert L., Swikert, Max A., and Bisson, Edmond E., "Effects of Sliding Velocity and Temperature on Wear and Friction of Several Materials," Lubrication Engineering, Vol. 11, No. 3, May-June, 1955.
9. Clark,D. S., and Varney, W. R., Physical Metallurgy, Van Nostrand, Princeton, New Jersey, 1952.
10. Bisson, Edmond E., "The Influence of Solid Surface Films on the Friction and Surface Damage of Steel at High Sliding Velocities," Lubrication Engineering,Vol. 9, No. 2 April 1953.
11. Godfrey, Douglas, and Bisson, Edmond E., "Bonding of Molybdenum Disulfide to Various Materials to Form a Solid Lubricating Film," Vol. 1, The Bonding Mechanism, NACA TN 2628, 1952.
12. Peterson, M. B., and Johnson, R. L., "Factors Influencing Friction and Wear with Solid Lubricants," Lubrication Engineering, Vol. 11, No. 5, Sept.-Oct. 1955.

CHAPTER 4

SLIDING AT HIGH TEMPERATURES

1. EFFECT OF TEMPERATURE ON LUBRICANTS

Liquid Lubricants

Liquid lubricants can basically be separated into two types, (1) the inert and (2) the polar or reactive lubricants.

Inert Lubricants

The inert liquids depend almost entirely on a hydrodynamic film to separate the sliding metal surfaces. If the film becomes too thin due to slow sliding speeds, low viscosity, or high pressure, metal to metal contact occurs and the galling criteria such as using insoluble pairs discussed in Chapter 1 Criteria for Scoring is particulary important.

Into this inert group fall a large number of organic and synthetic lubricants of many different chemical compositions and varying characteristics. For the most part they depend on the separation of the sliding metal surfaces by a hydrodynamic film. If the film becomes too thin metal to metal contact and scoring occurs. The advantages of synthetic and natural compounds of this type are the widely varying characteristics of viscosity, oxidation and corrosion resistance obtainable.

The most common liquid lubricants are those of the hydrocarbon type obtained from crude oil. Above 300^{o}F oxidation becomes a major problem for hydrocarbons. It should be noted that some paraffin base lubricating oils will form soaps with iron.

The fluorocarbons are similiar in structure to the hydrocarbons with fluorine replacing the hydrogen atoms. As a result they are extremely oxidation resistant but the viscosity is very temperature senstive.

The silicones are similiar in structure to the hydrocarbons with silicon replacing the carbon. They are somewhat more oxidation resistant than hydrocarbons and have viscosities little affected by temperature. If boundary lubricated sliding occurs the silicones are almost completely useless and metal transfer can be equal to that for unlubricated surfaces.[1]

There are many other synthetic liquid lubricants now being tested and synthesized. For the most part they are for rolling bearings and show little promise for the more stringent requirements of sliding bearings at high temperatures. The liquid lubricants and they properties are listed in Figure 1.

Reactive Liquid Lubricants

The polar and fatty acid lubricants tend to bond to the metal surfaces and thus help prevent metal to metal contact when boundary or thin-film lubrication occurs.

The fatty acids are oil or greaselike hydrogen-carbon-oxygen compounds that in some reactions behave as acids by reacting with bases and forming soaps. The resulting soaps have a higher melting point (usually approximately 50-100°F) than the parent acid. (See Figure 2) It is usually necessary for an oxide film to be present on the metal surface before this reaction occurs. Inert metals such as platinum, gold, and silver do not show this behavior with the fatty acids.

These fatty acids and resulting soaps are chain like compounds with the chain length and melting point proportional to the molecular weight. As a result of the metal-acid

Type	Approx. Max. Temp.	Remarks
Hydrocarbons	300°F	Petroleum base, fairly boundary lubricant, fair viscosity characteristics.
Fatty Acids	300°F	Mostly of organic origin, prone to oxidation and gumming. Solid at room temperature. Max. temp. for soap 300°F. Excellant boundary lubricant.
Silicones	400°F	Stable to 400°F. Very good viscosity characteristics, poor boundary lubricant.
Fluorocarbons, Chlorocarbons	400°-600°F	Excellent oxidation resistance, poor viscosity characteristics.
Polyphenyl Ethers	700°F	Under development for use at 700°F. Principally for rolling bearings. Solid at room temperature.

Figure 1

Liquid Lubricants

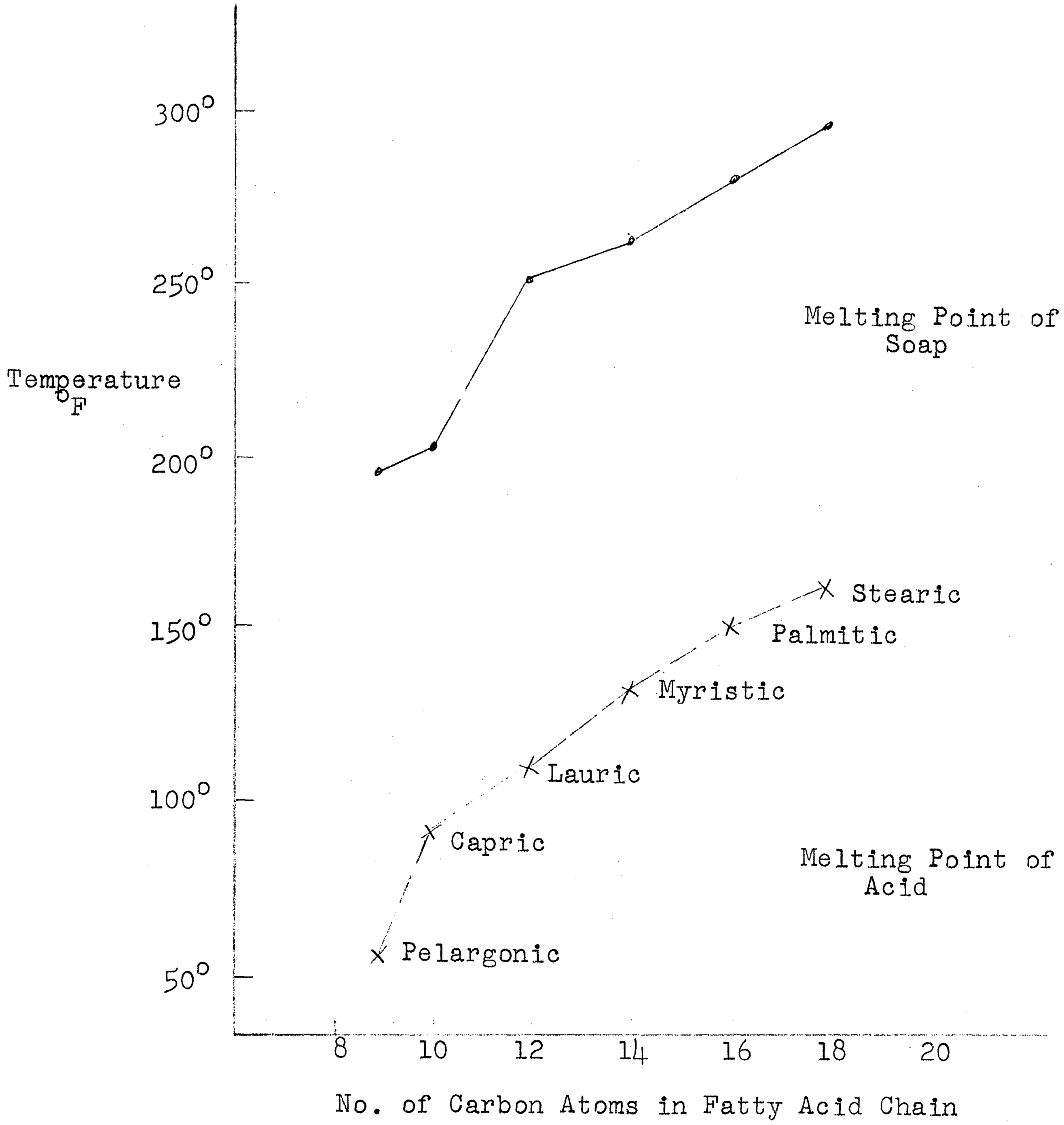


Figure 2
Breakdown Temperature of Metallic Soaps on Steel Surfaces

reaction the chainlike molecules are firmly anchored by a chemical bond to the metal surface. Thus at intermediate temperatures of 200^{o}-300^{o}F it is possible to have a fluid lubricant with a solid lubricant bonded to the metal surfaces. When boundary or thin-film lubrication occurs metal to metal contact with the resulting galling and welding is minimized. (See Figure 3) At the melting point of the soap formed however, the soap desorbs and the boundary protection is lost since the molecules are no longer bonded to the surface. The highest melting point soap commonly formed by a fatty acid reaction with iron is that of stearic acid with a melting point of 295^{o}F.[1]

Solid Lubricants

Solid lubricants are particularly valuable for application where low velocity, high pressures, or high temperature render liquid lubricants ineffective.

A solid lubricant should possess two important characteristics; (1), it should have a low shear strength to give a low coefficient of friction, (2), it must adhere to the surfaces. Wear, not friction, is usually the main problem in design. A lubricant that forms a continuous adherent layer and thus prevents metal to metal contact will give low wear.

Most solid lubricants have a layer-lattice crystal structure with layers of hexagonal platelets that are solidly linked in the basal plane but allow slip between layers. In this group are graphite and many chlorides and disulfides.[2]

Condition of surfaces	Friction coeff.	Pick-up (X 10^9 g/cm of track	Average mass individual p up fragments (approximate
Unlubricated	1.2	20,000	10^{-7}gm
Lubricated with:			
Liquid silicone	1-1.2	10,000	10^{-7}
Liquid paraffin	0.2-0.3	500	10^{-9}
Solid Paraffin	0.1-0.2	100	5×10^{-10}
Solid alcohol	0.15	50	3×10^{-10}
Fatty acid (inadequate reaction)	0.5-0.6	20	10^{-10}
Fatty acid(heavy reaction) Forming copper soap	0.05	1	10^{-10}

Figure 3

Friction and Pick-up on Boundary Lubricated Copper Surfaces in Sliding Contact (2 Kg load)

Some other solid lubricants are effective even though not layer lattice structures because of their ability to cling to metal surfaces. Iron sulfide and silver iodide are examples.

The maximum temperature for molybdenum disulfide and graphite before some oxidation beings in air is approximately 750^0 and 825°F respectively. Effective lubrication can occur to higher temperatures however. Molybdenum disulfide is effective in vacuum or inert dry atmospheres up to its dissociation temperature 1440°F. Graphite must have an adsorbed layer of water vapor or oxygen to be effective and is a poor lubricant in vacuum or inert dry gases.

One of the most promising solid lubricants is lead monoxide, PbO[4]. A study of Peterson and Johnson on the lubricating abilities of several heavy metal oxide at 1000°F showed PbO to have a coefficient of friction of .08 with no surface damage to the disk or rider of Inconel. Some difficulty caused by reversion of the PbO to high friction Pb_3O_4 was experienced for temperatures between 700°-900°F. Above 900^0 the PbO is a stable and continuous film. A low friction, low wear lubricating film was obtained to 1300°F. At room temperature the friction coefficient rises to 0.29.

The charts of Figures 4 and 5 give the characteristics of many potential solid lubricants.

The necessity for adherence of the lubricant above all other considerations, such as structure, is seen to be a prerequisite for wear prevention. Much of the ability of lead oxide as a lubricant is believed to be due to the solubility of lead oxide

Materials	Typical friction coefficent at 75°F.	Remarks
MoS_2	0.06	Oxidizes above 750°F
CdI_2	.06	Absorbs moisture
WS_2	.08	Oxidizes above 700°F similar to MoS_2
Ag_2SO_4	.14	Forms thick tenacious surface films
Na_2SO_4	.41*	Does not lubricate sliding metals
PbI_2	.28	Good film formation
Graphite	.06 - .10 (in moist air only)	Will not lubricate in dry air
PbO	.23	Lubricates well at 1300°F.
BN	.41*	Does not lubricate sliding metals
Teflon	.06	Excellent lubricant below 550°F. (A Plastic
AgI	.19	Good film formation (not layered lattice)

* 0.41 was the limit of the friction indicator in these runs. Actual friction coefficient was higher.

Figure 4

Lubrication of Steel Surfaces by Powdered Solid in Low Speed (5.7 FPM) Sliding

Load, 40 lb. (Three 3/16 in. rad. hemispheres on a flat). Atmosphere, dry air unless noted.

Lubricant	Method of application	Coefficient of friction	MHO hardness at room temperature *	Surface condition after test
None	-----	0.41	-----	Severe galling
PbO	Powder	0.08-0.12	2	Nearly continuous lubricant film; no damage
PbO	Evaporated metal film, then oxidized	0.07	2	Nearly continuous lubricant film; no damage
Pb_3O_4	Powder	0.21	2.3	Discontinuous film some metal contact; some rider wear
Ag	Evaporated	0.25	-----	Film worn and penetrated, allowing contact of base metals
Bi_2O_3	Powder	0.31	-----	Very thin film with local metal contacts, formed after some galling
CdO	Powder	0.34	3	Very thin film with local metal contacts
In_2O_3	Evaporated metal film, then oxidized	0.36	-----	Very thin film with local metal contacts
WO_3	Powder	0.41	2.5	Failed surface; galled
Sb_2O_3	Powder	0.41	2.5-3	Failed surface;solid flowed plastically to form irregular film, allowing metal contact
PdO	Powder	0.41	-----	Severe galling
BN	Powder	0.41	-----	Severe galling
Tacl	Powder	0.41	1-1.5	Severe galling

* Handbook data

Figure 5

Coefficients of Friction for Various Lubricants at 1000°F.

(Lubricants listed in order of decreasing lubricating effectiveness) with the base metal oxide to form a tightly bonded coating.[4] It should be noted that PbO can be highly corrosive to some metal at temperatures above 1000°F.[4]

New lubricants, both solid and liquid, are constantly being developed and evaluated for their high temperature properties. At the present state of this relatively new science none have proved entirely adequate. Many combinations are being tired in an effort to overcome the deficiencies of individual lubricants.

Various schemes such as resin binders to protect and retain solid lubricants have been attempted. So far, powdered metal compacts seem to be the only materials with the necessary high-temperature strength and stability.

2. HIGH TEMPERATURE SLIDING

The properties of all solid and liquid lubricants generally become inadequate as the temperature is increased above 400°-500°F. For this reason increasing attention is being paid to the sliding characteristics of metal at high temperatures. Comparatively little is now known on the subject. It is only in the last decade that extensive investigation has been done.

High temperatures add the following complications to those ordinarily incountered in sliding.

1. Lubricants decompose, oxidize or are ineffective.
2. High temperatures increase the welding tendencies of the contacting metals.
3. Softening of the matrix metal may occur at operating temperatures.
4. The oxidation rate increases and oxide films form rapidly.

With the lack of suitable lubricants the formation of oxide films becomes a factor of dominating importance.

Desirable Characteristics for Oxide Films

The function of oxide films of often reducing friction by decreasing metal to metal contact has been discussed in another chapter. Even more important is the reduction of wear. The following properties of oxide films are necessary to minimize wear:

1. The oxide film must be adherent.
2. The oxide base metal strength properties must be similar.

Adherence of the oxide film to the base metal is to a large degree dependent on the oxide-metal specific volume ratio. If the volume ratio is equal to or greater than one the oxide forming on the base metal can be a dense adherent scale that prevents further oxidation of the base metal.

$$\frac{V \text{ oxide}}{V \text{ metal}} > 1$$

If the volume ratio is less than one the relatively brittle oxide film will be in tension and tend to crack or shear off.

Metals with favorable volume ratios that have been found to form adherent oxides are copper, beryllium, zinc, aluminum, silicon, lead, chronium, molybdenum, iron,cobalt and nickel.

Similar mechanical properties of the oxide and base metal substrate are necessary to prevent the oxide from breaking up under load. If the metal substrate yields plastically under the oxide film the relatively brittle film will break up and allow metal to metal contact. Figure 6 lists the pure metal and oxide properties for several common engineering metals. Of particular interest is the column "load at which metal to metal contact occurs." Here it is seen how a weak substrate, (low hardness), as for tin and aluminum causes the oxide to break up under low stress and allow metal to metal contact. High temperatures will agravate this by softening the metal matrix more than the oxide.

Figure 5 also shows that the coefficient of friction can be less for sliding contact with some metal-to-metal contact than for continuous oxide to oxide sliding. Oxide to oxide bonds can be quite strong. This and some of the subsequent data emphasize again that friction and wear are not necessarily directly correlated.

Friction and Wear at High Temperatures

Little direct work has been published on this subject but the oxide-metal film volume and strength criteria previously mentioned appear to be applicable to the work of Peterson and Florek[5] and that of E. P. Kingsbury and E. Rabinowicz[6].

In order to determine wear of sliding surfaces, it would be desirable to have an easily measured quantity as a parameter. It is generally accepted that coefficient of friction is an indication of wear. This, however, is not a direct relationship, but measuring friction force is an easier task than measuring actual metal removed. For this reason, most researchers will indicate wear, by friction factors.

Metal	Hardness, kg/mm^2		Load at which appreciable metal contact occurs, g	Coefficient of Friction	
	Metal	Oxide		Metal on metal	Oxide on oxide
Gold	20	----	0	2	---
Silver	26	----	0.003	1	0.8
Tin	5	1650	0.02	1	1
Aluminium	15	1800	0.2	1.2	0.8
Zinc	35	200	0.5	0.8	1.2
Copper	40	130	1.0	1.6	0.8
Iron	120	150	10.	0.6	1.0
Chromium	800	----	>1000.	---	0.4

Figure 6

Coefficient of Friction of Clean and Oxide-Coated Metals at 77° F.

Among the pure metals which were tested by Peterson and Florek[10] were iron, copper, aluminum, zirconium, nickel, and molybdenum. Several alloys were also tested including Inconel, Hastelloy B tool steel, and 1020 steel. The results of their work are shown in Figure 7-10.

The run made with iron revealed that the sliding process did assist in forming a run-in film. From the standpoint of engineering applicability it is also interesting to note that the tool steel and 1020 steel curves are very similar to the iron curve.

The results of the copper run illustrate that the film is being removed around 300°F. due to the thermal softening of the substrate giving a rise in friction.

Aluminum showed a gradual decrease in friction followed by a rather large rise. This rise in the coefficient of friction is believed to be due to the softening of the substrate causing the oxide to break up. However, it has also been suggested that the oxide is actually changing its structure as the temperature approaches 650°F.

Zirconium illustrates the non-adherence of a film. Indeed this was found to be true by Peterson and Florek as powdered oxide was found on the sliding surfaces.

The curves for Inconel and Hastelloy B show the relationship between alloys and constituents. The oxide film will result at a temperature where the least oxidant resistant constituent begins to oxidize.

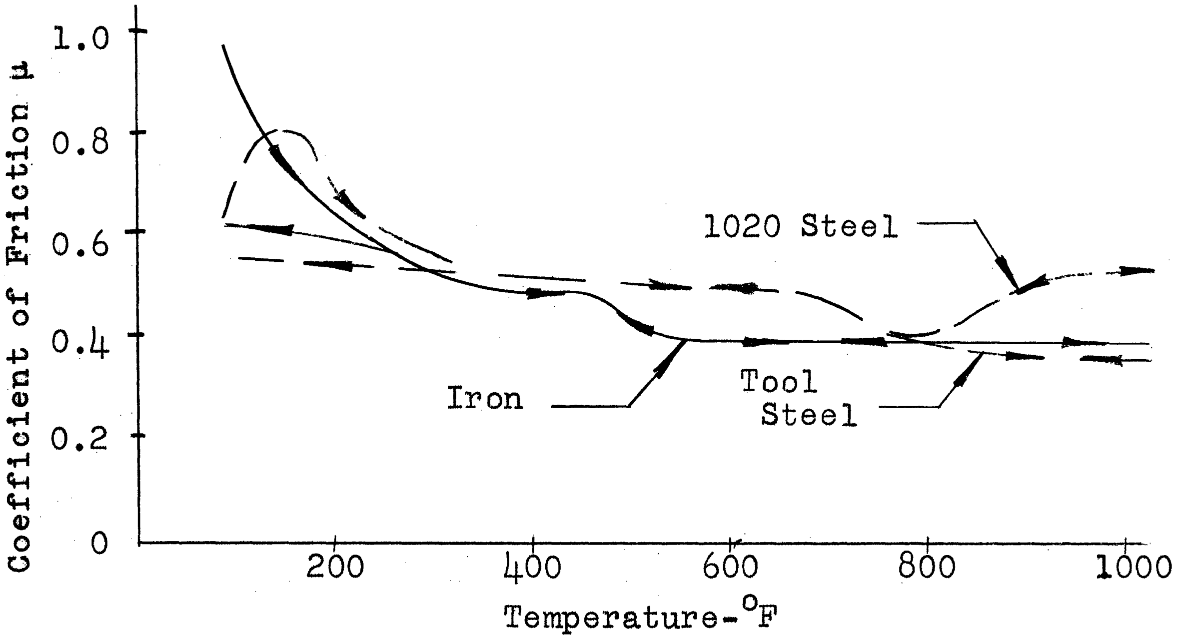


Figure 7

Effect of Temperature on Coefficient of Friction

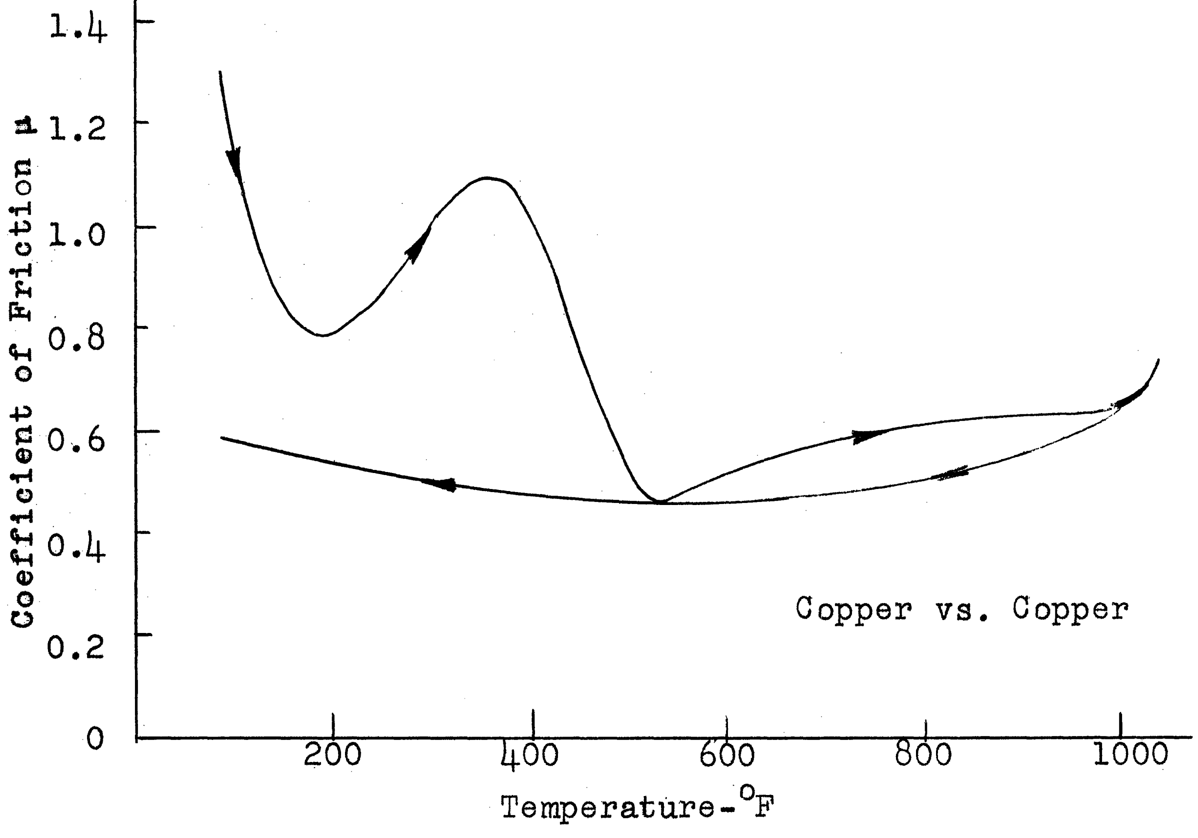


Figure 8

Effect of Temperature on Coefficient of Friction

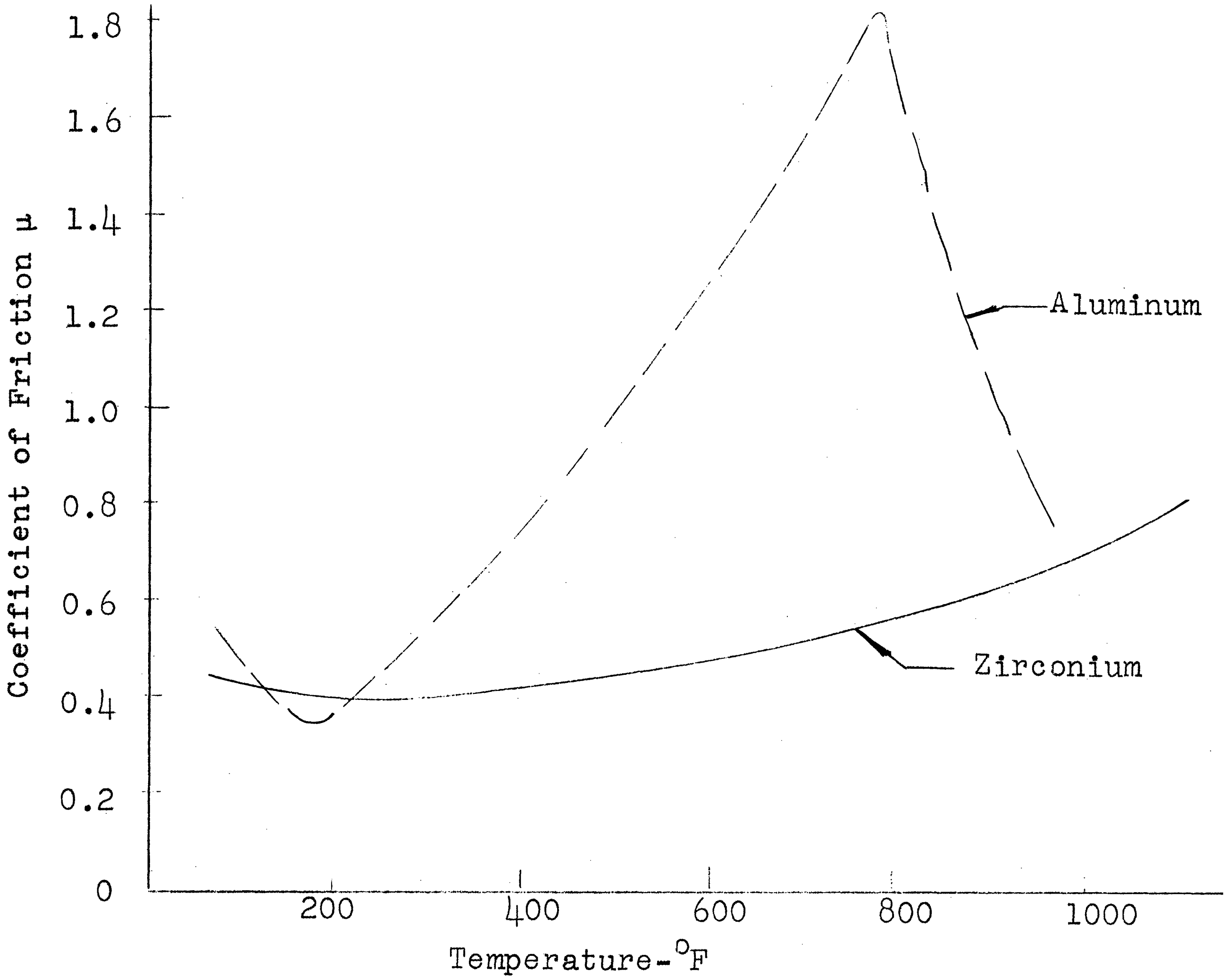


Figure 9
Effect of Temperature on Coefficient of Friction

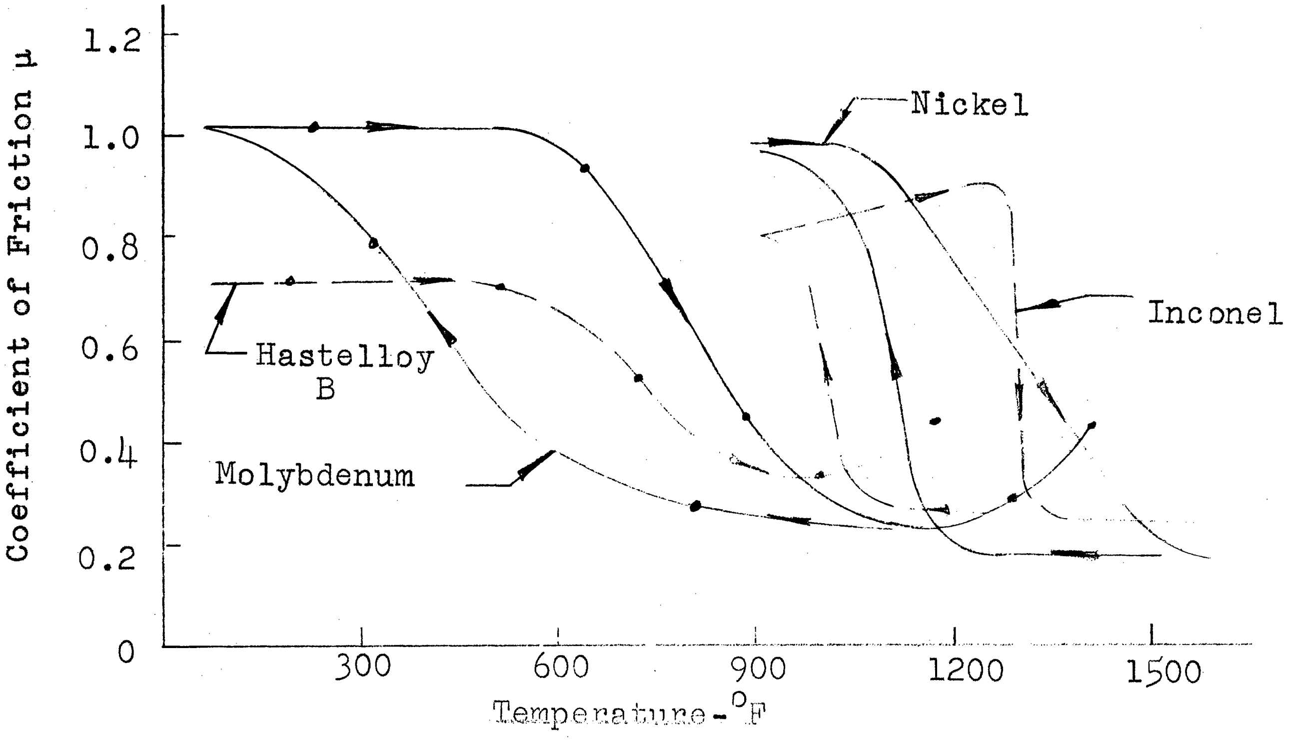


Figure 10
Effect of Temperature on Coefficient of Friction

The overall results indicate that it is necessary to increase the temperature to promote oxide formation in some instances while sufficient oxide is formed at room temperature in other cases. It is also apparent that some of the oxides are stable at low temperatures whereas others show reversibility in the tests indicating the instability of the oxide at low temperature.

For cobalt a rise in friction at the phase change temperature of 788°F has been noted and this is probably caused by the 0.3% volume expansion as cobalt transforms to a face centered structure from a hexagonal close packed structure. This expansion may have caused the oxide film to break up. The fact that on cooling the friction did not rise indicates this since the film would be compresssed as the shrinkage to a H.C.P. structure took place. From this it appears that a third criteria for wear control must be added. To avoid removal of the oxide film the metal or alloys used for sliding should not have a phase change anywhere in their operating temperature range.

Rabinowicz and Kingsbury in an apparatus basically similar in effect to that of Peterson and Florek, investigated the high temperature sliding friction and wear of copper, zinc, 1020 steel, and titanium. It was found that for steel and copper the wear rate dropped to a low value at temperatures above 500°F. The coefficient of friction also dropped with rising temperature. For zinc and titanium the coefficient of friction was almost unchanged but the wear rate increased steadily with rising temperatures. The curves are plotted in Figures 11 and 12.

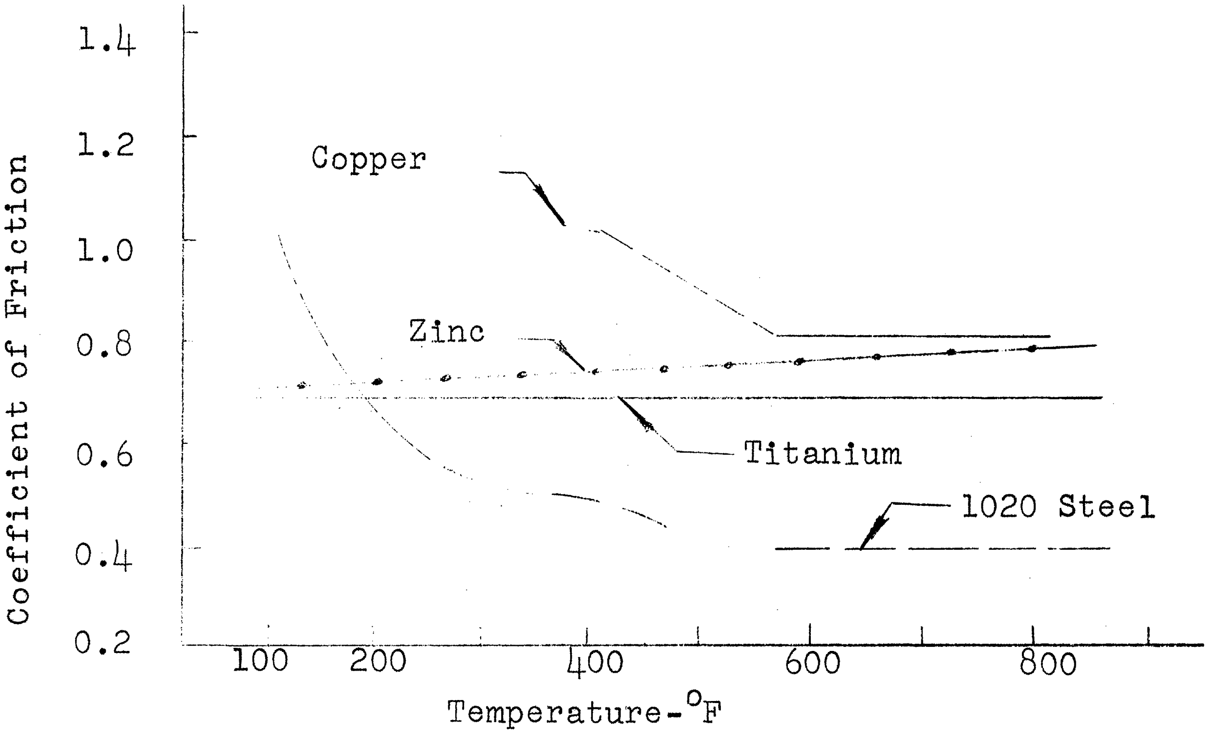


Figure 11

Effect of Temperature on Coefficient of Friction

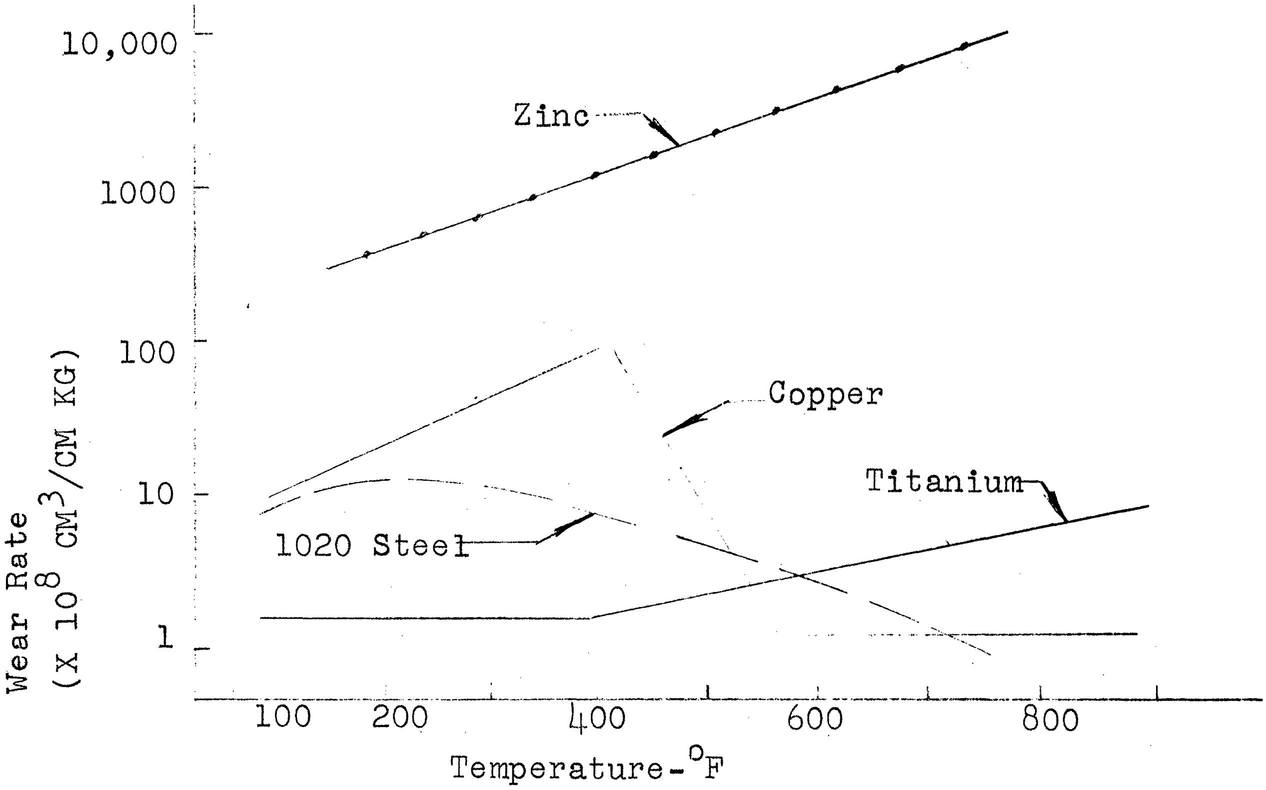


Figure 12

Wear Rates as a Function of Temperature

Formation of an Oxide on Alloys

There are no clear cut rules for determining whether a particular element in an alloy will form a protective oxide film. This subject is being investigated intensely in the field of high temperature oxidation. Such factors as the electron valence of the metals, atom size, and relative diffusion rates of the metals enter to complicate the picture. One tentative relation by Wagner[7] is:

$$N - (\text{Minimum}) = \frac{1}{16\ Z\ C}\left(\frac{\pi\ Kp}{D}\right)^{1/2}$$

N - Mole Fraction of ME

D - Diffusion rate of ME

Z - Valence of metal ME

C - Number of gram atoms of ME/unit volume

16 - The atomic weight of oxygen

Kp - Oxidation rate of <u>pure</u> ME in terms of weight increase $= \frac{M^2}{t}$

Above N(min) ME oxide exclusively can form and will protect the base metal if the oxide film fulfills the conditions for adherency (volume ratio and relative strength) mentioned earlier. For further information on this subject the reader is referred to references 7 and 8.

Thus from the evidence indicated an oxide film can be of great aid in reducing wear if it is adherent. Figures 11 and 12 show that the wear rates for steel and copper drop drastically at 600°F. This has been attributed to the oxide layer formed.[6]

BIBLIOGRAPHY

1. Maurice J. Sinnott, The Solid State for Engineers, John Wiley and Sons, 1958.
2. Robert L. Johnson, Possibilities in the Field of Dry Lubricants, SAE PREPRINT 1958 N18B.
3. D. H. Killeffer, A. Linz, Molybdenum Compounds, Their Chemistry and Technology, Interscience Publishers, New York, 1953.
4. M. B. Peterson and R. L. Johnson, Solid Lubricants for Temperatures to 1000°F., Lubrication Engineering, April 1957.
5. Engineering Approach to Surface Damage, 1958 Summer Conference Course, University of Michigan
6. E. P. Kingsbury, E. Rabinowicz, Friction and Wear of Metals to 1000°C., Trans. J. Basic Eng., Vol.18, Series D, No.2, June 1959.
7. O. Kubaschewski, B. E. Hopking, Oxidation of Metals and Alloys, New York Academic Press, 1953.
8. Benjamin Lustamn, The Resistance of Metals to Oxidation at Elevated Temperature, Metals Handbook, 1948, Pg. 223.

CHAPTER 5

CONTACT LOADING

1. PRESSURE DISTRIBUTION BETWEEN TWO BODIES IN CONTACT

In order to investigate the stresses caused by the contact of two bodies pressed against each other by a force one should first determine the pressure distribtuion between these two bodies. It is then possible to obtain the magnitudes of the stresses at any point of the contact zone.

General Case

According to Hertz the intensity of pressure (p) over the surface of contact is represented by the ordinates of a semi ellipsoid constructed on the surface of contact and given by

$$p = p_o \sqrt{1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}} \qquad 1$$

where (p_o) is the maximum pressure and (a),(b) are the semi axes of the boundary (Figure 1)

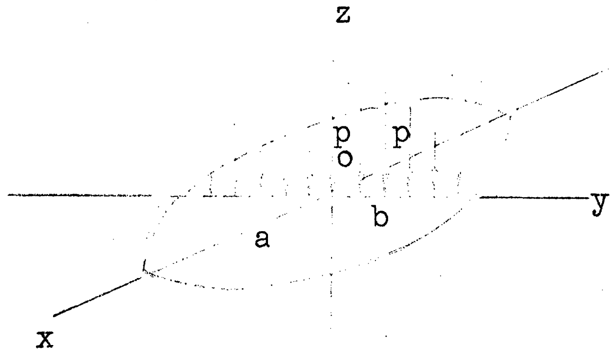


Figure 1
Pressure distribution between two bodies in contact

The maximum pressure (p_o) is clearly at the center of the surface of contact. Since total load (P) is equal to the volume of this semi ellipsoid we can write

$$P = \frac{2}{3}\pi abp_o \qquad 2$$

from which the maximum pressure (p_o) is found to be

$$p_o = \frac{3}{2}\frac{P}{\pi ab} \qquad 3$$

To calculate (p_o) we must know (a) and (b) both of which depend on the shape, applied force and the material properties of the two bodies. (a) and (b) are given by Timoshenko as

$$a = m\sqrt{\frac{3}{4}\frac{P\Delta}{A+B}}$$

$$b = n\sqrt{\frac{3}{4}\frac{P\Delta}{A+B}} \qquad 4$$

where P = Total force

$\Delta = \frac{1-\mu_1^2}{E_1} + \frac{1-\mu_2^2}{E_2}$

μ_1, μ_2 = Poisson's ratio

E_1, E_2 = Modulii of elasticity

$A + B = \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R'_1} + \frac{1}{R_2} + \frac{1}{R'_2}\right)$

$B - A = \frac{1}{2}\left[\left(\frac{1}{R_1} - \frac{1}{R'_1}\right)^2 + \left(\frac{1}{R_2} - \frac{1}{R'_2}\right)^2 + 2\left(\frac{1}{R_1} - \frac{1}{R'_1}\right)\left(\frac{1}{R_2} - \frac{1}{R'_2}\right)\cos 2\psi\right]^{\frac{1}{2}}$

R_1, R'_1, R_2, R'_2 = Minimum and maximum radii of curvatures at the point of contact

ψ = The angle between the planes containg the curvatures $(1/R_1)$ and $(1/R_2)$.

m, n = Constants depending on the ratio $\frac{B-A}{B+A}$

Using the notation $\cos\theta = \frac{B-A}{B+A}$ the values of (m) and (n) are given in the table on the next page.

30°	35°	40°	45°	50°	55°	60°	65°	70°	75°	80°	85°	90°
2.731	2.397	2.136	1.926	1.754	1.611	1.486	1.378	1.284	1.202	1.128	1.061	1.0
0.493	0.530	0.567	0.604	0.641	0.678	0.717	0.759	0.802	0.846	0.893	0.944	1.0

Two spheres in contact

In the special case of two spheres in contact we have

$$A + B = \frac{1}{R_1} + \frac{1}{R_2}$$

$$B - A = 0$$

and the ratio of $\frac{B-A}{B+A}$ is equal to zero, giving $\cos\theta = 0$ and $\theta = 90^\circ$. Thus from the table we have

$$m = n = 1.000$$

Substituting these results into equation (3) and (4) we get

$$a = b = \sqrt[3]{\frac{3}{4}\,\frac{P\,\Delta}{(1/R_1 + 1/R_2)}} \qquad 5$$

$$p_o = 0.578\sqrt[3]{\frac{P\,(1/R_1 + 1/R_2)^2}{\Delta^2}} \qquad 6$$

Sphere on a flat plate

Equations (5) and (6) can also be used for the case of sphere on a flat plate. Taking ($R_2 = \infty$) the above equations become

$$a = b = \sqrt[3]{\frac{3}{4}\,PR_1\Delta} \qquad 7$$

$$p_o = 0.578\sqrt[3]{\frac{P}{R_1^2\,\Delta^2}} \qquad 8$$

Sphere in a spherical socket

For the case of sphere in a spherical socket the sign of R_2 is negative and the equations (5) and (6) for the radius of contact area and the maximum pressure become

$$a = b = \sqrt{\frac{3}{4}\,\frac{P\,\Delta}{(1/R_1 - 1/R_2)}} \qquad 9$$

$$p_o = 0.578 \sqrt[3]{\frac{P\,(1/R_1 - 1/R_2)^2}{\Delta^2}} \qquad 10$$

Two cylinders in contact (Axes parallel)

In the case of two cylinders in contact with the axes being parallel the area of contact approaches a rectangle and the distribution of pressure (p) over this area is represented by a semi-elliptical prism with the equation (Figure 2)

$$p = p_o \sqrt{1 - \frac{y^2}{b^2}} \qquad 11$$

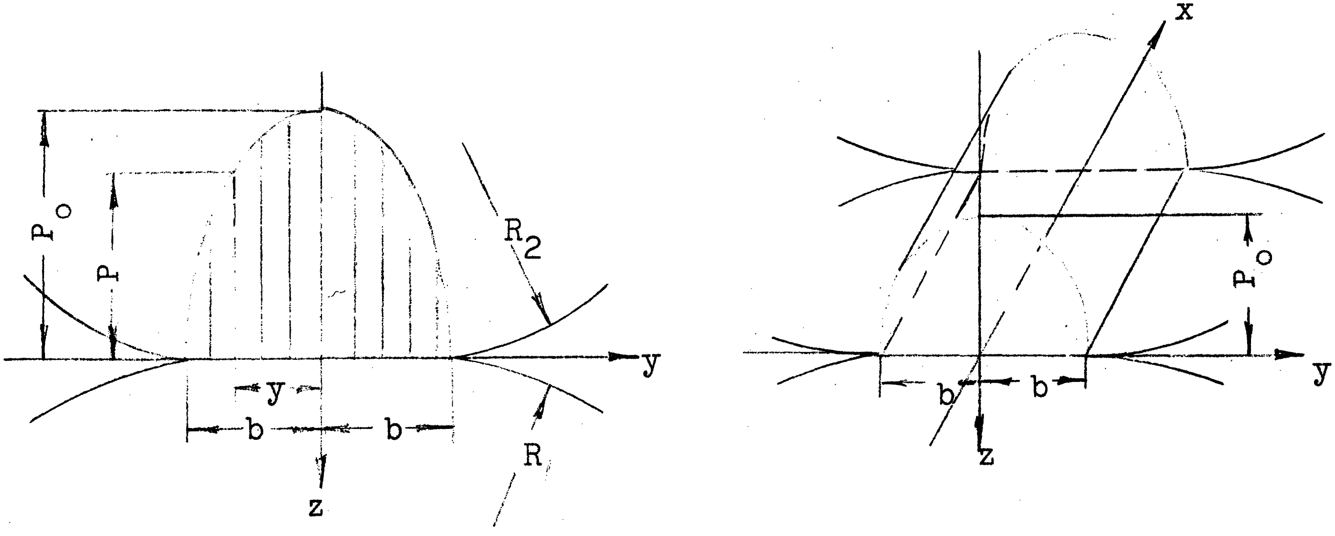


Figure 2

Pressure distribution on contact surface of two cylinders

From the volume of this semi-elliptical prism we have

$$P = \frac{1}{2}\,\pi\, b a p_o \qquad 12$$

where (a) is the length of the cylinders. From the equation (12) we get

$$p_o = 2\,\frac{P}{\pi\, a b} \qquad 13$$

The equation for (b), as given by Timoshenko, is

$$b = 1.13 \sqrt{\frac{P}{a} \frac{\Delta}{(1/R_1 + 1/R_2)}} \qquad 14$$

Substituting this value of (b) into equation (13), the expression for the maximum pressure (p_o) becomes

$$p_o = 0.564 \sqrt{\frac{P}{a} \frac{(1/R_1 + 1/R_2)}{\Delta}} \qquad 15$$

Cylinder on a flat plate

As in the case of spheres, equation (14) and (15) can be used for the case of cylinder on a flat plate. Taking ($R_2 = \infty$) equations (14) and (15) become

$$b = 1.13 \sqrt{\frac{PR\Delta}{a}} \qquad 16$$

$$p_o = 0.564 \sqrt{\frac{P}{aR\Delta}} \qquad 17$$

Cylinder in a circular groove

For the case of cylinder in a circular groove the sign of R_2 is negative and the equations for the width of contact and the maximum pressure become

$$b = 1.13 \sqrt{\frac{P}{a} \frac{\Delta}{(1/R_1 - 1/R_2)}} \qquad 18$$

$$p_o = 0.564 \sqrt{\frac{P}{a} \frac{(1/R_1 - 1/R_2)}{\Delta}} \qquad 19$$

2. STRESS VARIATION IN CONTACT ZONE

After establishing the distribtuion of pressure between the two bodies in contact, we can analyze the variation of stresses in the contact zone. Although, in some cases, only the components of the stresses for the points on (z) axis (where x and y coordinates are zero) will be given, it should be understood that the contact zone includes any point in the body with the coordinates (x,y,z,).

Stresses due to a concentrated normal load (P)

Stresses (in the yz-plane) due to a concentrated normal load (P), as given by E. I. Radimovsky, are (Figure 3).

$$\sigma_y = \frac{2P}{\pi z} \sin^2\theta \cos^2\theta$$

$$\sigma_z = -\frac{2P}{\pi z} \cos^4\theta$$

$$\tau_{yz} = -\frac{2P}{\pi z} \sin\theta \cos^3\theta$$

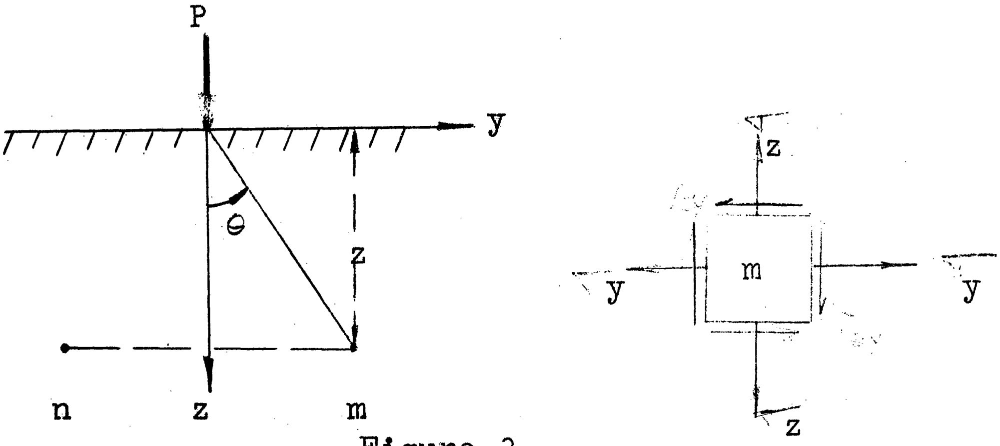


Figure 3

Single concentrated force acting perpendicular to a plane

In Figure 4 the values of (σ_z) and (τ_{yz}) are plotted for points in the yz-plane located at a depth (z) below the surface. These values show how the stresses vary along the line m-n due to the concentrated load, (P). It can be seen that the shear stress, (τ_{yz}), has different sign on each side of (z) axis. Thus,

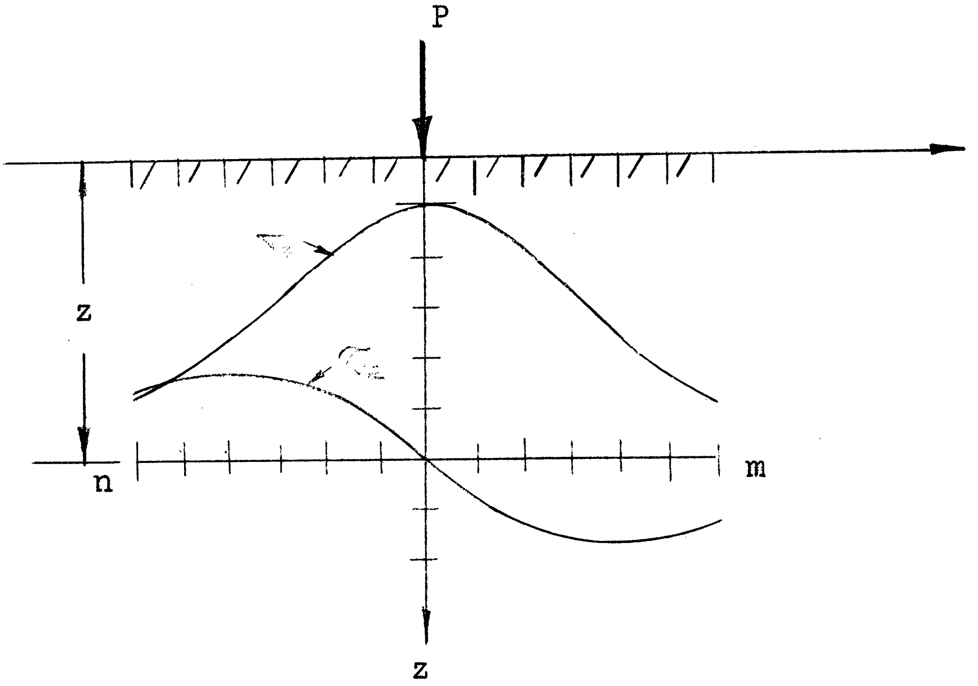


Figure 4

Variation of stresses ($_{z}$) and ($_{yz}$)
due to concentrated load P.

if the load (P) was moving on the surface, the shear stress (τ_{yz}) at a fixed point in the yz-plane would first increase, then become zero as the load passes over the point, and then change its sign and increase in magnitude.

Stresses due to concentrated tangential load (Q)

According to J. O. Smith and C. K. Liu the equations for the stresses in yz-plane due to concentrated tangential load (Q) are

$$\begin{aligned} \sigma_y' &= -\frac{2Q}{\pi z} \cos^3\varphi \sin\varphi \\ \sigma_z' &= -\frac{2Q}{\pi z} \cos\varphi \sin^3\varphi \\ \tau_{yz}' &= -\frac{2Q}{\pi z} \cos^2\varphi \sin^2\varphi \end{aligned} \qquad 21$$

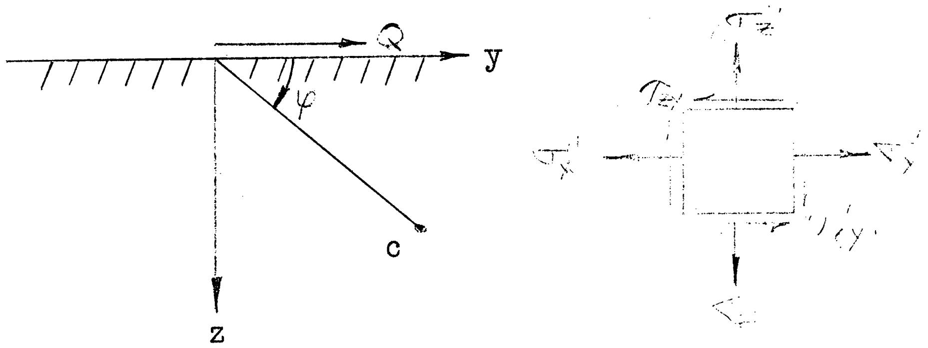


Figure 5
Single concentrated tangential force on a plane

It can be seen from the above equations that (τ_{yz}) and (σ'_z) have identical forms, so do (σ_y) and (τ'_{yz}). Thus, Figure 4 can be used for (σ'_z) which shows that in the case of concentrated tangential loading (σ'_z) changes sign. From equations (21) above it can also be seen that as (φ) varies from 0° to 180° (σ'_y) also changes sign but (τ'_{yz}) does not.

Stresses due to contact of two bodies of general shape

Stress (on the z-axis) due to contact of two bodies of general shape are computed by H. R. Thomas and V. A. Hersch. The magnitudes of these stresses are plotted in Figure 6 for a (b/a) ration of 0.86609. Since (because of the symmetry for points on z-axis) these stresses are principal stresses the shear stress components τ_{xy}, τ_{xz}, and τ_{yz} are zero. But on planes bisecting the angles between y and z, x and z, and x and y axes there are shear stresses which are given by

$$\begin{aligned} \tau_{xz}(45^\circ) &= \tfrac{1}{2}(\sigma_x - \sigma_z) \\ \tau_{xy}(45^\circ) &= \tfrac{1}{2}(\sigma_x - \sigma_y) \qquad 22 \\ \tau_{yz}(45^\circ) &= \tfrac{1}{2}(\sigma_y - \sigma_z) \end{aligned}$$

Since for the particular case shown in Figure 6 τ_y $(45^\circ)=$ $\tau_{xz}(45^\circ)$ and $\tau_{xy}(45^\circ) \cong 0$, only $\left[\tau_y \; (45^\circ)\right]$ is plotted. As it is seen $(\tau_{yz}{}^{45})$ is a maximum at a depth of about 0.44a.

Stresses due to contact of two spheres

The equations for the stresses, for points on the z axis, due to contact of two spheres as computed by H. R. Thomas and V. A.. Hersch are given below.

$$\sigma_x = \sigma_y = \frac{2a\left(\frac{1}{R_1} + \frac{1}{R_2}\right)}{\pi \, \Delta} \left[(1+\mu)\left(\frac{z}{a} \cot^{-1} \frac{z}{a} - 1\right) + \frac{1}{2} \, \frac{a^2}{a^2 + z^2}\right] \qquad 23$$

$$\sigma_z = - \frac{2a\left(\frac{1}{R_1} + \frac{1}{R_2}\right)}{\pi \, \Delta} \left(\frac{a^2}{a^2 + z^2}\right)$$

Since (σ_x) and (σ_y) are equal equations (22) become

$$\tau_{xz}(45^\circ) = \tau_{yz}(45^\circ) = \frac{1}{2} \, (\sigma_x - \sigma_z)$$
$$\tau_{xy}(45^\circ) = 0$$

Variations of these stresses are plotted for various depths in Figure 7. The magnitudes of shear stress components $\tau_{xz}(45^\circ)$ and $\tau_{yz}(45^\circ)$ become maximum at a depth equal to about half of the radius of the surface of contact. The magnitude of this stress (for $\mu = 0.3$) is about $0.31p_0$.

Stresses due to contact of two cylinders

In the case of two cylinders in contact two different motions exist; one is pure rolling, and the other is combined rolling and sliding. It has been proved that the stresses due

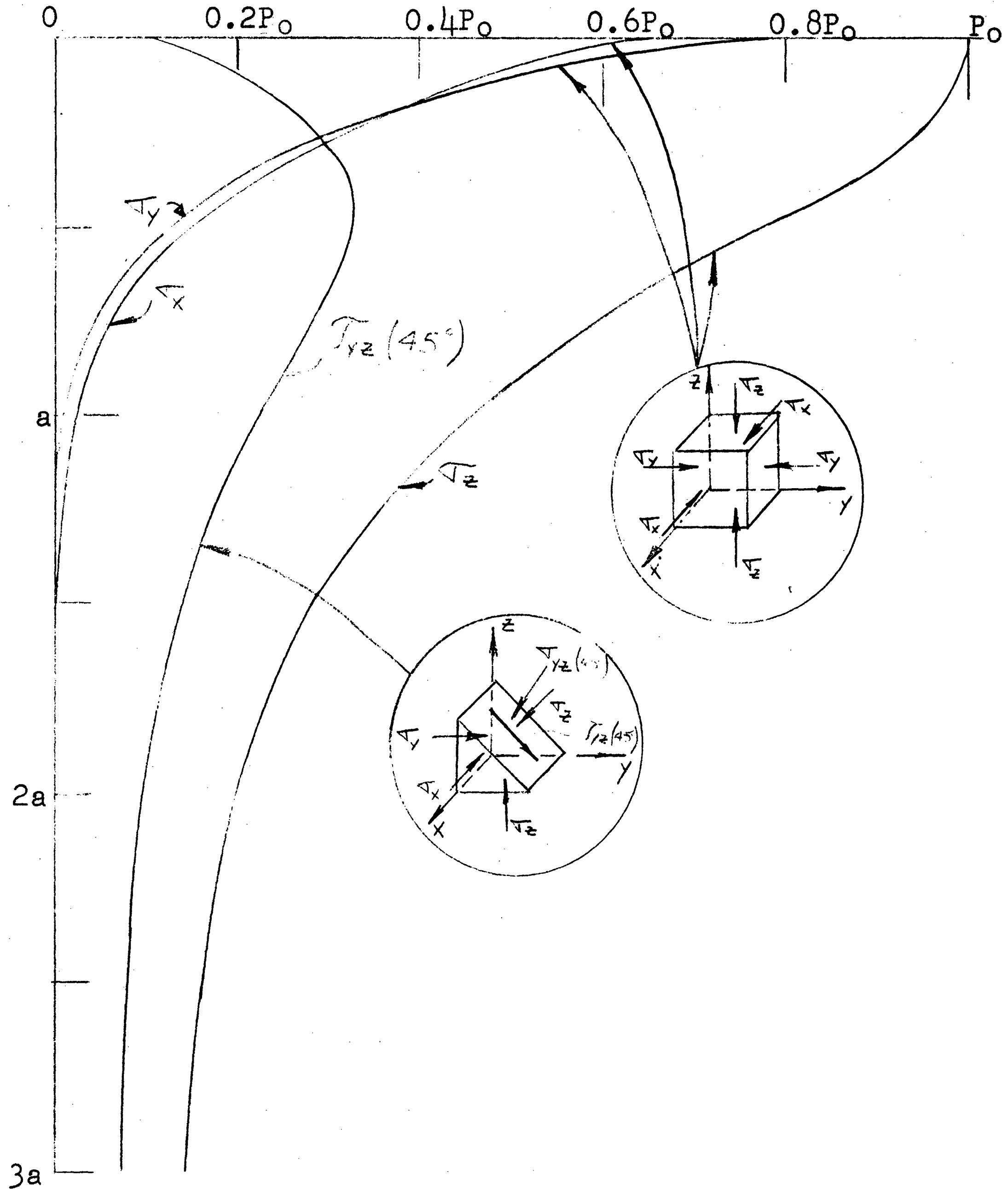


Figure 6

Variations in stress components due to contact of two bodies of general shape (b/a=0.86609) with the distance from contact surface.

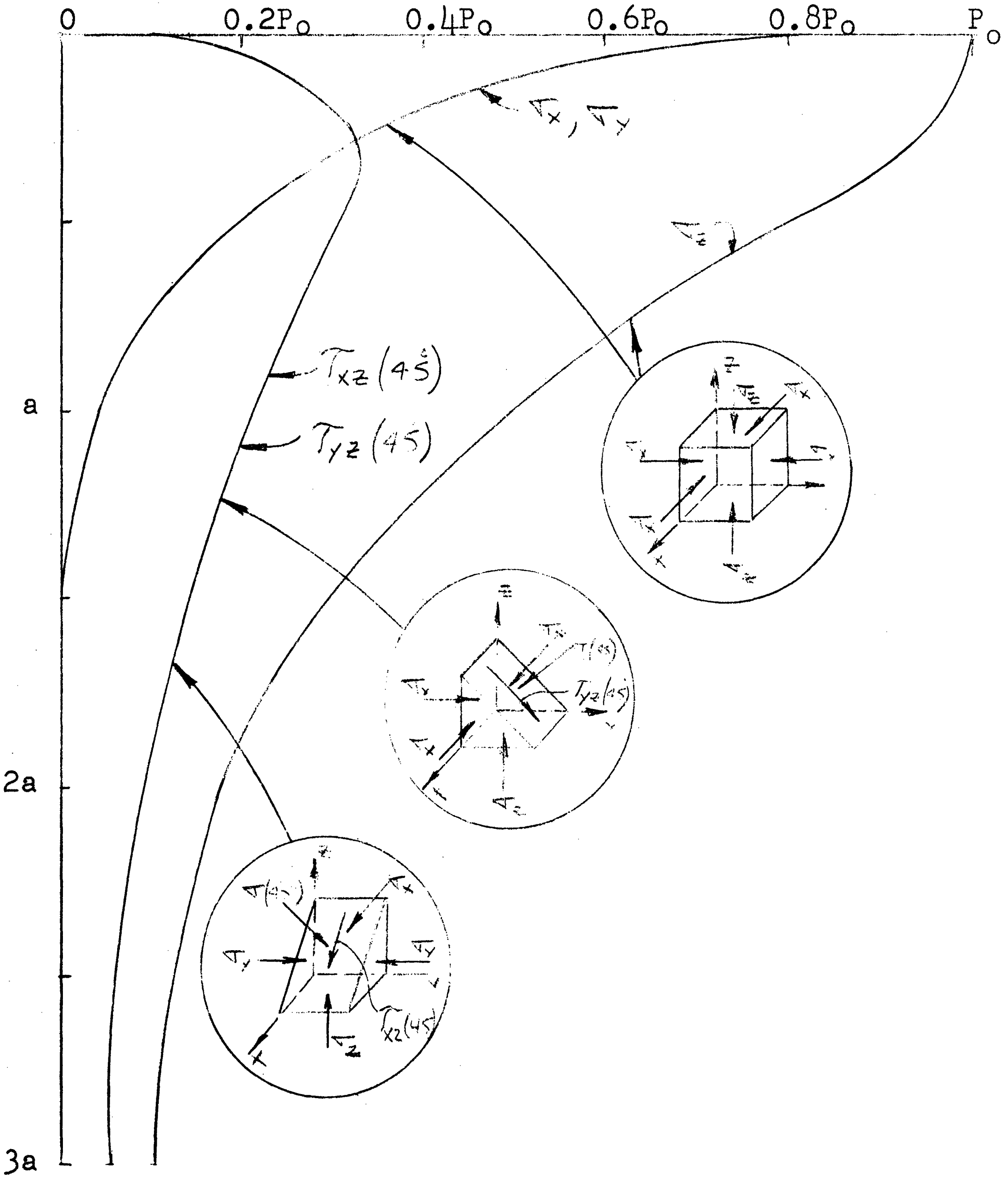


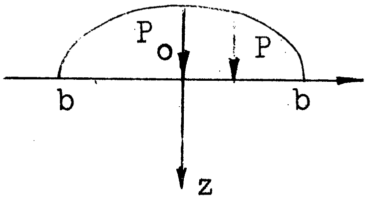
Figure 7

Variations in stress components due to contact of two spheres with the distance from contact surface.

to combined rolling and sliding can be obtained by the superposition of the stresses due to pure rolling, (normal load alone), and pure sliding,(tangential load alone). Hence, we can examine these two cases separately and superimpose them to obtain the stress distribution due to combined rolling and sliding action.

Case I. Stresses due to normal load with an elliptical distribution.

As it was shown in Section 1 pressure distribution between two cylinders in contact is represented by a semi-elliptical prism and given by equation (11).



$$P = P_o\sqrt{1 - \frac{y^2}{b^2}}$$

Figure 8

Elliptical distribution of normal load.

The calculation of stresses in contact zone due to normal load that is distributed as given in equation (11) (Figures 2 & 8) results in a stress variation as shown in Figure 9, 10, 11, 12 and 13. Figure 9 is drawn to show the stresses for points on z-axis. It is similar to the Figures 6 and 7 that were plotted for the cases of two bodies of general shape and two spheres. In the case of two cylinders, because of the differences in the magnitudes of normal stresses σ_{x_n}, σ_{y_n}, and σ_{z_n} the shear stress components on the bisecting planes differs from each other. Since the magnitude of the shear stress on the plane bisecting the axes y and z is greatest, it is the most important one, and as shown in Figure 9 this component (τ_{yz_n}(45°)(*)) has a maximum at a depth of (0.786b) and the magnitude of this maximum shear is about ($0.3p_o$).

Figures 10,11,12, and 13, show the variation of stresses at a point (with the coordinates x,y,z), in the contact zone. As it is seen from these figures, except the shear stress component (τ_{yz_n}), the magnitudes of normal stresses σ_{x_n}, σ_{y_n}, and σ_{z_n} are smaller for the points away from the load and are maximum for the points under the load (These maximums are plotted in Figure 9). As it was in the case of the concentrated normal load, the shear stress componet (τ_{yz_n}) changes sign.

Instead of being stationary, if the load was moving at a fixed point in the body the normal stresses would have zero magnitude at first and as the load approaches the magnitudes of these stresses it would increase, become maximum as the load passes over the point and then decrease and become zero as the load moves further away from the point. In the case of shear stress component (τ_{yz_n}) this is different, it changes sign as the load moves over the surface. The coordinates of points where this stress reaches a maximum are ($z = 0.5b$) and ($y = + 0.85b$), and the magnitude is ($0.256p_o$). Hence it is clear that if the load is moving the most critical stress is the shear stress

---

(*) Subscript (n) denotes stresses due to normal load, (t) denotes stresses due to tangential load, and superscript (*) denotes stresses due to combined normal and tangential load.

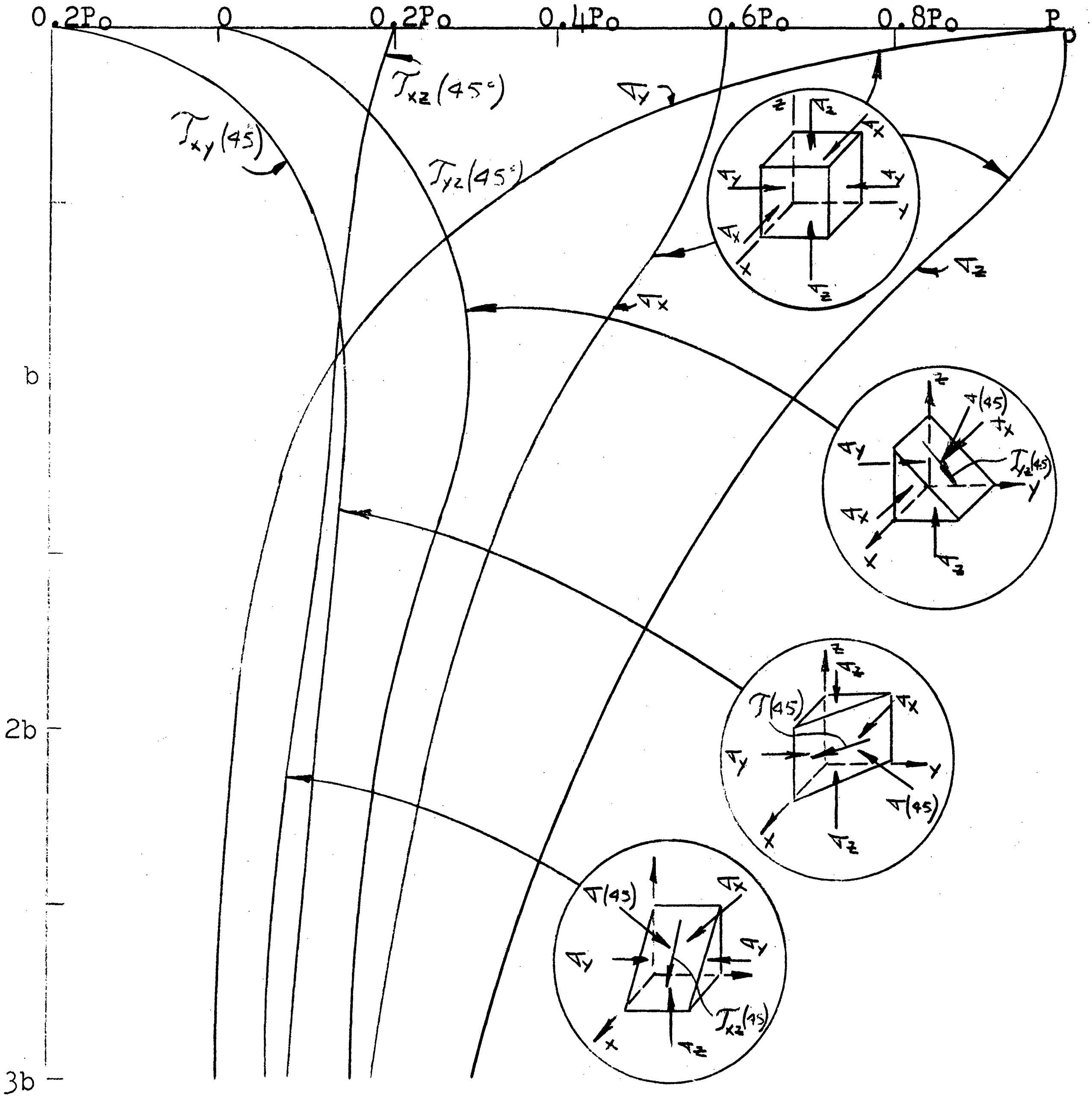


Figure 9

Variations in stress components due to contact of two cylinders with the distance from contact surface.

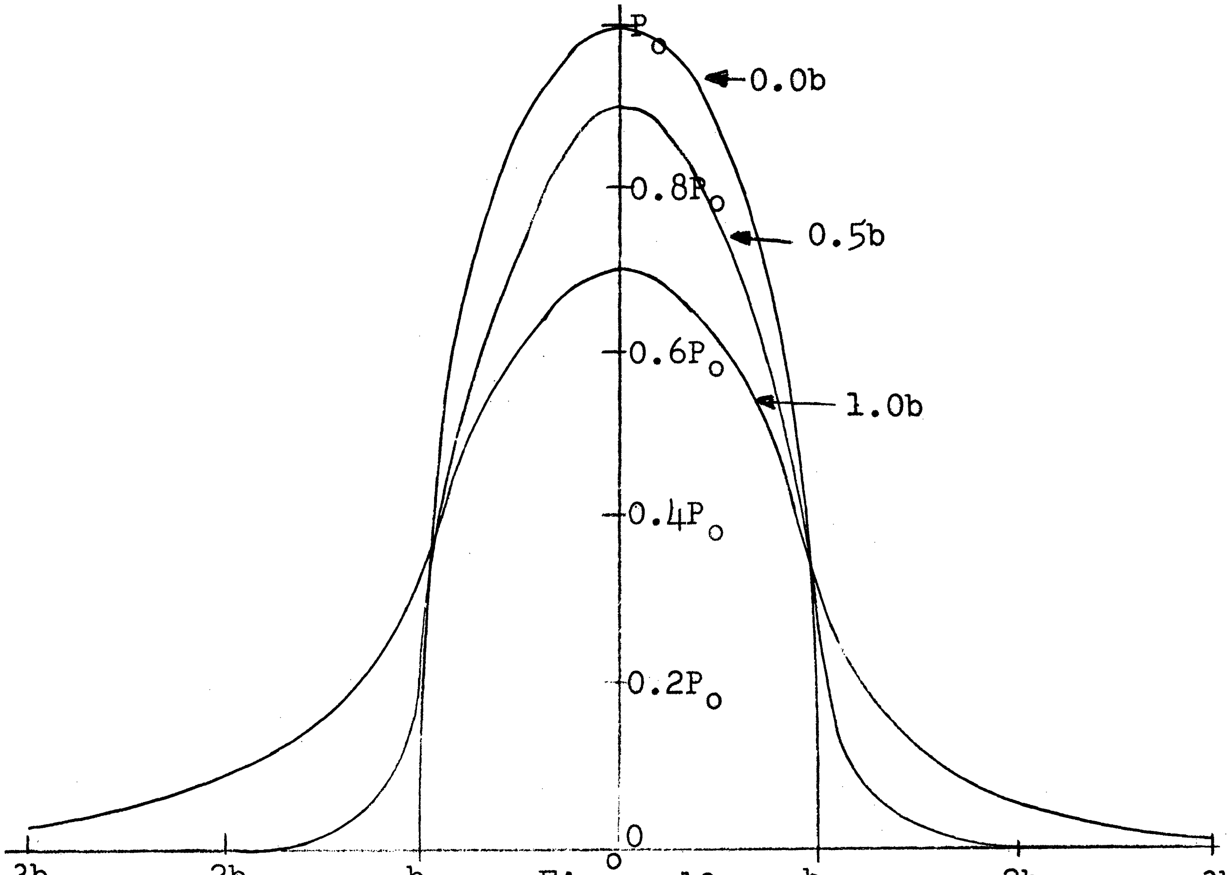


Figure 10

Curves showing the values of stress (σ_{z_n}) for various values of y, at the surface and at various depths below surface. (Figure 8 for loading)

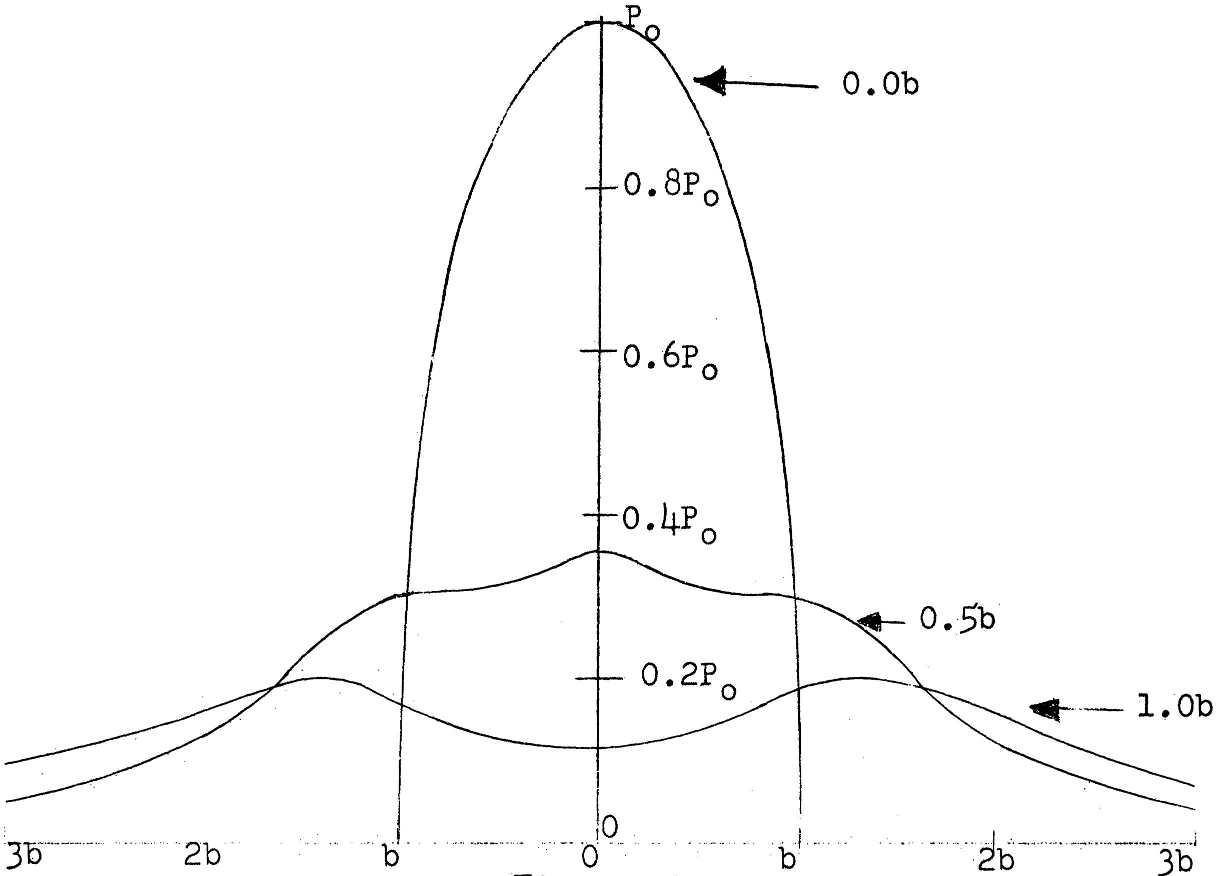


Figure 11

Curves showing the values of stress (σ_{yn}) for various values of y, at the surface and at various depths below surface. (Figure 8 for loading)

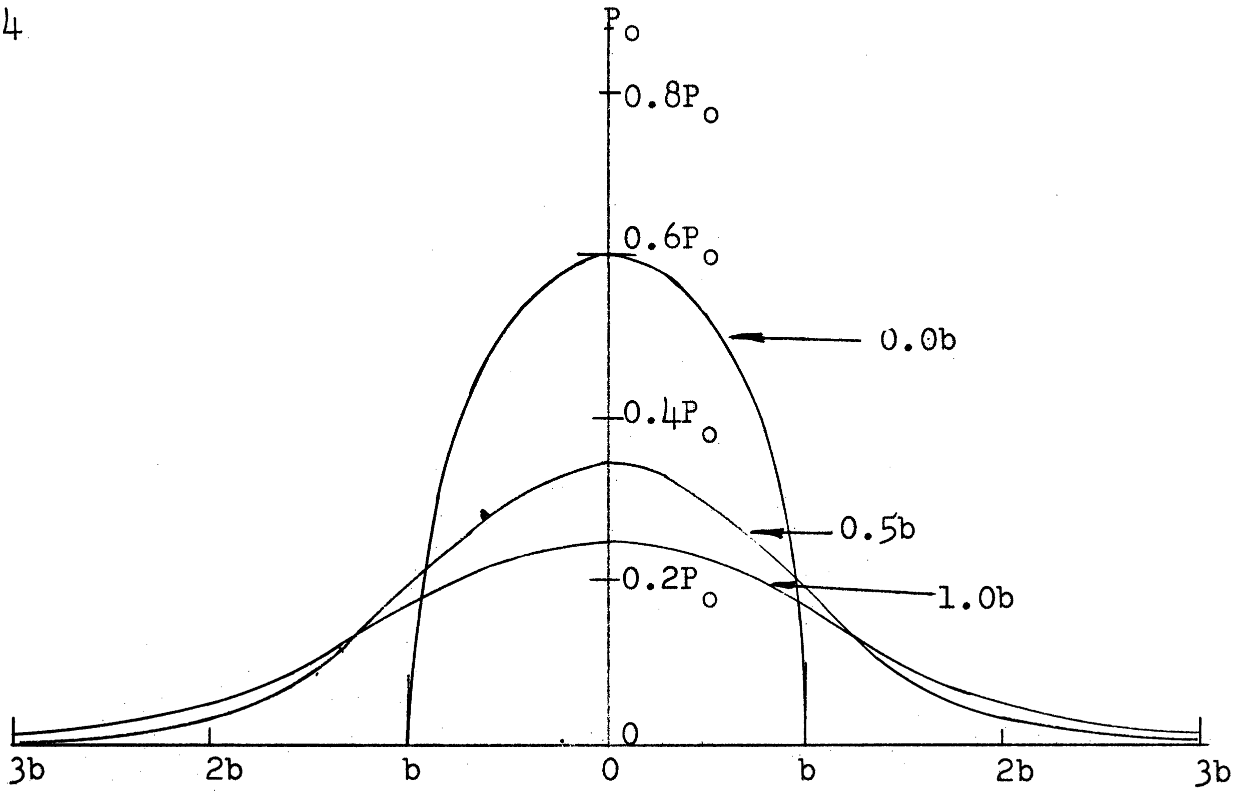


Figure 12

Curves showing the values of stress (σ_{x_n}) for various values of y, at the surface and at various depths below surface. (Figure 8 for loadin

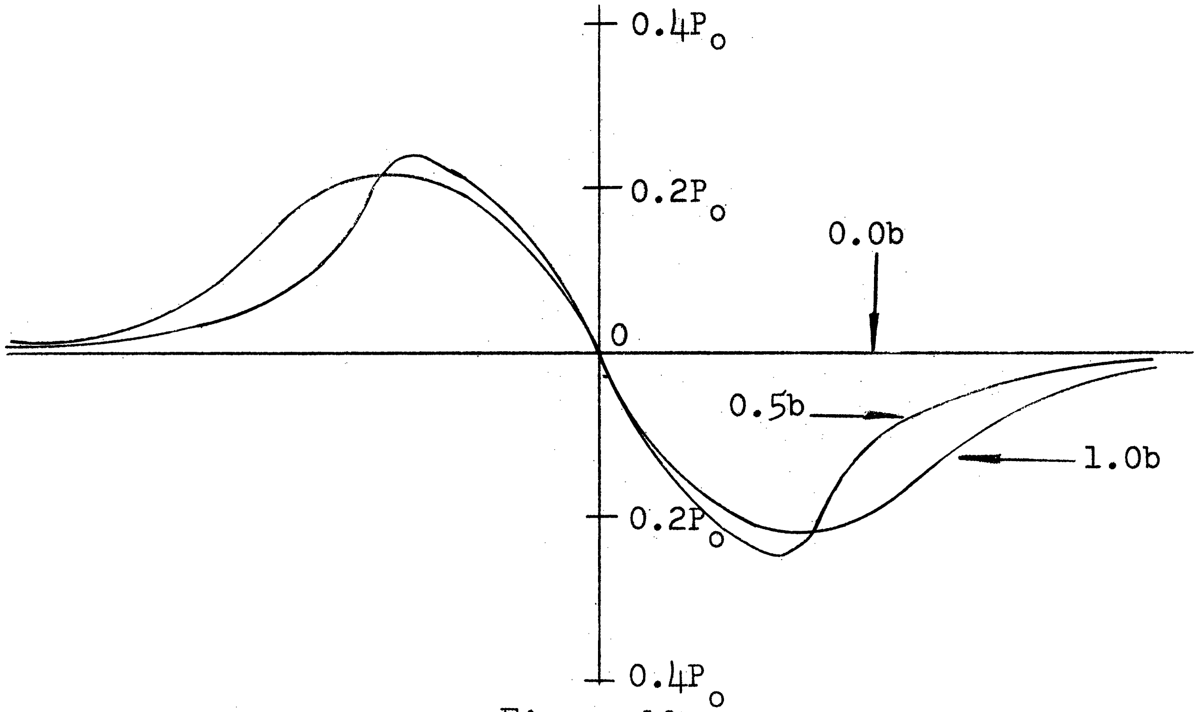
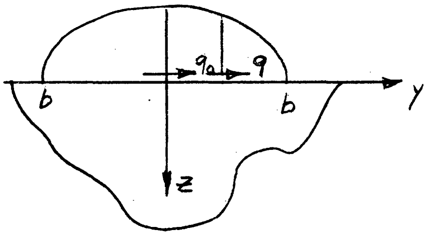


Figure 13

Curves showing the values of stress (τ_{yz_n}) for various values of y, a the surface and at various depths below surface.(Figure 8 for loading component (τ_{yz_n}) which changes sign and thus, has a complete stress reversal compared to the $\tau_{yz_n}(45^o)$ under the load which has a smaller range.

Case II. Stresses due to tangential load with an elliptical distribution



$$q = q_o \sqrt{1 - \frac{y^2}{b^2}}$$

Figure 14

Elliptical distribution of tangential load

If we assume a load that is distributed as shown in Figure 14 stress variation in the contact zone due to this load distribution (as given by J. O. Smith and C. K. Liu), would be as shown in Figure 15,16 and 17. Because of the similarity in stress equations it can be seen that (τ_{yz_n}) and (σ_{z_t}) have identical forms, so do (σ_{y_n}) and (τ_{yz_t}). Hence in this kind of loading the stresses that change sign are (σ_{y_t}) and (σ_{z_t}).

Case III. Stresses due to combined sliding and rolling action

It was proved by Mindlin that when rolling and sliding occur together ($q_o = fp_o$) where (f) is the coefficient of friction of the surfaces ~~of~~ contact. Hence in order to determine the stresses due to combined rolling and sliding action we can multiply the stresses given in Figure 15,16 and 17 by the coefficient of friction thereby expressing them in terms of (p_o) and add them to the

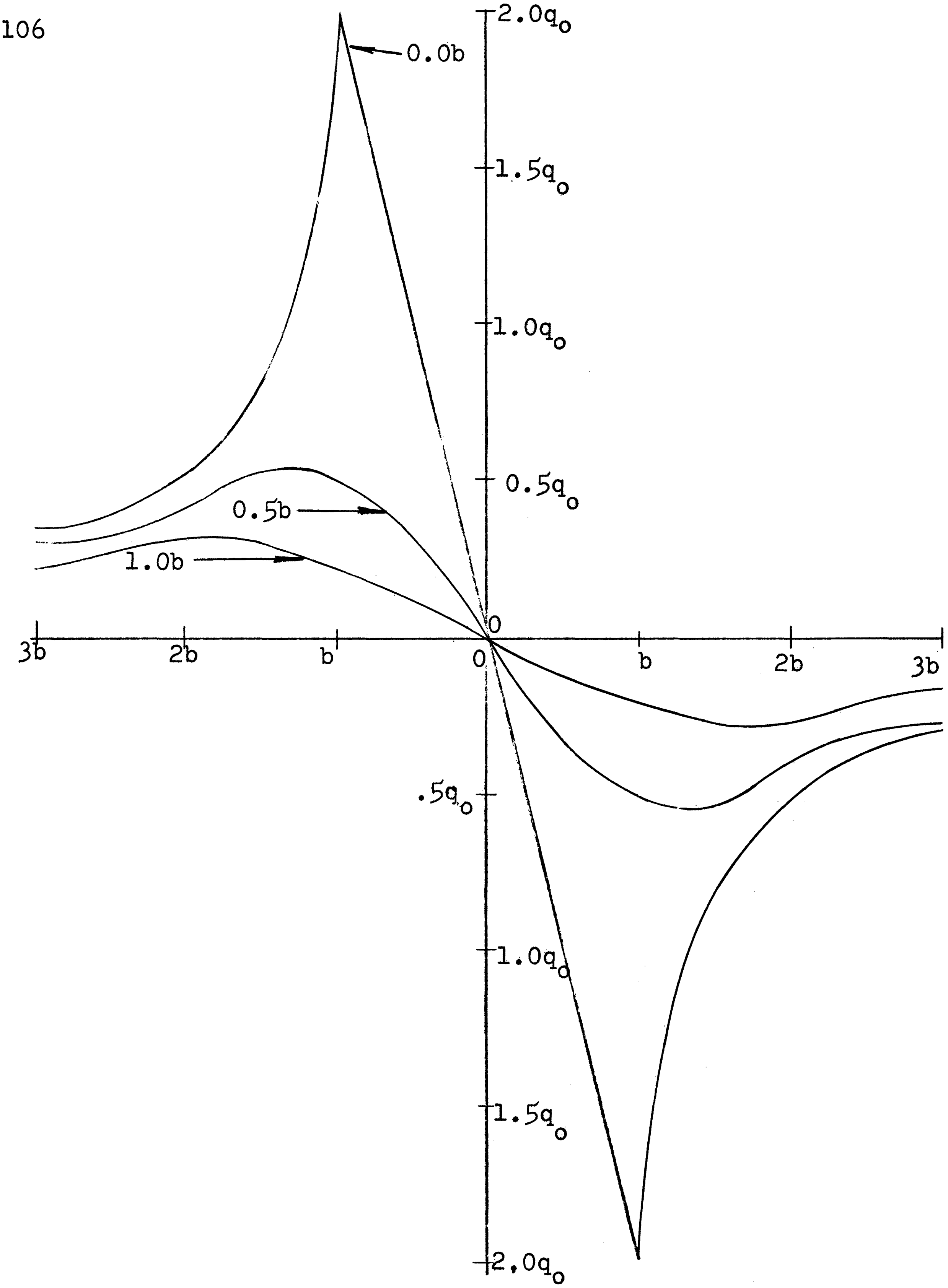


Figure 15

Curves showing the values of stress (σ_{y_t}) for various values of y, at surface and at various depths below surface. (Figure 14 for loading)

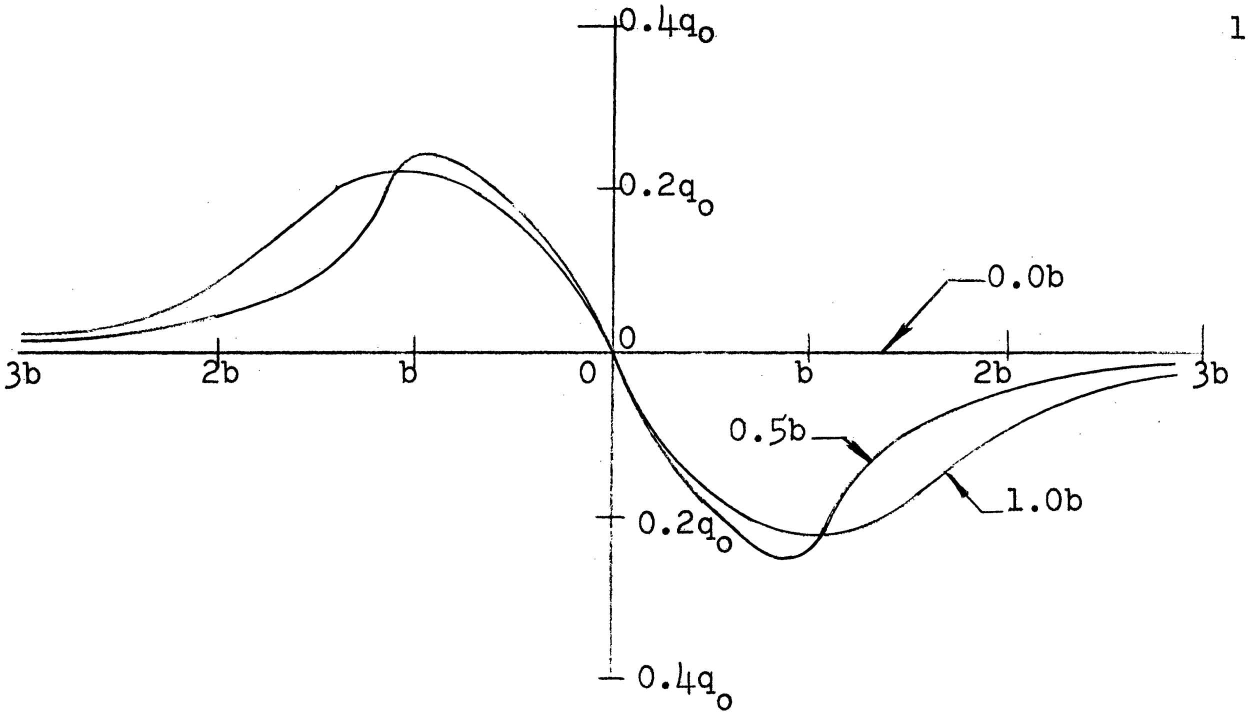


Figure 16

Curves showing the values of stress (σ_{z_t}) for various values of y, at the surface and at various depths below surface.(Figure 14 for loading)

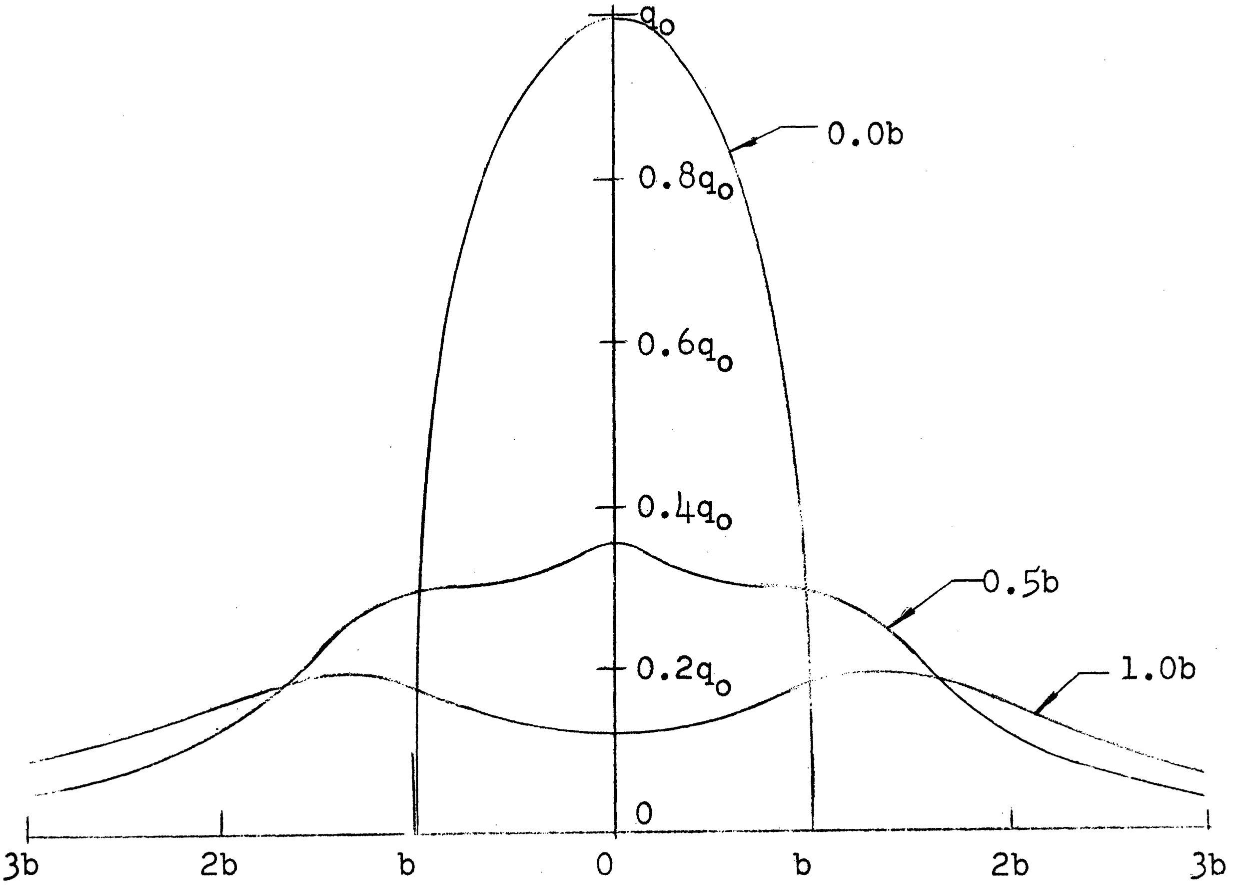


Figure 17

Curves showing the values of stress (τ_{yz_t}) for various values of y,at the surface and at various depths below surface. (Figure 14 for loading

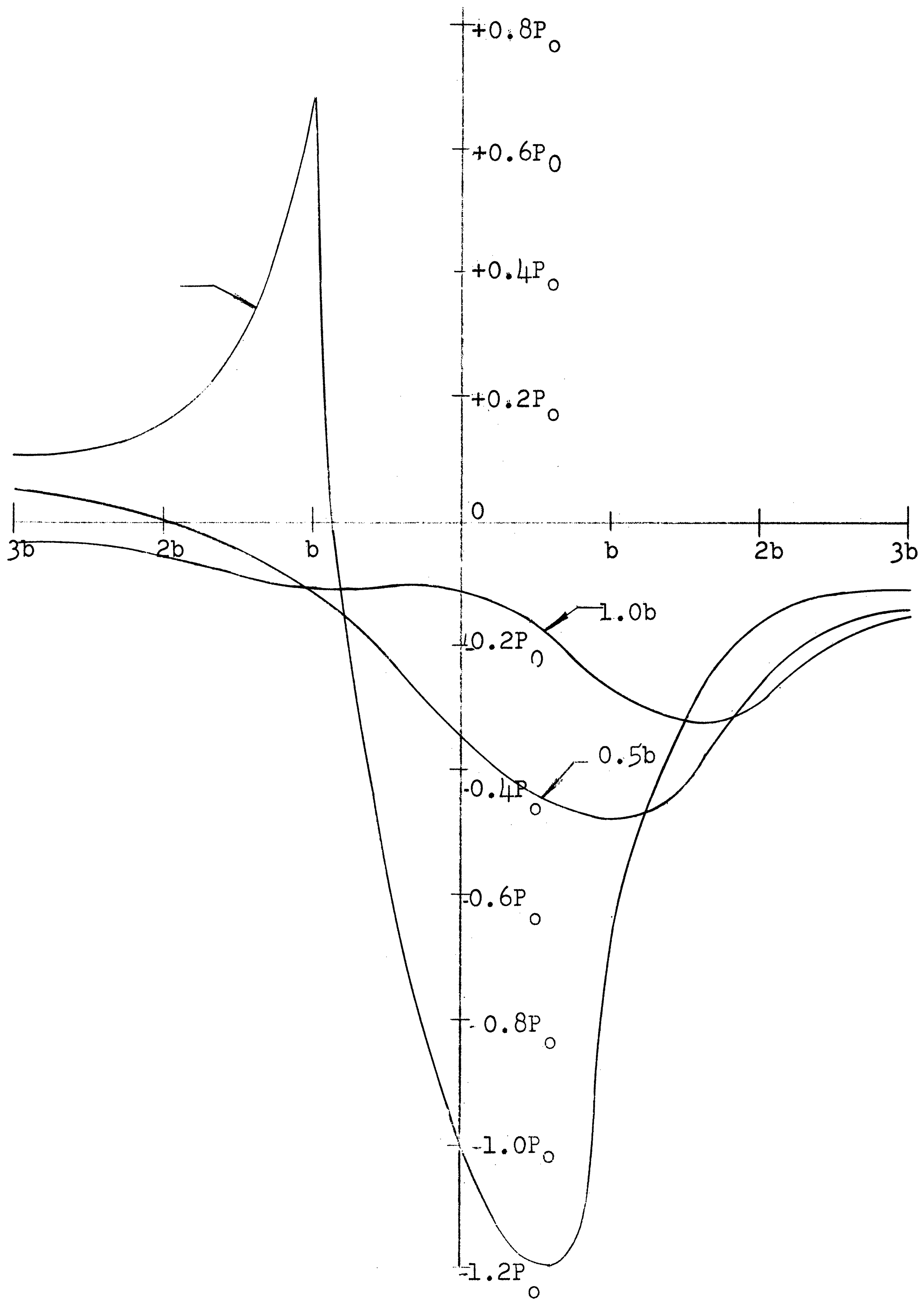


Figure 18

Curves showing the values of stress (σ_y^*) for various values of y, at the surface and at various depths below surface. (f = 1/3)

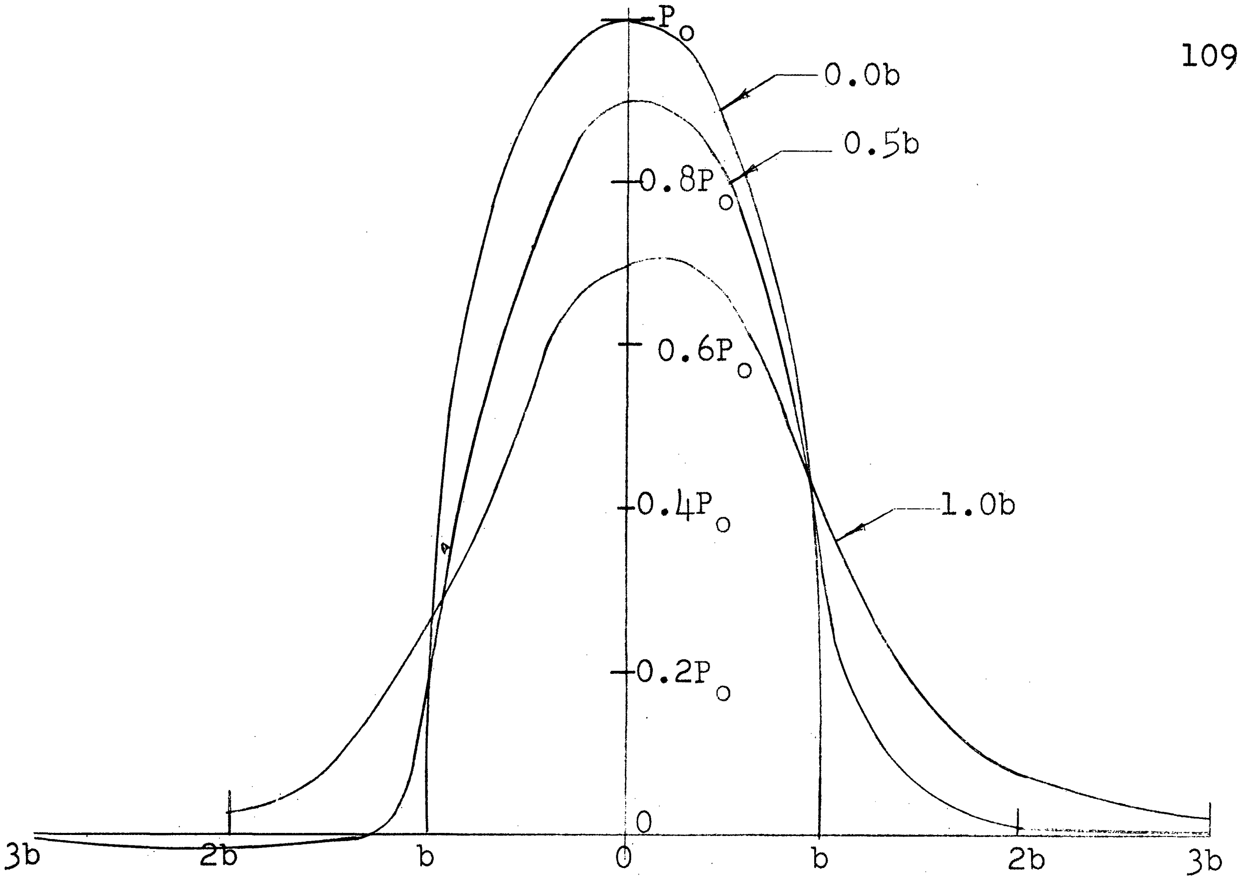


Figure 19

Curves showing the values of stress (σ_z^*) for various values of y, at the surface and at various depths below the surface. (f = 1/3)

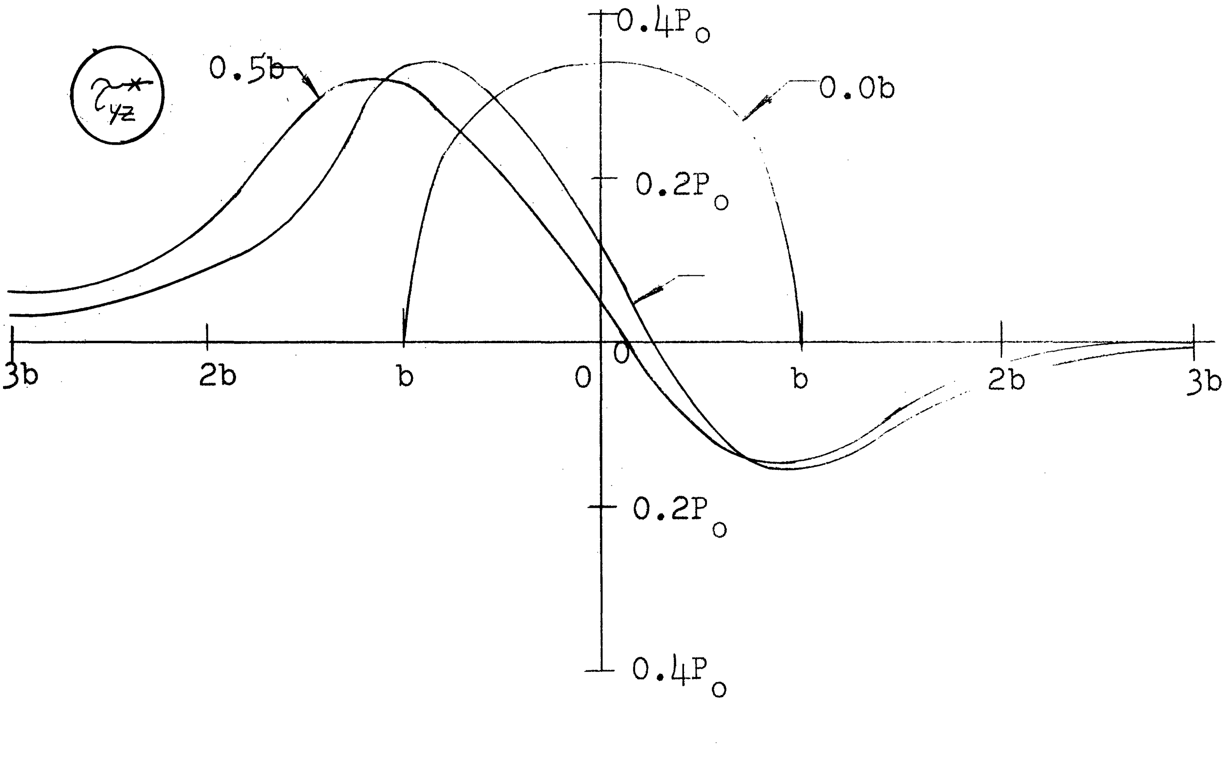


Figure 20

Curves showing the values of stress (τ_{yz}^*) for various values of y, at the surface and at various depths below the surface. (f = 1/3) stresses given in Figure 10,11 and 13. This done for a coefficient of friction of (1/3) and stress variations shown in Figure 18, 19 and 20 were obtained.

As it can be seen from Figure 18 that the normal stress (σ_y^*) changes from $0.67p_o$ tension to $-1.20p_o$ compression. When the range of this normal stress is compared with the range that occurs when there is no friction force, which is from zero to $-1.0p_o$ in compression, it is noted that there is an 87% increase in magnitude ($f = 1/3$), and there is a reversal of stress which is of considerable importance in explaining how a progressive crack starts and propagates. In Figure 21 (σ_y^*) is again plotted for various coefficients of friction.

From the theory of elasticity we can also find (τ_x^*) which is given by

$$\sigma_x^* = \mu(\sigma_z^* + \sigma_y^*)$$

and by making use of Mohr's circle we can determine three principal stresses at a point with the coordinates (y and z). Calling these principal stress σ_1^*, σ_2^*, and σ_3^* in decreasing order in magnitude, from theory of elasticity we have

$$\tau_{max}^* = \frac{1}{2}(\sigma_1^* - \sigma_3^*)$$

Therefore, knowing the principal stresses at each point we can determine the maximum shearing stress at that point which is on a plane bisecting the directions of maximum (σ_1^*) and minimum (σ_3^*) principal stresses. When this is done it is found that at a point A ($z = 0$, $y = 0.3b$) (See Figure 22) the maximum values

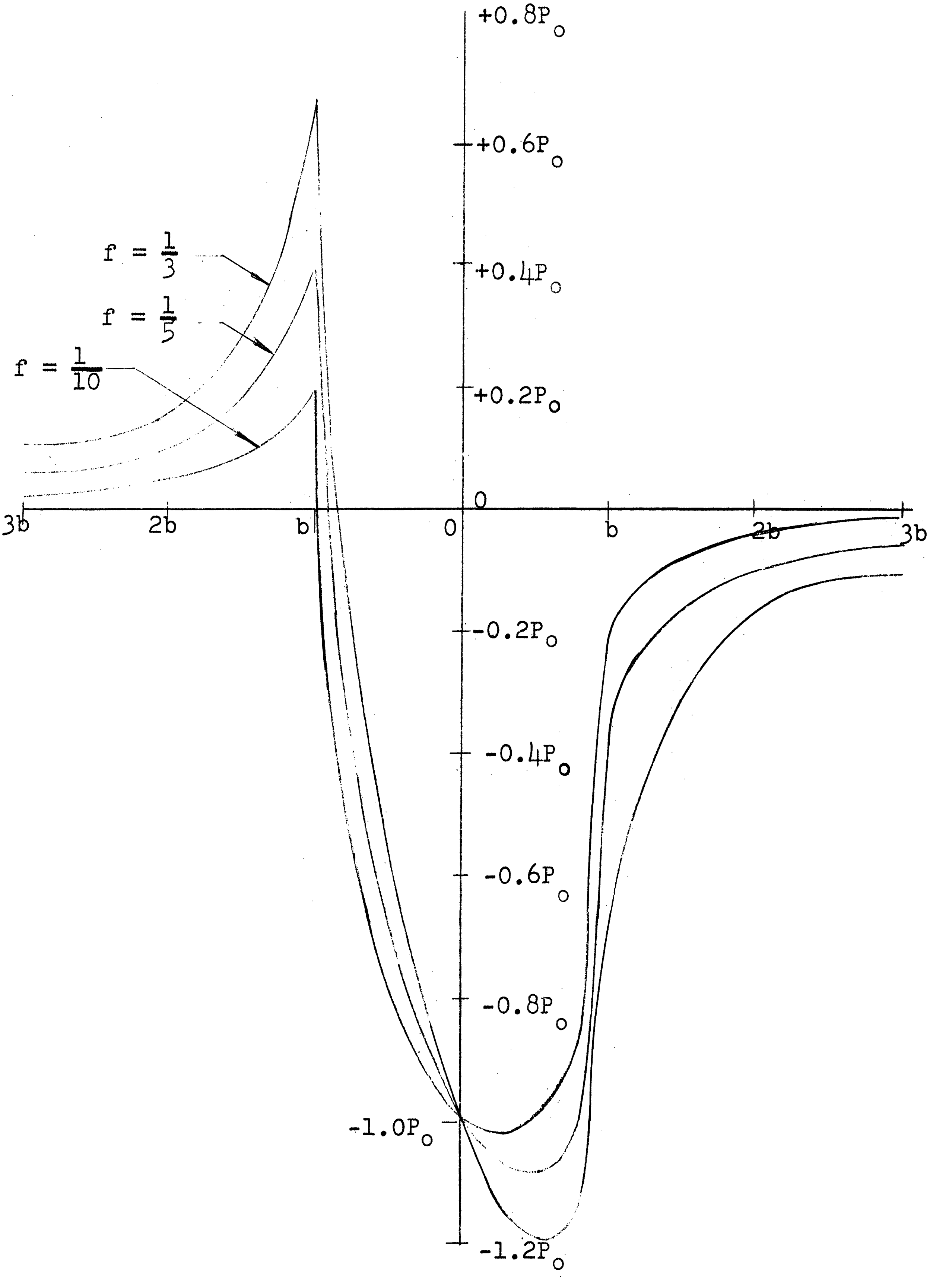


Figure 21
Curves showing the values of stress (σ_y^*) at surface (z = 0) for various values of y, and coefficient of friction f.

of three principal stresses occur (for $f = 1/3$, $\mu = 0.25$). These values of principal stresses are

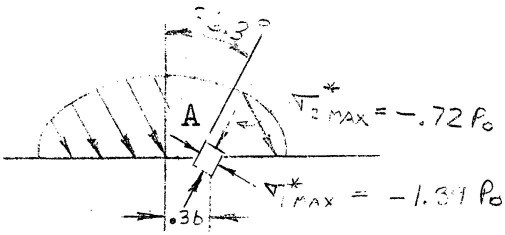


Figure 22
Location of maximum principal stresses

$$\sigma^*_{1_{max}} = -1.39p_o$$

$$\sigma^*_{2_{max}} = -0.72p_o$$

$$\sigma^*_{3_{max}} = -0.53p_o$$

By using equation (24) it is found that the maximum value of (τ^*_{max}) also occurs at the point A and has magnitude

$$\tau^*_{max} = 0.43p_o$$

It is clear that the magnitude of this stress is greater than the magnitude of maximum shearing stress that is found when no friction exists. The magnitude of this stress was ($\tau_{max_n}=0.30p_o$) and it was occuring at a depth 0.786b underneath the surface. Therefore tangential force caused by a coefficient of friction of 1/3 increases the maximum shearing stress by 43%. Furthermore, the location of the maximum shearing stress is changed by the addition of tangential force from beneath the surface to the point A in the surface. It was found by J. O Smith and C. K. Lui that when the coefficient of friction becomes greater than (1/9) the maximum shearing stress occurs at a point in the surface, but when (f) is less than (1/9) this stress is underneath the surface.

BIBLIOGRAPHY

1. Roark, R. J., Formulas for Stress and Strain, McGraw-Hill Book Compancy, Inc., New York , 1954.
2. Buckingham, E., Analytical Mechanics of Gears, McGraw-Hill Book Company, Inc., New York, 1949.
3. Timoshenko, S., Strength of Materials, D. Van Nostrand Company, Inc., New York, 1955.
4. Timoshenko, S., and Goodier, J. N., Theory of Elasticity, McGraw-Hill Book Company, Inc., New York, 1951.
5. Way, Steward,"Pitting Due to Rolling Contact", Trans. ASME, Vol. 57, 1935.
6. 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", Jour. of Applied Mechanics, Vol. 20, June 1953.
7. Radimovsky, E. I., "Stress Distribution and Strength Condition of Two Rolling Cylinders Pressed Together", University of Illinois Engineering Experiment Station Bulletin Series No. 408, Vol. 50, No. 44, 1953.
8. Buckingham, E., and Talbourdet, G. J., "Recent Roll Test on Endurance Limits of Materials", Mechanical Wear (Symposium), ASM, 1950.
9. Faires, V. M., Design of Machine Elements, The Macmillan Company,New York, 1941.
10. Thomas, H. R., and Hoersch, V. A., "Stresses Due to the Pressure of One Elastic Solid Upon Another", University of Illinois Engineering Experiment Station Bulletin No. 212, 1930.

CHAPTER 6

SURFACE ENDURANCE OF MATERIALS

1. EXPERIMENTAL LOAD STRESS FACTORS

Values of strength for materials are determined experimentally, and surface endurance strengths are no exception. The successful design of a machine part will have its strength greater than its stress by a desired factor of safety. To accomplish this, the contact stress values may be computed from the Hertzian equations, and the strengths may be determined experimentally. Also, pitting is a function of stress and strength, and it is against this type of wear that parts in rolling contact are designed. It is logical that a basic relationship should exist between these parameters.

To find this relationship, tests were conducted to determine fatigue strength as a function of wear, and this relationship was then incorporated as a function of stress. The result is a factor which is constant for a given material at a given life.

Definition of Load-Stress Factor

The load-stress factor is defined as the limiting load for wear, and is equal to the load in pounds per inch of contact length multiplied by the sum of the reciprocals of the rolling parts' radii. It is given by the equation

$$K_1 = W \frac{1}{R_1} + \frac{1}{R_2}$$

From the Hertzian stress theory, the maximum normal compressive stress present in rolling contact is

$$\sigma_z^2 = 0.35 \frac{W(1/R_1 + 1/R_2)}{1/E_1 + 1/E_2}$$

Substituting the load stress factor K, the stress equation becomes

$$\tau_{z\,max}^{2} = \frac{0.35\ K}{1/E_1 + 1/E_2}$$

To convert τ_z to a maximum allowable stress value for no wear. K must be related to the surface endurance strength. This is accomplished experimentally in the following manner.

1. Two cylinders are rotated together at a constant speed and load until surface failure appears.
2. The load is changed and the test repeated with two similar cylinders. In each test the number of cycles to failure is recorded.
3. These repeated tests provide data for an endurance curve in terms of load for cylinders of given material and physical properties.
4. To convert the endurance curve values to the experimental load stress factor, the radii of the test rolls and the limiting load (W) from the endurance curve are substituted in the equation $K = W(1/R_1 + 1/R_2)$, and it is solved for K

From this brief discussion, the inherent relationship of K to strength and stress is evident

Load-Stress Factor Charts

Since the numerical values of K are directly proportional to the endurance strength of a material, they are plotted on charts with standard S-N coordinates. Reference is made to a paper by W. D. Cram[2] for a collection of 26 such charts, presenting values of K for many common metals used in engineering design.

Two of these charts are reproduced as Figures 1 and 2 of this chapter. From these figures there are three parameters that must

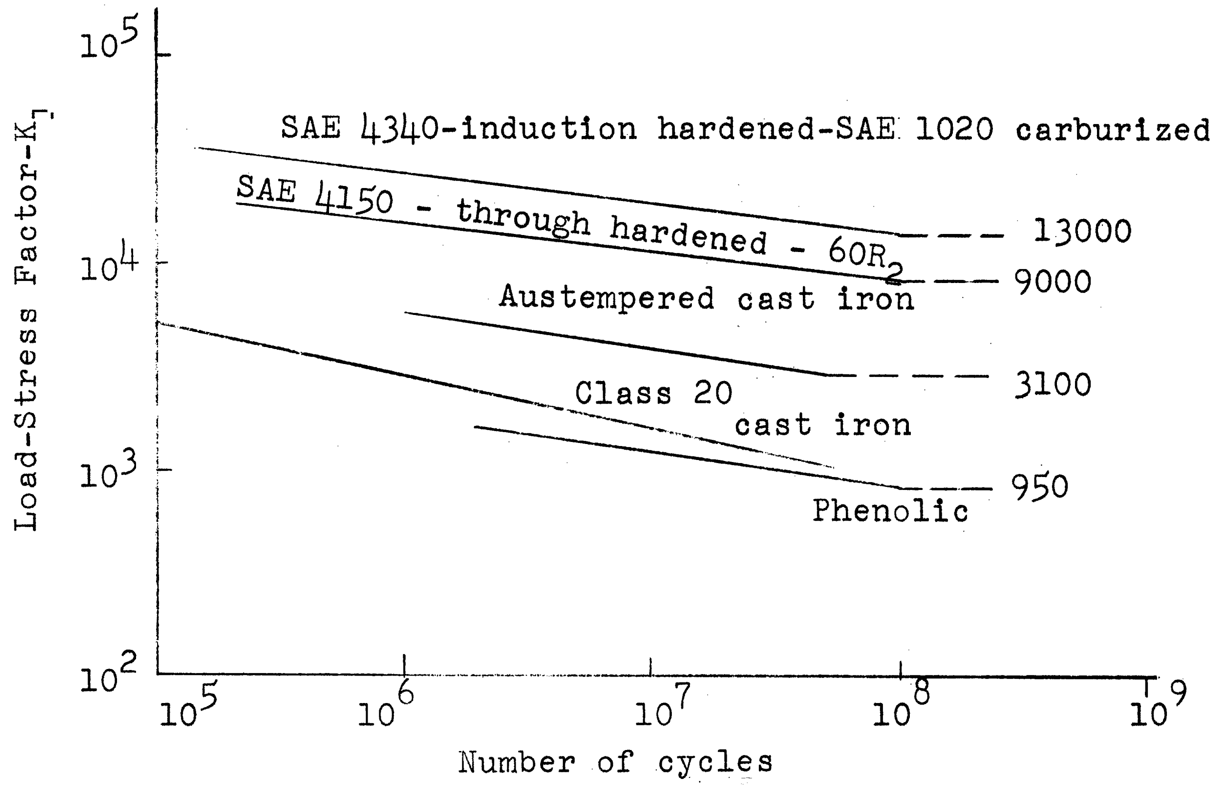


Figure 1

Relative Load Stress Factors

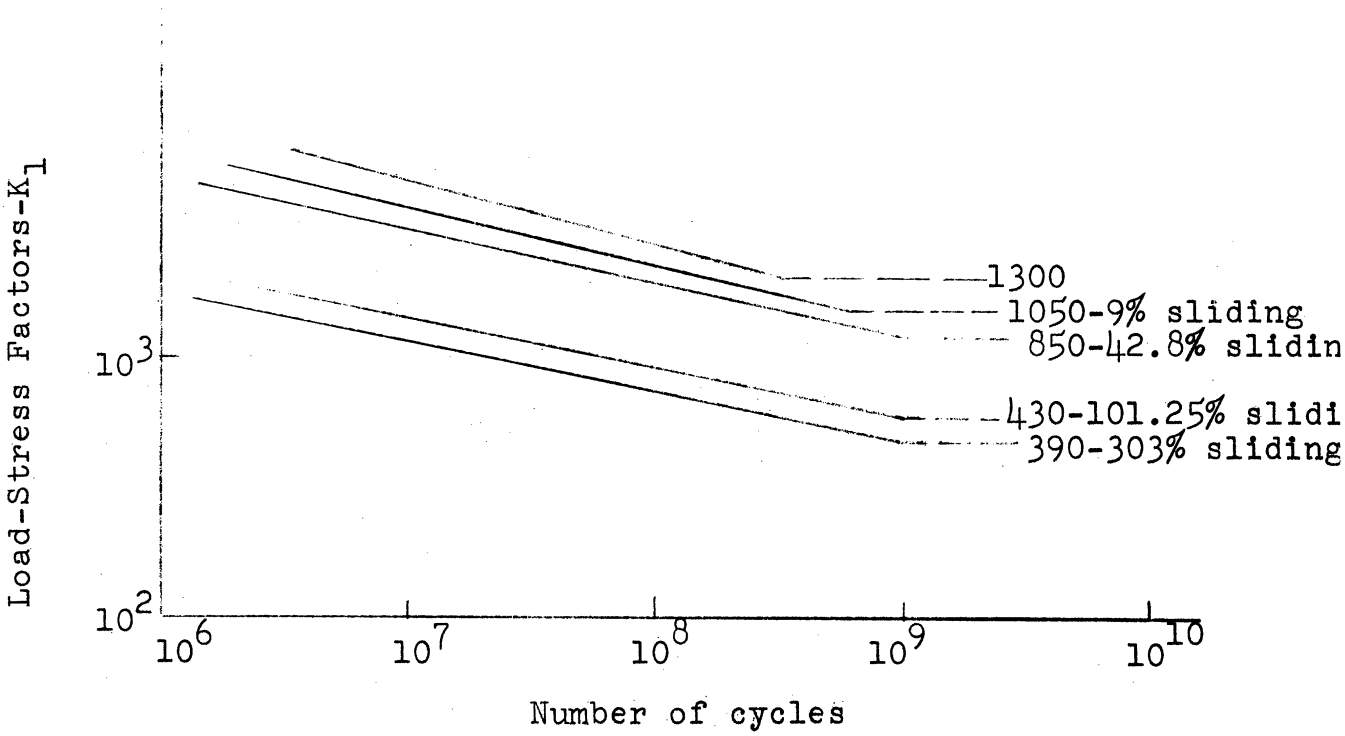


Figure 2

Gray Iron Casting, ASTM A48-48, Class 20, 130-180 Bhn and Same be determined to obtain a value of K for use in design. They are:

1. The material of the contacting surfaces.
2. Percent sliding of the surfaces at the point of contact if the application is not pure rolling.
3. The desired life of the parts in cycles of stress application. (Revolutions in the case of cylinders)

Conversely, the physical conditions of the application (load and dimension) may determine the value of K. In this case, the charts are used to select a suitable material, or life for the parts.

Figures 3, 4, and 5 also list the load stress factors for numerous other materials.

Application of Load-Stress Factors

The load stress factor is a tool to be used in design of parts subjected to rolling or rolling and sliding contact, such as gears, cams, ball and roller bearing, etc. It may be used to determine size, load, material or maximum allowable stress with a few simple calculations. To illustrate the application of the load stress factor, a basic problem is solved to determine the maximum allowable load on a cam and roll surface.

Sample problem

Assume: ASTM A48-48 Class 20, Gray cast iron
Radius of roll, $R_1 = 1$ inch
Radius of curvature, cam, $R_2 = 4$ inches

Service for infinite life
No sliding

Material	K_1 Cylinders	B (Gears) 14 1/2°	20
Cast iron with 30% steel scrap	2190*	137*	18
Nickel cast iron	1369	85	11
Nickel cast iron, heat-treated to 300 Br.	1643	102	14
Nickel cast iron, heat-treated to 350-400 Br.	1917	120	16
Nickel cast iron, hot-quench	3286	205	28
Nickel cast iron, hot-quench, 9% sliding	2465	154	21
Chrome-nickel cast iron	1506	94	12
Molybdenum cast iron	1917	120	16
Molybdenum cast iron, hot-quench	3697	231	31
Phosphor bronze, sand-cast	1177	73	10
Phosphor bronze, chilled	2054	128	17
Phosphor bronze, overheated when cast	205	12	1
Phosphor bronze, sand cast, with nickel cast iron, hot quench	2730	171	23
For 1,000,000 cycles			
SAE-2515 steel, .030-in. case	16,720	1045	143
SAE-2515 steel, .040-in. case	25,780	1610	220
SAE-2515 steel, .055-in. case	11,840	740	101
SAE-1040 steel, induction-hardened	17,900	1118	153
For 10,000,000 cycles			
SAE-2515 steel, .030-in. case	13,920	870	119
SAE-2515 steel, .040-in. case	19,680	1230	168
SAE-2515 steel, .055-in. case	9760	610	83
SAE-1040 steel, induction-hardened	13,900	868	118
For 100,000,000 cycles			
SAE-2515 steel, .030-in. case	11,660	728	99
SAE-2515 steel, .040-in. case	16,540	1033	141
SAE-2515 steel, .055-in. case	8180	511	69
SAE-1041 steel, .055-in. case	10,780	673	92

* Probably much higher than for average of this material

Figure 3

Experimental Load-Stress Factors[1]

(Mated with hardened steel unless otherwise noted)

Earle Buckingham Analytical Mechanics of Gears First Edition
McGraw - Hill Book Company, Inc. New York 1949.

Drive Roll	Driven Roll	Ultimate Strength	Elastic Limit- psi	Brinell Hardness	Endurance Limit-psi	Endurance Limit-psi	
Hardened and Ground Steel	C.I. with steel scrap as cast	35,200	15,000	223	21,000	87,000	
	C.I. with steel scrap heat treated	45,950	35,750	255	25,000	87,000	
	Nickel cast iron as cast	35,400	12,600	217	16,000	69,200	
	Nickel cast iron heat treated drawn	41,700 -	24,000 -	246 375	- -	75,800 81,900	
	Nickel cast iron heat treated	46,500	25,000	287	19,000	107,300 92,900	10% sliding
	Nickel chrome-cast iron as cast heat treated	39,000 44,700	15,100 31,600	234 243	- - -	72,400 72,400	
	Moly.cast iron as cast heat treated	- 49,200	- 22,400	- 290	- 22,000	82,000 113,700	
Phosphor Bronze	Sand Cast Bronze Chilled Bronze Cast Iron heat treated			80 80 -		67,600 82,800 83,300	
Nickel Bronze	Steel hardened and ground			80		75,700	
SAE 2515 Steel Carburized	SAE 2515 steel carburized Case Depth .030" .040" .060"	core: 180,000	core: 163,000	core: 38 R_c case 58 R_c		247,000 294,000 207,200	w/10% sliding at 10^8 cycles
SAE 1040 Induction Hardened	SAE 1040 Induction hardened					237,000	at 10^8 cycles

Figure 4

Surface Endurance Limits

From "Analytical Mechanics of Gears" by Buckingham

Table 2

	SLIDING	
	0%	9%
	FACTORS	
Gray iron casting, ASTM A48-48 Class 20, 130-180 Bhn and same.	1300	105
G. M. Meehanite, 190-240 Bhn and same.	1950	150
Nodular iron casting, ASTM-A339-51T, Grade 80-60-03, 207-241 Bhn and same.	3400	185
Gray iron casting, ASTM A48-48, Class 30, heat treated (austempered), 270-290 Bhn and same.	4200	340
1.05 Carbon Tool Steel, Hardened Through to 60-62 Rc and:		
Gray iron casting, ASTM A48-48, Class 20, 160-190 Bhn, phosphate coated.	1300	
Gray iron casting, ASTM A339-51T, Grade 20, 140-160 Bhn.	1000	90
Nodular iron casting, ASTM A339-51T, Grade 80-60-03, 207-241 Bhn.	2000	175
Gray iron casting, ASTM A48-48, Class 30, 200-220 Bhn.	1600	
Gray iron casting, ASTM A48-48, Class 35, 225-255 Bhn.	2300	210
Gray iron casting, ASTM A48-48, Class 30, heat treated (austempered) 255-300 Bhn.	3100	250
Gray iron casting, ASTM A48-48, Class 30, oil quenched, 270-415 Bhn.	2200	
S. A. E. 1020 Steel, 130-150 Bhn.	1700	135
S. A. E. 4150 Steel, heat treated to 270-300 Bhn, phosphate coated.	12000	790
S. A. E. 4150 Steel, heat treated to 270-300 Bhn.	9000	670
S. A. E. 4150 Steel, heat treated to 270-300 Bhn, flash chromium plated.	6300	
S. A. E. 6150 Steel, heat treated to 270-300 Bhn.	1850	
S. A. E. 1020 Steel, carburized to .045" depth of case, 50-58 Rc.	13000	850
S. A. E. 1340 Steel, induction hardened to 45-55 Rc.	10000	800
S. A. E. 4340 Steel, induction hardened to 50-55 Rc.	13000	900
S. A. E. 65 Phosphor Bronze, 67-77 Bhn.	1000	
High strength yellow brass, extruded.	2000	
S. A. E. 39 Cast Aluminum, 60-65 Bhn, and Gray iron casting, ASTM A48-48, Class 30, oil quenched, 340-360 Bhn.	300	
Zinc Die casting.	320	26
Random fibre cotton base phenolic.	1000	95
NEMA Grade L, laminated phenolic.	880	83
Linen base, laminated phenolic.	800	70
Graphitized laminated phenolic.	1000	

Figure 5

Experimental Load Stress Factors

(Infinite Life)

Find: Maximum allowable load on face of cam and roll surface for no significant wear.

Solution: From Figure 2, K = 1300

$$K = W\ (1/R_1 + 1/R_2)$$

$$W = \frac{1300}{(1/1 + 1/4)} = 1040 \text{ lb./in. of face width}$$

While this is a simplified example, the basic application of K is clearly shown. However, under actual operating conditions there are several modifications that must be made.

1. Lubrication - The experimental values of K were determined with adequate lubrication, completely free from abrasives. If this condition doesn't exist in actual operation, K must be lowered.
2. Deflection - If the cam deflects under loaded conditions, the effective face width of contact may be reduced. This would increase the actual load per inch of face width.
3. Sliding - The state of pure rolling seldom exists in machine parts. An actual or estimated value for percent sliding will reduce K as shown in Figure 2.
4. Loading - A cam often is subjected to impact loading, which will raise the maximum stress several times its normal value. To compensate for impact loads the value of K is reduced proportionately.

When the above factors are observed, the experimental load stress factor becomes a useful and accurate tool for design.

2. GEAR DESIGN

The most extensive use of the load-stres factor today is in the design of gear teeth. Buckingham's original and subsequent work to establish experimental load-stress factors was primarily for design of gears against wear.

The procedure outlined in the preceding section is applicable to gear design with two modifications:

1. From the Hertzian stress equations, Buckingham has derived a relationship that permits the substitution of the gear pitch diameters for the radii of the test rolls.[1]

   $$B = K_1 \frac{\sin \phi}{4}$$

   K_1 from Cram's charts[2]

   ϕ pressure angle of gear teeth

   In many design manuals tables of B are presented to solve for the limiting wear load. The values, B, are the factors in these tables.

2. Buckingham also derived an equation to solve for the total limiting load for wear.[1]

   $$F_W = B\,D_p\,b\,Q$$

   $B = \dfrac{K_1 \sin \phi}{4}$

   D_p = pitch diameter

   b = face width of the tooth

   $Q = \dfrac{2\,VR}{1+VR}$ VR = velocity ratio

F_W is the total permissible load for conditions of no significant wear. The value of b is the effective face width, or width actually in contact. Deflection of the gears, their shafts, and the flatness of the face must all be considered in determining this width. To be satisfactory, the maximum load F_W must be greater than the dynamic load on a gear tooth.

3. COMPARISON OF MATERIALS

Some materials show a much greater resistance to wear in rolling or combined rolling and sliding contact than do others.

Endurance strength, surface hardness, surface finish and grain structure determine the degree of resistance shown.

Materials such as medium carbon steels tend to cold work under repeated load application. This increases the endurance strength and improves the surface finish, thus increasing the wear resistance. On the other hand, cast iron does not cold work, has a lower endurance strength than steel, and shows relatively poor wear resistance.

The effect of surface finish is especially noted when sliding is present. A smooth surface materially reduces the effect of scoring.

A comparison of several materials is presented in Figure 1, with the load-stress factor representing a measure of wear resistance.

Steel

The best resistance to wear was shown by carburized 1020 and induction hardened 4340 steels. From Figure 1, values of K as high as 13,000 were displayed by these materials. Two influencing characteristics affect this high value.

1. Smooth surface - machining and cold working give them a smooth surface which reduces both the local stress concentrations at the surface and the effect of scoring. The stress concentrations are a source of fatigue cracks that can cause pitting.
2. Hardened case - carburizing and induction hardening increase the case (surface) hardness and endurance strength. With the case depth equal to twice the depth to the point of maximum shear, destructive (deep) pitting is prevented.

The ability of the steel to further cold work during operation also increases its wear resistance.

Cast Iron

In general, cast iron showed poorer wear resistance than did steel. Two forms of cast iron are indicated on Figure 1, and shall be briefly considered.

1. Austempered - through an interrupted quench, bainite is formed instead of ferrite and pearlite. This bainitic structure had the best wear resistance of the cast irons, primarily because of its strength and ductility.
2. Class 20 - Very poor wear resistance results from the brittleness of class 20 iron. No appreciable cold working takes place under repeated loads, so class 20 irons have a low endurance strength and high wear rate.

Non-Metallic Materials

Laminated phenolic materials display a moderate resistance to wear. However, in applying the load-stress factor, speed is a major factor. These materials deform a great deal under loading, and the energy of deformation is transformed to heat. Heat is not conducted away as fast as in metals, resulting in marked temperature rise as the speed increases. In phenolic materials, the wear rate increases with temperature.

Relative wear resistance of phenolic materials is shown in Figure 1.

BIBLIOGRAPHY

1. Buckingham, E., Analytical Mechanics of Gears, New York, McGraw-Hill Book Co., Inc., 1949.
2. Cram, W. D., Experimental Load Stress Factors, Page 57 of Reference 4.
3. Lipson, C., Pitting of Gear Teeth, Consulting report, 1950.
4. Lipson, C., and Colwell, L. V., Engineering Approach to Surface Damage, Ann Arbor, Michigan, University of Michigan Press, 1958.
5. Smith, J. O., and Chang Keng Lui, "Stresses Due to Tangential and Normal Loads on An Elastic Solid with Application to Some Contact Stress Problems," ASME Paper No. 54-Lub-14.
6. Timoshenko, S., Theory of Elasticity, McGraw-Hill Book Co., Inc., 1934.

CHAPTER 7

EFFECT OF INCLUSIONS

1. GENERAL DISCUSSION OF FATIGUE

The topic of this section is the concept of fatigue, what is known about it, and why metals are weak in fatigue. Following this is a discussion of the actual method of fatigue failure due to crack propagation.

Facts Known About Fatigue

The concept of fatigue is well known to engineers, and many of the variables which effect fatigue strength are known. However, the explanations of the effect of these variables are not all completely known. Those which are known are the stress-fatigue life diagram or S-N diagram, the effect of the mean stress upon the allowable or safe range of repeated stress, the effect of corrosion upon fatigue life, the notch effect, and the effect of combined stresses. The explanations of these variables are as follows:

1. S-N Diagram - The allowable or safe range of repeated stress varies lineraly(or nearly lineraly) with the logarithm of the number of cycles of stress repetitions till failure occurs. Also, the most commonly used metal for machine design, steel, exhibits the property known as an endurance limit. That is, below a certain range of stress, failure will not occur.
2. The effect of the mean stress upon the allowable or safe range of repeated stress: as the mean stress in increased, the safe range of cyclic stress diminishes. This is shown by the Modified Goodman Diagram.
3. The Notch Effect - The effect of a notch is known quantatively via the mechanism of concentration of stress and strain.
4. The effect of corrosion - Corrosion action tends to sharpen notches, or create them if none exist. In this way, it localizes the strain or stress.

5. The effect of combined stresses - The shear strain energy theory, for ductile metals, describes the effect of combined stresses upon the life of a part.[1]

Why Metals Are Weak in Fatigue

Fatigue failure is known to be failure which originates at a localized point, and which progresses gradually through the rest of the part until complete failure occurs. Yielding and necking, found in static failures, is not found in fatigue failure. Also, the repeated stresses which cause fatigue failure are or can be below the yield point of the materials. With this in mind; an attempt to explain why metals are weak in fatigue will be made.

In general, commerical metals are made up of a general grouping of dissimilarly shaped and randomly oriented crystalline grains. Thus the metals cannot be considered to be homogeneous. The elastic strength of any given crystal could be exceeded due to loads applied to a part, and it is possible that the stress on the total section (referred to as the nominal stress) due to this load is well below the yield point of the material in the aggregate condition. This process causes localized deformation, and the deformations of these crystals are called slip bands by the literature. In a static test, at first only a few randomly located grains form slip bands, but as the stress increases (gradually) many slip bands are formed and most the grains are damaged.

Under repeated or fatigue loading, at stresses below the yield point, the effects of these slip bands are localized to a few crystal zones of great stress concentration. Therefore the most important factors which lead to fatigue failure originate in localized or concentrated regions, as compared to static failure of many areas. When the stressing is repeated, the deformation and crystal breakup continues and eventually very small (submicrocopic) cracks appear. These cracks are usually connected to or associated with the slip bands in the early stages of failure. As the failure continues, the small cracks may lengthen by intersecting others and a small or microscopic crack develops. From the discussion above, one can conclude that the failure mechanism is formed by an accumulation of effects which may be considered as chance effects. The nucleation and growth of cracks is not similar to static failure. The weaker grains or crystals are microscopically readjusted, and these readjustments are multiplied in their severity by the repeating load. Fractures develop at a relatively low nominal stress (in comparison with the strength of the material in static tension). The nature of the heterogenity of the polycrystalline metal lets the response of a member to an external load be regarded as similar to a physical model built of small elements (crystals) that have different sizes and shapes. The individual elements strength is dependent upon its size and orientation. Under loading of the member, the damage development is a statistical process largely dependent upon chance and the location of the weakest crystals in the high stress zones. In other words, the damage may be visualized as developing at varying rates in the localized crystals with each repeated load application. Thus to the engineer, the terms fatigue life and limit have meaning in terms of the probability of failure at a predetermined number of cycles.

Metals contain many heterogeneous constituents and it is not surprising that they exhibit a weakness when fatigue loaded because these elements lower the resistance to mechanical stressing. These heterogeneous constituents are:

1. Random orientation of crystals of varied sizes;
2. Various phases of different elastic characteristics mixed together.

The mixed phases can consist of inclusions which have different physical properties than the matrix of the major phases. Discontinuities and segregation streaks are also evidence of mixed phases. Stresses are probably formed at the boundaries or interfaces of the various constituents since they do have different mechanical and physical properties.

Almost always, fatigue cracks begin at the outside or exposed surface of the metal. The only exceptions to this are metals with special treatments such as nitriding where the surface metal is much stronger than the core material. It appears that the exposed surface is weaker than the core, since the interior or core grains have bonds with adjoining grains and those at the surface do not. However, it must be noted that the stress at the surface of bars is maximum for bending and torsion. The ideal crystalline structure has the lattice extending indefinitely in all directions, whereas crystals may have a free surface (those at the outside of a bar) at which the positions of atoms for equilibrium must be changed. Therefore imperfection of the atom arrangement is found at the surface layer. Indirect proof of this surface weakness due to atom mis-arrangement can be found in the process of surface work, such as shot peening. After shot peening of a surface, the fatigue strength may go up. The shot peening increases only the strength of the outer layers of metal, therefore one can conclude that the inner metal strength was higher than the surface metal strength.[2]

Statistical Nature of Fatigue Data[2]

Metals undergo readjustments on a microscopic scale throughout the cycles of repeated stressing. The instability and chance changes in structure result in diverging data in phenomenological endurance limit tests, (i.e) the scatter of data during the experimental determination of endurance limit of metals.

A metal surface, when viewed through a microscope, reveals a distinctly heterogeneous appearance. Irregularly shaped grains, complex alloying constituents, oxides, nitrides, and other contaminants are literally all over the surface. Now we recognize the improbability of fixing a definite fatigue life to a particular part chosen as a random sample even when detailed test data are available for the sample material. Each elementary crystal and also its neighboring cross links have a characteristic strength behavior. When a load is applied to the member, the part may respond in a variety of ways depending upon the surface condition at each point. When the cumulative damage reaches a certain stage, a small microscopic crack appears on the surface and the crack may propagate rapidly to neighboring crystals. Thus, it is apparent why all metals are weak in fatigue. The real problem is how can we predict the most probably strength of a part.

Many factors affecting strength such as the imposed service conditions and the properties of the metal must be carefully considered.

The scatter of data obtained in a fatigue test program results from the inherent characteristics of the material and not from inaccuracies in experimental determinations. One can not speak of the fatigue limit in terms of a single value. The S-N curve must be considered a family of curves with some functional relationship. The series of curves constructed from the scatter of data are shown in Figure 1. Each curve represents a definite probability of failure. The mathematical probability of failure should not be surprising because probability looms large in the study of many physical laws. The kinetic theory of gases and the average velocity of a neutron represent two cases where probability distributions are employed.

The conventional S-N curve as normally obtained with 10-15 specimens should be regarded as a crude approximation. Such a small sampling can not possibly establish a normal distribution; nor can damage line data be obtained in this manner. Here as in many applications of science to industry, economic considerations preclude accumulation of sufficient statistical information to design a single part for limited life with a predictable degree of certainty. In addition, many variables and uncertainties appear in practical service conditions precluding a precise determination of service history. The exact life of an individual member such as an airplane wing, a turbine blade or a gun tube

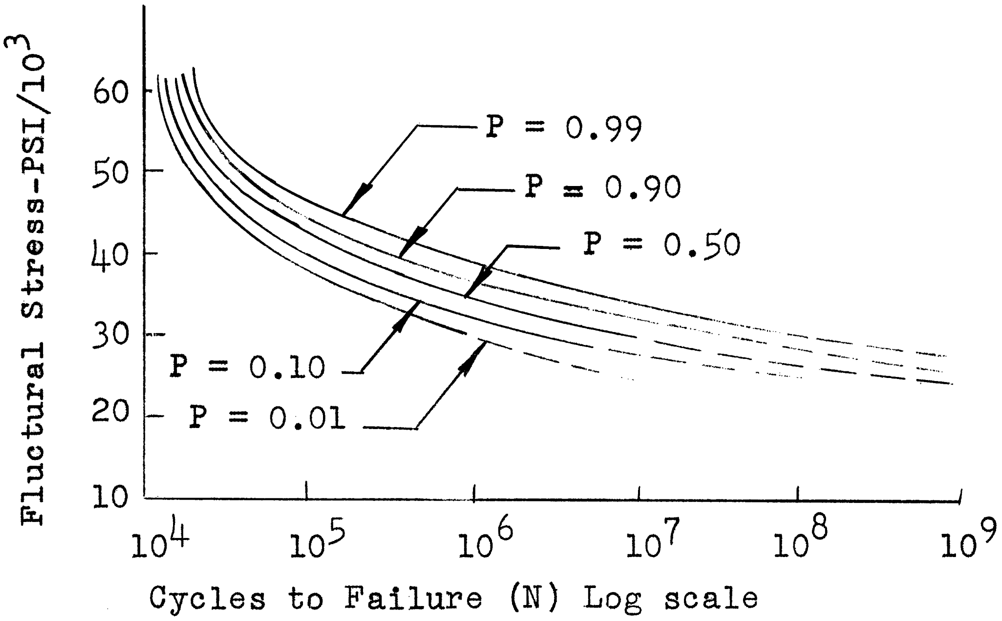


Figure 1

Composite SN curve for various probabilities of failure P derived from data for small unnotched specimens of 75-S-T Al.

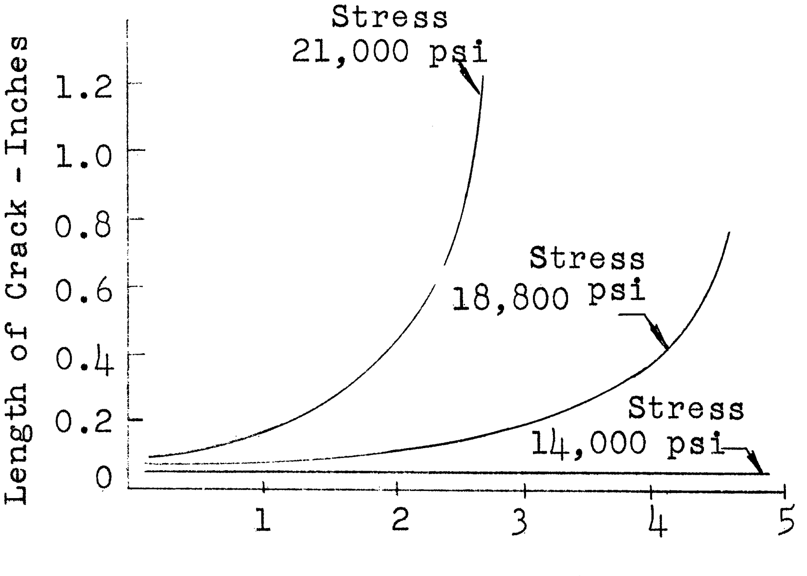


Figure 2

Condition of Crack Propagation in 0.45% C Steel cannot be precisely determined; however, enough data can be obtained to base design upon infinite life data provide an endurance limit does indeed exist for the material under consideration. Other uncertainties are accounted for by the use of suitable safety factors.

Phenomenological Evidence of Fatigue Damage-Crack Propagation

The fatigue process may be characterized by the following three stages:[2]

> Stage 1-during which crystals slip and fragmentation occurs accompanied by work hardening until the limit of cold work is reached.
>
> Stage 2-during which the statistical distribution of the crystalline lattice takes place (ie. breaking of the high energy atomic bonds in the severely cold worked striation of slip within the crystals resulting in the formation of sub-microscopic cracks.
>
> Stage 3-during which the sub-microscopic cracks coalesce to join visible spreading cracks which ultimately result in fracture.

True fatigue damage appears during stage 2 and 3, since work hardening does not necessarily damage the part.

Once visible cracks appear, they spread very rapidly. Figure 2 represents the rate of crack propagation in a 0.45% carbon steel. The propagation of the crack depends upon the magnitude of the imposed stresses. At 21,000 psi the crack propagates very rapidly while at 14,000 psi the crack does not propagate at all. Hence, the stress level is an important consideration in prevention of failure by propagation of a crack.[3]

Sub-Surface Fatigue

Most commonly reported fatigue failures start at the surface of a machined part. Sub-surface failure in fatigue occurs in several important applications. Sub-surface fatigue in this paper will imply damage originating at least several thousandths of an inch below a nominal surface. An attempt will be made to extrapolate surface conditions which are particularly applicable.

Many parts of failed machinery indicate the following sub-surface fatigue failures;

1. Excessive number of inclusions.
2. Internal fissures incurred by heat treatment or cold workings, some of microscopic size.
3. Overstressing in service by high vibrational stresses at points of weakness such as the case core junction.

2. INCLUSIONS

Types of Inclusions[4]

Three kinds of clear cut inclusions exist, but many sub-variants are also known. The three main divisions are:

1. Non-metallic
2. Metallic
3. Inter-metallic

The first type consists of foreign matter in the wrong place such as dirt. Sufficient care in selecting raw materials and in processing metals or alloys should reduce the number of such inclusions. Examples of non-metallic inclusions are: entrained particles of slag, flux, oxides, phosphides, blowholes, etc. Such bodies are usually introduced inadvertently while the metal is in the molten stage.

Metallic inclusions are introduced intentionally to facilitate machining. Inclusions which are intermetallic compounds arise neither from carelessness nor by the intentional acts of man.

Blow Holes

Blow holes are caused by the following:

1. Decrease in solubility during solidification
2. Reaction between oxides and carbon
3. Reaction between liquid metal and moisture in sand
4. Entrapment of air in molds.[5]

Impurities

Impurities may be present in solution with metals or as separate mechanical particles. Inclusions of ferrous alloys are divided into the following groups:

1. Oxides
2. Sulfides
3. Silicates

Many inclusions of a more complex intermediate composition exist and every commercial metal contains some characteristic inclusions. Impurities that occur in continuous films or long rows of inclusions are most undesireable, since discontinuities of this kind are of a weak and brittle nature.[5]

Origin of Flakes

Wide spread attention to this defect in steel was attracted by two railroad accidents in 1911 and 1916 respectively. The failures were caused by sub-surface fatigue of the rails. In the cross-sectional area underlying the failure, small microscopic cracks were discovered. These flakes (or transverse fissures) were attributed to hydrogen gas which was disolved in the molten metal. Since the solubibility of hydrogen decreases with temperature, free gas evolved at high pressure causing minute cracks in the metal during cooling.[5]

Shape of Inclusions

In cast metals the shape of all inclusions are about the same in all polished cross sections. In wrought metals, as the orientation of the polished surface is changed, the shape of inclusions vary. Mechanical working at high temperatures deform the plastic inclusions. The inclusions assume elongated shapes in the longitudinal direction of mechanical working and flat globular shapes in the transverse sections.[5]

Orientation of Inclusions

Inclusions oriented normal to the direction of loading are the most harmful. Inclusions located in fillets or other geometric discontinuties may cause premature failure because inclusions are local stress raisers.[2]

If a fatigue crack appears at an inclusion, it normally progresses from inclusion to inclusion. Highly heat treated steels are particularly sensitive to inclusions, but soft steels are relatively insensitive.[2]

Modern methods of non-destructive testing, if properly used, make it relatively simple to detect the more damaging types of inclusions. The fact that a material contains inclusions should not form the only basis for rejection. The type, metallic or non-metallic, location, size, shape, location and type of loading must all be considered. It is wise, however, to avoid the use of "dirty" steels in such designs as crankshafts, piston pins and landing gears.[2]

The results of repeated fatigue tests longitudinally and transversely are tabulated in Table 1 below. Transverse and longitudinal designate a specimen taken from rolled steel transverse to the direction of rolling and another specimen in the direction of rolling. The transverse fatigue limit is only 52% as great as the fatigue of steel A in the longitudinal direction. This property of steel is called anisotropy. Steel B exhibits similar characteristics.[6]

The transverse fatigue life of forgings made by conventional steelmaking practice is less than the longitudinal fatigue limit largely because elongated inclusions are present. Elimination of large inclusions of measureable size (vacuum steel, Table 1) improves the strength in the transverse direction considerably.[6]

For the transverse specimen fatigue failures were initiated by inclusions becuase the long axis of the inclusion was parallel to the direction of maximum shear stress and perpendicular to the direction of the principal stress.

Table 1

Fatigue Limits in Repeated Bending[6]

	Steel (A)	Steel (B)	Vacuum Steel
Longitudianl Fatigue Limit psi	128,000	116,000	139,000
Transverse Fatigue Limit psi	67,000	79,000	120,000
Ratio $\frac{\text{Trans.}}{\text{Long.}}$	0.52	0.68	0.86
Hardness R_c	28	27	29

Distribution, Size and Number of Inclusions

No way has been found to put the varied types of fracture, single and multiple nuclii, large and small inclusions, depths below the surface, etc. all on a common basis so that a single general quantitiative correction between inclusion size and its effects onfatigue life can be determined. Microscopic examination of aircraft quality 190 ksi steel reveals a conglomeration of sizes, numbers and depth below the surface.[7]

Beneficial Effects of Inclusions

In some commercial metals non-metallic inclusions are beneficial. When a greater speed of machining is desired, inclusion forming elements are purposely introduced into the alloy, as in the case of sulphur added to low carbon steel or tellurium to copper. Long records of wrought iron's resistance to corrosion in air and in soil are explained by the beneficial effect of slag inclusions and the absence of manganese.[5]

The Relation of Inclusions to Notches

Surface inclusions (or inclusions which are open to the surface) are in effect, notches. Their stress raising ability can be considered from the theoretical view point. The Battelle Staff states that experience shows that the effect of an inclusion is similar to that of a hole of like dimensions. This statement considers the fact that the inclusion material is weak as compared to the host metal, and does not sustain any of the shearing load carried by the rest of the surface metal (when in torsion). With the above statements in evidence, the study of inclusions through published literature dealing with stress concentration factors is possible.

In both torsion and bending, the surface (or shallow subsurface) is the most highly stressed local area and any surface roughness or discontinuities have the effect of stress raisers. An inclusion at the surface can be considered to be a discontinuity. In general, the discontinuities or inclusions in the surface are more serious in harder steels since most hard materials are very notch sensitive. The stress concentration factor however, is not a function of the metal or its hardness, theoretically, but is a function of the size, shape, and location of the notch (or inclusion).

Much published material for calculation of stress concentration factors can be found for many types of notches. However, most of these formulas are theoretically derived and cannot be used in their present form for stress concentration factors for inclusions. To show this, one might consider the following formular for the stress concentration factor (s.c.f.) produced by a keyway in a circular shaft under torsion loading.

$$K' = 1 + \sqrt{\frac{t}{r}}$$

t = depth of keyway

r = radius in the corner at the bottom of the keyway

A longitudianl inclusion exposed at the surface of a circular shaft could be considered as a keyway, and an s.c.f. could be calculated for it. (This formula holds for a shaft which is large in respect to the size of the depth and length of the keyway). In forgings, the inclusions are of the elongated type as discussed earlier, and therefore from the formula one can see that the factor K' increases as the radius decreases In fact, it will become infinite when the radius goes to zero and the notch becomes sharp. This is obviously not possible, since it would mean failure under any load at all.

Another interesting fact about the formula is that an inclusion of equal depth and radius (t = r) can have only one value of K', that is 2.00. Thus one would conclude that the size of the inclusion has no effect upon the s.c.f. and this is not so. As proof, Figure 3, shows this conclusively. Cummings, Stulen, and Schulte[8] made an investigation of 4340 steel as to inclusion content and effect upon fatigue limit. They found for 4340 steel of 140,000 psi tensil strength, that a repeated bending stress of 86,000 psi, the fatigue life increased with steels with decreasing size of spherical inclusions. They also noted

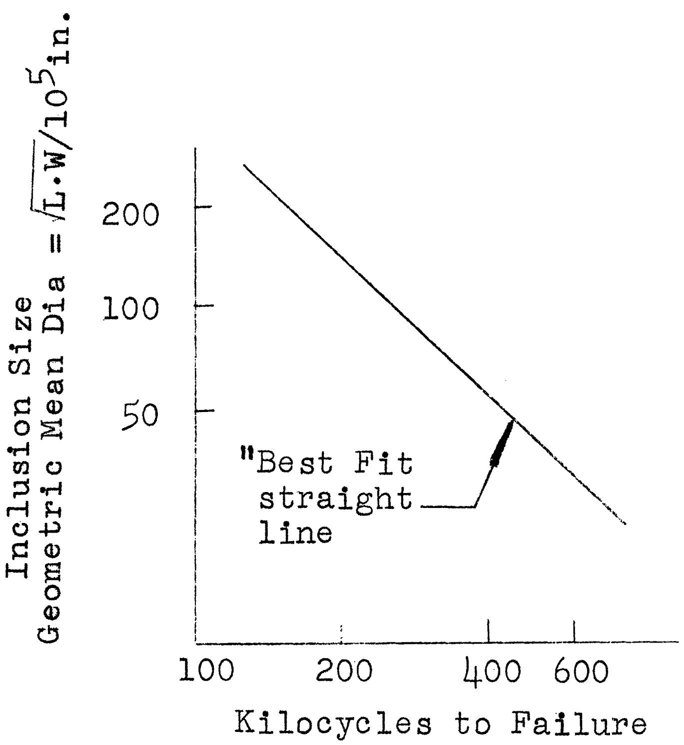
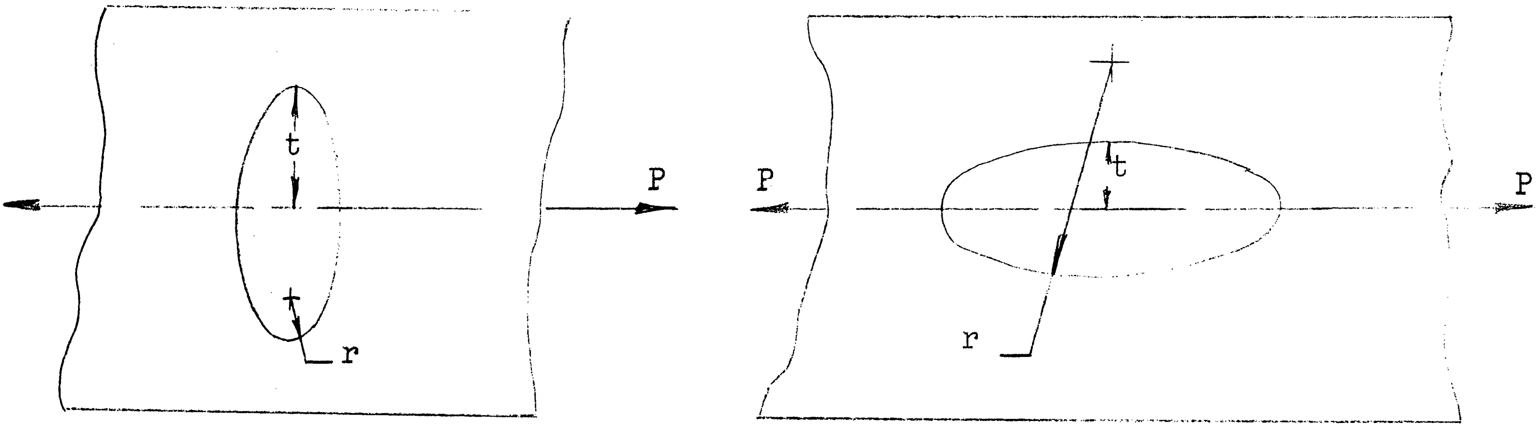


Figure 3
Inclusion Size vs. Fatigue Life

SAE 4340 Steel
140 KSI T_oS_o
Stress Level= 86 KSI
L = Length
W = Width



Perpendicular (transverse)

Longitudinal

Figure 4
Orientation of Elliptical Inclusions on Bars

that this behavior was not true of all types of 4340 steels or at all stress levels, but the fact that the formula does not apply in one case gives reason to believe that it will not work in other cases. Stewart and Williams[9] found that the fatigue strength of soft steel (Mo-Vanadium) was dependent upon the inclusion size. In this case the inclusions were longitudinal. They took specimens from three areas of a large forged shaft, and these three areas had different sized inclusions. They found the worst case, i.e. that specimen with the heaviest inclusions, reduced the fatigue strength by 20%. This test was run with repeated torsional loadings.

With the above discrepancies in mind, a new correction for the $K' = 1 + \sqrt{\frac{t}{r}}$ formula was developed. The formula is now:

$$K = 1 + \frac{K' - 1}{1 + \frac{\pi}{\pi - W}\sqrt{\frac{r'}{r}}}$$

K = Technical SCF
K'= SCF for ideal material $= 1 + \sqrt{\frac{t}{r}}$
W = flank angle of notch
r = radius of curvature of notch
r'= elastic notch sensitivity = .0189 inch for steel

This formula seems to work well for the transition range from weakly curved notches to pointed ones as it coincided with experimental evidence.

The flank angle can be considered to be constant for inclusions at a value of zero since this would give a maximum value of K therefore being conservative. With this fact accounted for, the formula is:

$$K = 1 + \frac{K' - 1}{1 + \sqrt{\frac{.0189}{r}}}$$

This formula now will explain the fact that small inclusions are less damaging than large inclusions. Assuming that the inclusion is circular in the transverse section, the following are values calculated by the uncorrected and corrected formulas:

Radius of Cross Section (r)	Depth (t)	K'	K (corrected)
.25	.25	2.00	1.785
.10	.10	2.00	1.696
.03	.03	2.00	1.557
.005	.005	2.00	1.340
.001	.001	2.00	1.187

The above values are calculated for constant (r/t) and assume that half of the inclusion is machined away when finishing the shaft.

The next topic of interest is that of the orientation of inclusions. The type inclusion to be discussed is the longitudinal type and it is assumed that its size is small in comparison to the size of the shaft in which it is found. Again, if the inclusions are located on the outside surface of the shaft, they may be compared to a formula for a theoretical s.c.f. In this case, the inclusion may be compared to an elliptical hole with its major axis perpendicular to the stress application. Here, the dimension t is one-half of the major axis length and r is the radius at the ends of the major axis. The formula for this case is then:

$$K' = 1 + 2\sqrt{\frac{t}{r}}$$

The equation also works for an elliptical hole with its major axis along the longitudinal axis of the bar. When the elliptical hole is in the longitudinal direction, the radius r is very large in comparison to t (see Figure 4). The result is that the K' is small as compared to the perpendicular orientation. For the majority of longitudinal inclusions, the ratio of t/r is very small, with the result being that K' is 1.00 for general purposes. Thus one concludes that longitudinal inclusions would not lower the fatigue limit in bending or axial loading to any appreciable extent. This fact has already been mentioned.

Conclusions

In general, the conclusions that can be drawn are as follows:

1. In soft steels, inclusions of a longitudinal nature (such as in forgings) do not affect to any extent the fatigue life in bending or axial loading. However, in torsion they do reduce the fatigue strength by 20-25%, in high strength steels, longitudinal inclusions are deleterious for all modes of loading.

   Transverse inclusions reduce the fatigue strength by as much as 50% for both soft and hard steels.

   A number of inclusions arranged with respect to one another can be much more damaging that single inclusions or several inclusions with a random arrangement.

Small globular inclusions apparently do not reduce the fatigue properties significantly.

2. Theoretical considerations can be used for calculating quanititatively the effect of some inclusions upon the fatigue strength of metals. However, corrections must be made to adjust the theoretical case to the practical case.

3. There is, in some cases, correlation between the size of spherical inclusions and the fatigue life of metals. For a particular hard steel, SAE 4340, the logarithm of the life in cycles decreased linearly with the increase of the logarithm of the size of the inclusions. For softer steels, there apparently is a relation between the fatigue strength and the size of inclusions. Inclusions of large size can reduce the fatigue strength by as much as 20%.

4. Inclusions are very harmful in regions of high stress, such as on the surface of a shaft in torsion or bending. Reductions of 20% of fatigue strength can occur.

5. Apparently, when inclusions are located in planes or patterns of principle stress, the fatigue strength is drastically reduced, as much as 60% in some cases.

6. Inclusions are particularly undesirable in regions of high stress.

BIBLIOGRAPHY

1. Murray, W. M., Fatigue and Fracture of Metals, Technology Press of M.I.T. and John Wiley and Sons, 1952, Paper 1.
2. Sines and Waisman, Metal Fatigue, McGraw-Hill, Chapter 3, pp. 379-381, 1959.
3. Lessells, J. M., Strength and Resistance of Metals, John Wiley and Sons, Inc., Chapter 8, 1954.
4. Murray, W. M., Fatigue and Fracture of Metals, Technology Press M.I.T. and John Wiley and Sons Inc., 1952, Chapter 5, 12.
5. Polushkin, E. P., Defects and Failures of Metals, Elsevier Publishing Co., Chapters 3,5,10, 1956.
6. Ranson, J. T., "The Effect of Inclusions of the Fatigue Strength of 4340 Steel," American Society for Metals,Vol. 46, 1954.
7. Cummings, H. N., Stulen, F. B., Schute, W. C., "Relations of Inclusions to the Fatigue Properties of SAE 4340 Steel," Trans. American Society for Metals, Vol. 49, 1957.
8. Stewart, W. C., Williams, W. L., "Effects of Inclusions on the Endurance Properties of Steels", Journal, American Society of Naval Engineers, Volume 60, 1948, pp. 475-504.

CHAPTER 8

EFFECT OF SURFACE TEMPERATURE

1. PRINCIPLES

In order to keep the range of the subject within the scope of this chapter, it is necessary to lay a restricting foundation in the form of three basic assumptions.

The first assumption is that there is perfect thermal contact between the two sliding surfaces. Being primarily interested in cases of boundary lubrication it can be seen that this assumption lies within reason since with boundary (thin film) lubrication there is a high degree of metal to metal contact with only a molecular layer of lubricant so that essentially the lubricant has very little or no effect as a coolant.

In the event of contact between two rubbing surfaces the source of heat naturally lies in the contact area.[8] This heat source extends in one direction only. That is, the heat is generated only at the point of actual contact. Since this is a case of boundary lubrication, it can be seen that this is a case of solid friction. Therefore the heat generated is entirely disipated to either side of the contact surface by conduction into the two rubbing parts.

The second assumption made is that, within the contact area, the bulk temperature between the two contacting parts remains constant. This can be seen by taking a microscopic view of the actual contact area. Here the actual contact is between the peaks of the asperities. Since the asperities are considered to bear the entire load the real contact pressures are very high, and accordingly the friction heat intensity becomes great even at appreciably low sliding velocities. Thus the heat generated must enter into the machine part through narrow conducting areas, i.e. the asperities in contact. As a result of this thermal resistance there is a considerable rise in the local temperature and this temperature is limited to the contact surface and the points immediately surrounding it. Beyond this small contact area the bulk material provides a low resistant passage as compared to that of the primary resistance.

The third assumption is based on the fact that the heat is generated by the action of the two surfaces rubbing together and is a function of the relative sliding velocity. Therefore it must be assumed that the sliding velocities are sufficiently high for the temperature rise to take place within the contact area.

2. SCORING

Scoring of surfaces is gradually becoming more recognized as a direct function of surface temperature.

In the case of unlubricated surfaces this temperature is generally considered to be the softening or melting point of the materials. In the case of solid lubricant films, it is directly related to the softening point of these films.[1] More recently it has been found that surface temperatures far lower than melting points of the materials in the surfaces bring about the onset of scoring of liquid lubricated surfaces.[2] It is, as a result, becoming more recognized that non-reactive (no extreme pressure additives) mineral oils and very likely most other commercial oils and synthetics[3,4] have a critical temperature beyond which they are no longer capable of satisfactory lubrication. Many speculations on the exact mechanism over which the temperature has control have been made. The most promising of these hypotheses perhaps is one concerning the failure of the adsorption characteristics of the oil called desorption. Roughly this may be looked upon as the failure to "wet" surface.

Scoring takes on different appearances, which are principally functions of the characteristics of the material, its surface characteristics and the lubricant. The carburized and hardened surface scored in a non-surface reactive mineral oil is badly torn and appears "roped". The same material scored in an external pressure (EP) lubricant appears softer with more material flow. Very likely, the lack of tearing which occurs with EP oil is associated with the protective contamination of the surface by the additives and the higher surface temperature required to produce the failure. This same appearance, although to a lesser extent, will frequently be seen to occur on a surface which has been treated, such as one which has been Parko-Lubrited.

With regards to the maximum mean temperature due to frictional heating, T_f, Figure 1, that is produced, a qualitative examination of the formula derived by Blok will be revealing. The formula is shown below.

$$T_f = K \frac{f\ W_n\ (V_1 - V_2)}{(C_1\ \ V_1 + C_2\ \ V_2)\ \frac{b}{2}}$$

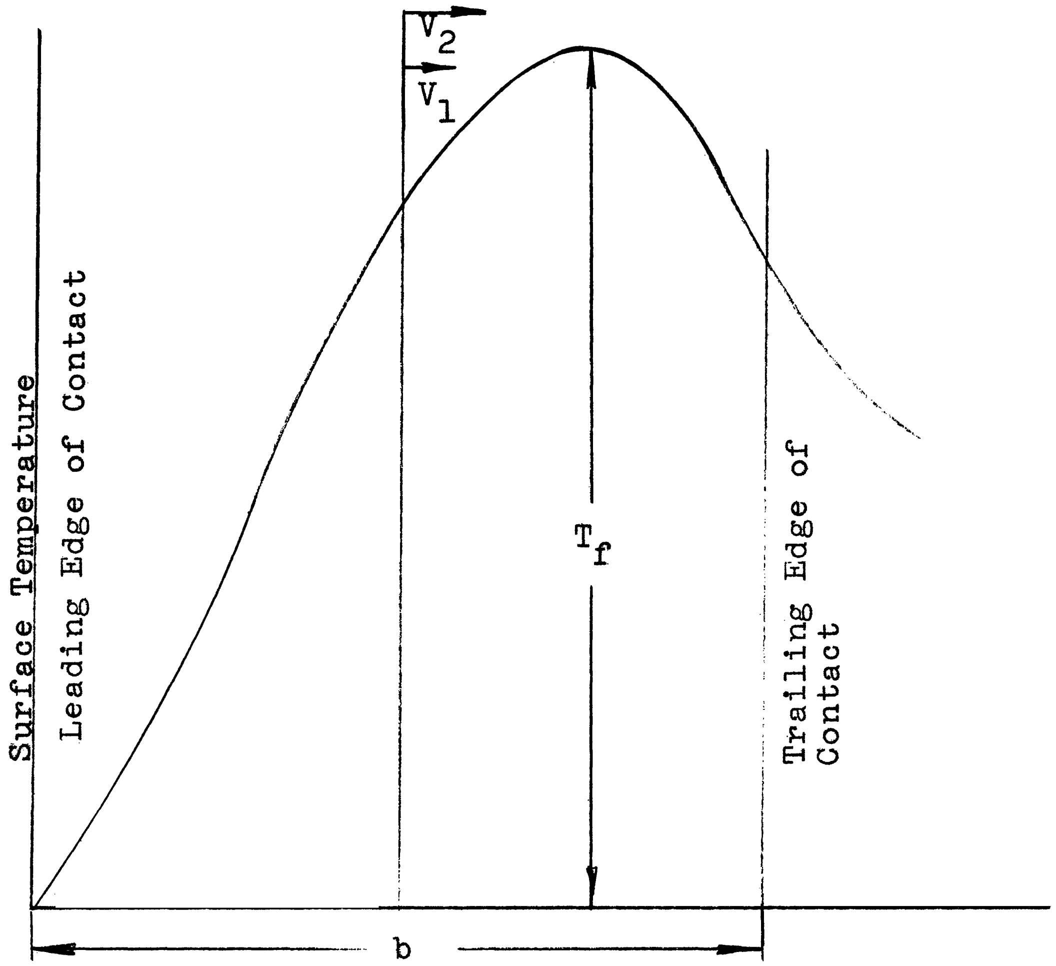


Figure 1

Shape of Surface Temperature Rise in Band of Contact

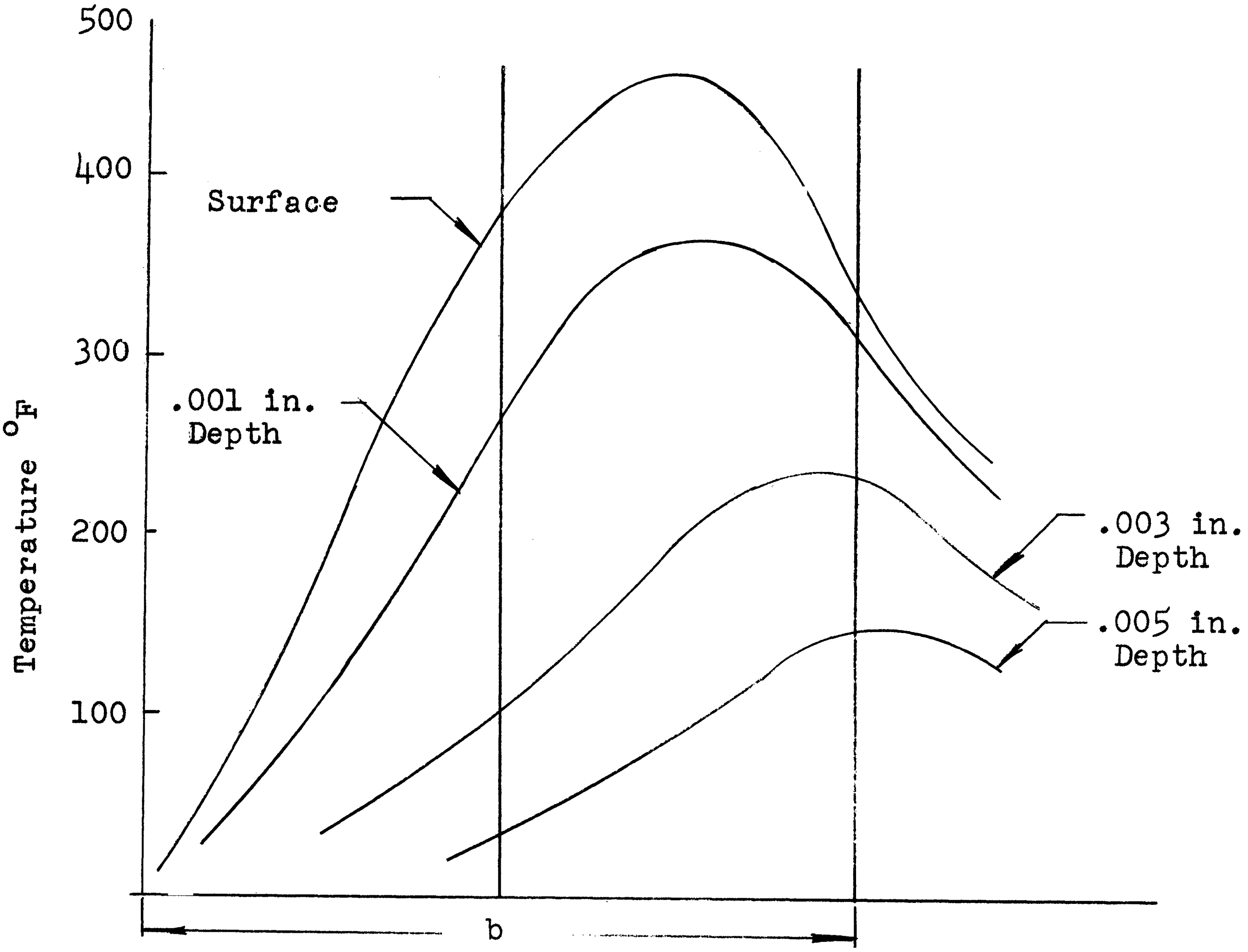


Figure 2

Temperature Flash on and Below Surface

K = Constant (1)

f = Coefficient of friction

W_n = Normal load per unit length

V_1 and V_2 = Surface velocities of cylinders in contact

C_1 and C_2 = Constants of material which include thermal conductivities, specific heats and densities

b = Width of band of contact

T_f = Maximum mean surface temperature (sometimes called "flash temperature".)

It is important to notice the velocity relationship that is obvious in the formula. From any given condition, if V_1 and V_2 are increased but the difference remains the same, then the resulting temperature, which is frequently called temperature flash, is reduced. This fact has been often overlooked in previous scoring criteria, such as PV and PVT, which are proportional to the heat intensity but not the surface temperature.

It will be found that if V represents the sliding velocity or the difference between V_1 and V_2, then the socring resistance will vary as $\sqrt{V}$. This velocity relation has been used, for instance, by T. B. Lane[5] to further verify the temperature criteria for scoring on gears.

Another interesting relation is found between maximum Hertz pressure, P_o, and temperature. It can be shown for instance that the maximum surface temperature varies as P_o 3/2 instead of directly proportional of P_o for a fixed values of V and geometry such as the roller radii. Increases in radii of curvature will, however, affect the surface temperature much less than might be expected and as in the case of gears where the use of larger radii of curvature such as in the case of gears where the use of larger radii of curvature such as in the adoption of higher pressure angles, will frequently be accompanied by higher values of V which can easily create no improvement in surface temperature and scoring resistance at all.

In a previous publication on gear tooth scoring,[6] the author showed the distribution of the load and speeds on gear teeth, and a formula was presented based on the temperature flash formula but which included an empirical factor of surface roughness. An assumption was made that the coefficient of friction was a constant for all of the tests. It was recognized at the time that it introduced elements which were foreign to the fundamental nature of the formula.

Since this work, investigations into the coefficient of friction as influenced by velocity have allowed us to use the formula below with improved success. Allowing for the similarity of material and simplifying, the formula can be written as:

$$T_T = T_B + K \frac{f\, W_n\, (V_1 - V_2)}{1 - \frac{S}{50} \quad \frac{b}{2}}$$

T_T = Total surface temperature

T_B = Bulk stable temperature of the part

S = Surface finish r.m.s. micro inches

It is apparent that we have not yet been able to remove the empirical value of S for surface finish. The work on the coefficient of friction is not complete, and it may be found that the surface finish will affect this value strongly.

As might be anticipated the coefficient of friction is reduced as the velocities increase; thereby giving more advantage to the higher speed gears than would have been predicted by the previous formula, which assumed friction was constant.

The reader will find additional verification to the critical temperature criteria throughout the literature.[5,7] It is important to note that the proponents of this hypothesis are increasing as time goes on.

3. THERMAL STRESSES

The work on thermal contact stresses done by B. W. Kelly is discussed in References 6 and 9. In this paper he calculates the temperatures at and below the surface of a material, and these results are shown in Figure 2.

Since the thermal stresses across the contact zone are proportional to the temperature, the stresses at the surface can be calculated by use of Figure 2. The resulting calculation is illustrated in Figure 3.

When this stress pattern is superimposed on the stress pattern of Figure 21 of Chapter 5 (f = .1) the complete stress range due to the normal load, frictional force, and thermal

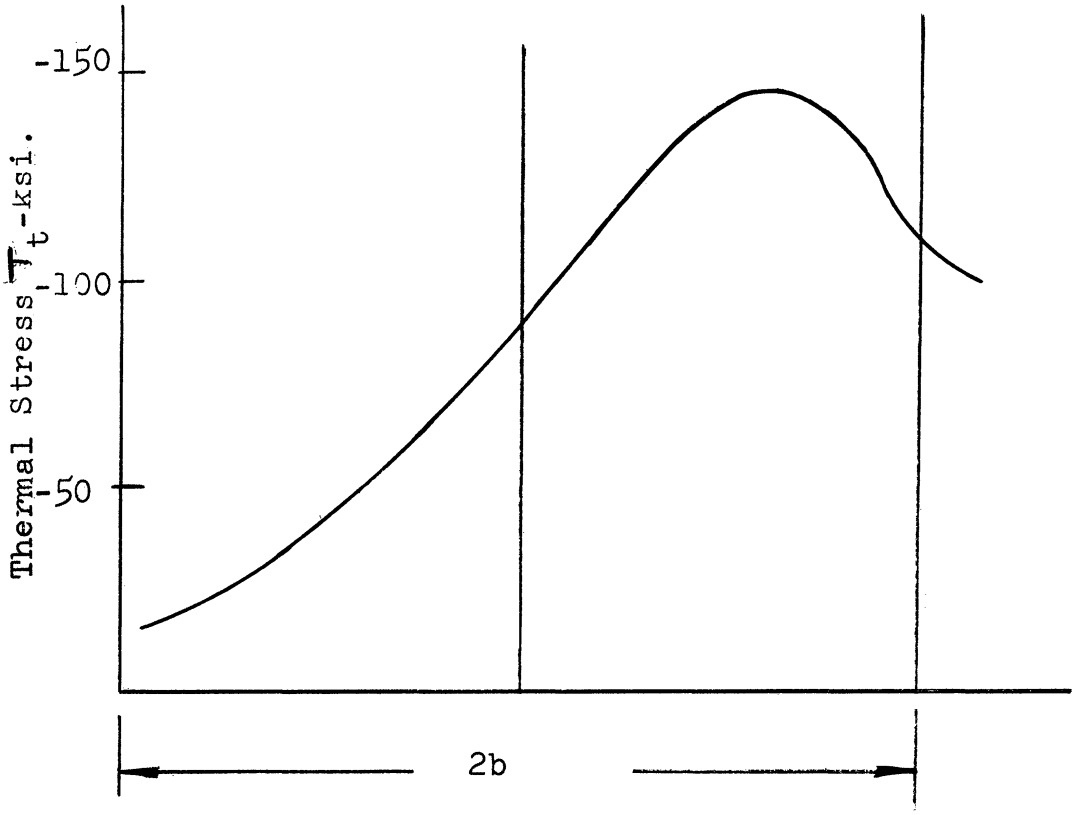


Figure 3

Thermal Stress Over Area of Contact

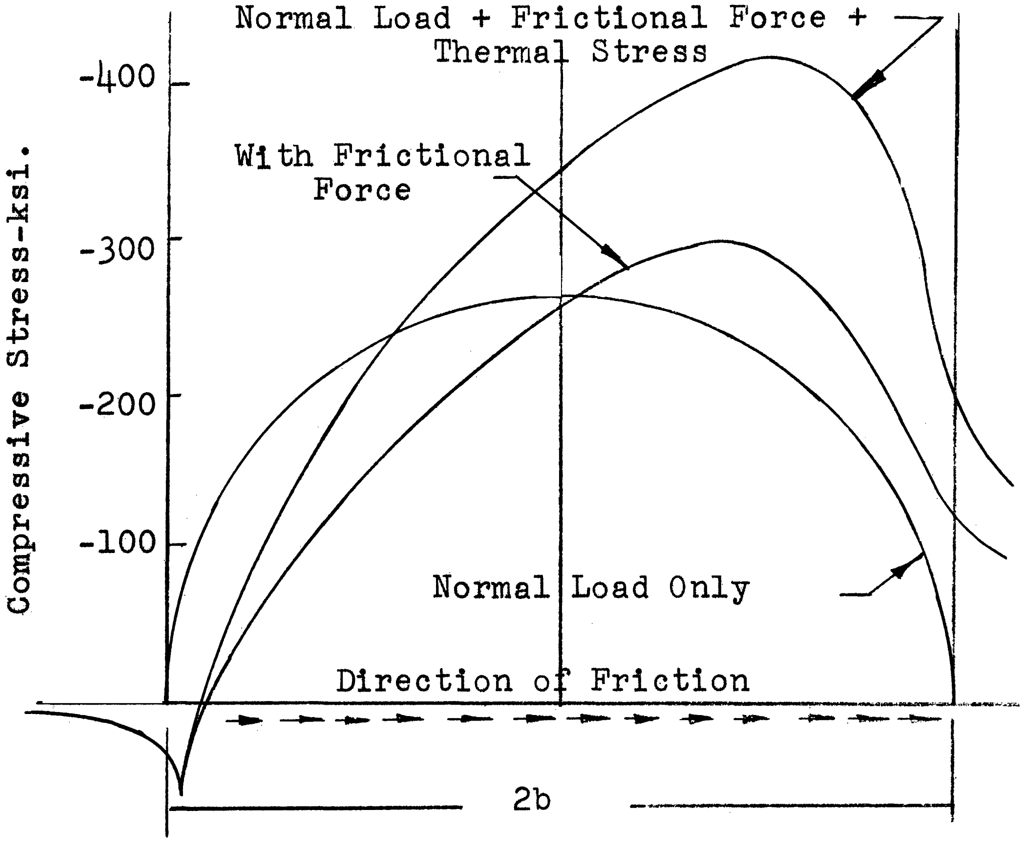


Figure 4

Surface Stress Range for Member in Negative Sliding
(See Figure 5 for Conditions)

stresses is the result. This is shown in Figures 4 and 5 for both positive and negative sliding. (Negative sliding occurs on the slower of two sliding parts.) Note how the thermal stresses act to increase the maximum compressive stress for a member in negative sliding and decrease it for a member in positive sliding. It can also be seen that the normal tensile stress is eliminated by the thermal stress indicating that fatigue failure should not show up for positive sliding members since they are stressed only in varying degrees of compression. On the other hand, members in negative sliding are subjected to both tensile and compressive stresses introducing fatigue characteristics and subsequently pitting.

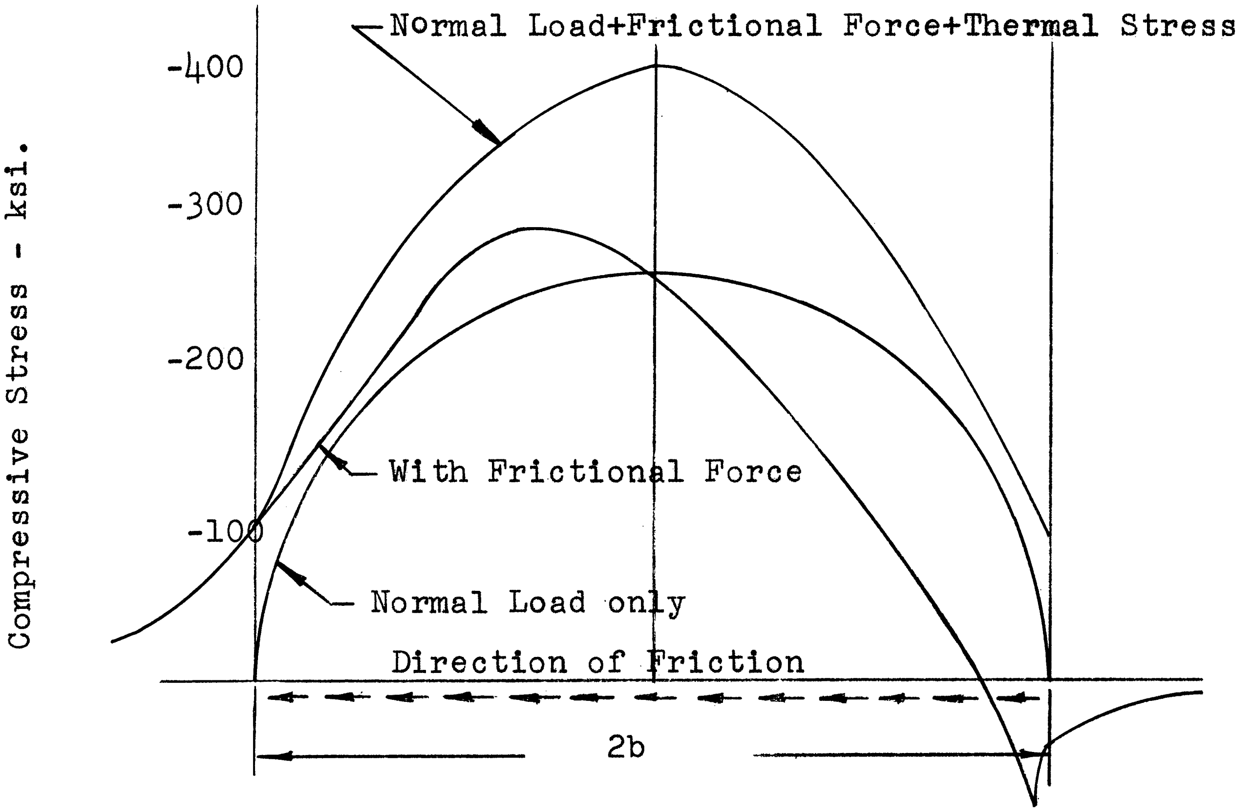


Figure 5

Surface Stress Range for Member in Positive Sliding

adii of Rollers	2.00 and 4.00 inches
$_1$	100 in./sec.
$_2$	150 in./sec.
Load W_N	23,000 Lb. in.
oefficient of Friction	0.1
pecimens	Steel on Steel

Data for Quantitative Study

BIBLIOGRAPHY

1. Bowden and Tabor, "Friction and Lubrication of Solids," Oxford University Press, 1950.
2. Rabinowitz, E., and Tabor, D., Proc. Roy. Soc., 208A, 455, 1951.
3. Murray, Johnson, and Bisson, Lubrication Engineering 10, 193, 1954.
4. Cowley, Ultee and West, "Influence of Temperature on Boundary Lubrication," ASLE Reprint 56LC-18, 1956.
5. Lane, T. B., and Hughes, J. R., "A Practical Application of the Flash-Temperature Hypothesis to Gear Lubrication," Proc. Third World Petroleum Congress, the Hague, 1951, Section 7.
6. Kelley, B. W., "A New Look at the Scoring Phenomena of Gears," SAE Transactions, Vol. 61, 1953.
7. Thomas, H. R., and Hoersch, V. A., "Stresses Due to the Pressure of One Elastic Solid on Another," University of Illinois Bulletin No. 212.
8. Blok, H., "The Dissipation of Frictional Heat", Applied Scientific Research, Section A., Vol. 5, 1955.
9. Kelly, B. W., "The Importance of Surface Temperature to Surface Damage", Engineering Approach to Surface Damage, University of Michigan Press, 1958.

CHAPTER 9

APPLICATION TO ANTI-FRICTION BEARINGS

1. BEARING MATERIAL

Material Properties Required for Rolling Loads

High Contact Fatigue Strength

Contact fatigue strength must be determined under dynamic rolling contact conditions, not by conventional static or fatigue tensile, compressive or bending tests. Flexural and rolling fatigue strengths have no direct relationship. The contact fatigue strength is always related to the life of the part (number of cycles) since it has no infinite life. (see Figure 6)

High Elastic Strength

The material must be elastic enough to spread the load over a sufficient contact area so that the maximum unit contact pressure does not exceed the unit contact fatigue strength of the material. Neither can the material be allowed to deform beyond its elastic limit, causing permanent setting or false brinnelling.

Resistance to Scoring

There is always some sliding between surfaces during operation. The balls or rollers slide a certain percent in the race in the direction they travel depending on the operating conditions. Also they rub against the retaining ring, and they slide minute amounts while rotating in the races due to the geometry of the contact area. Therefore the material must have resistance to scoring.

Structural Stability

In order to maintain structural stability the material must be tempered at a temperature in excess of the designed operating temperature. Hardness is reduced as the tempering temperature is increased. However, we do not want too high a tempering temperature because hardness decreases as the tempering temperature increases. Figure 1 shows the hardness of a few bearing materials at various temperatures.

For bearings to operate at elevated temperatures, the following metallic properties should be considered:

1. Hot hardness
2. Recovered hardness
3. Resistance to scoring
4. Coefficient of thermal expansion
5. Corrosion resistance
6. Purity

Low Level of Non-Metallic Inclusions

When evaluating the effect of non-metallic inclusions, an alloy or carbide segregations on the fatigue strength of a material, the quantity, distribution and orientation of the inclusions and the stress concentration sensitivity of the material must be considered. The average unit quality and size of the inclusions is a general measure of the quality of the material. However, the distribution and orientation of inclusions is just as important. They only inclusions which are detrimental are those located in the high stress regions. The relation between the direction of the inclusions and the principal stresses greatly determines the effect of the inclusions. Since inclusions are harmful because they form stress concentrations the sensitivity

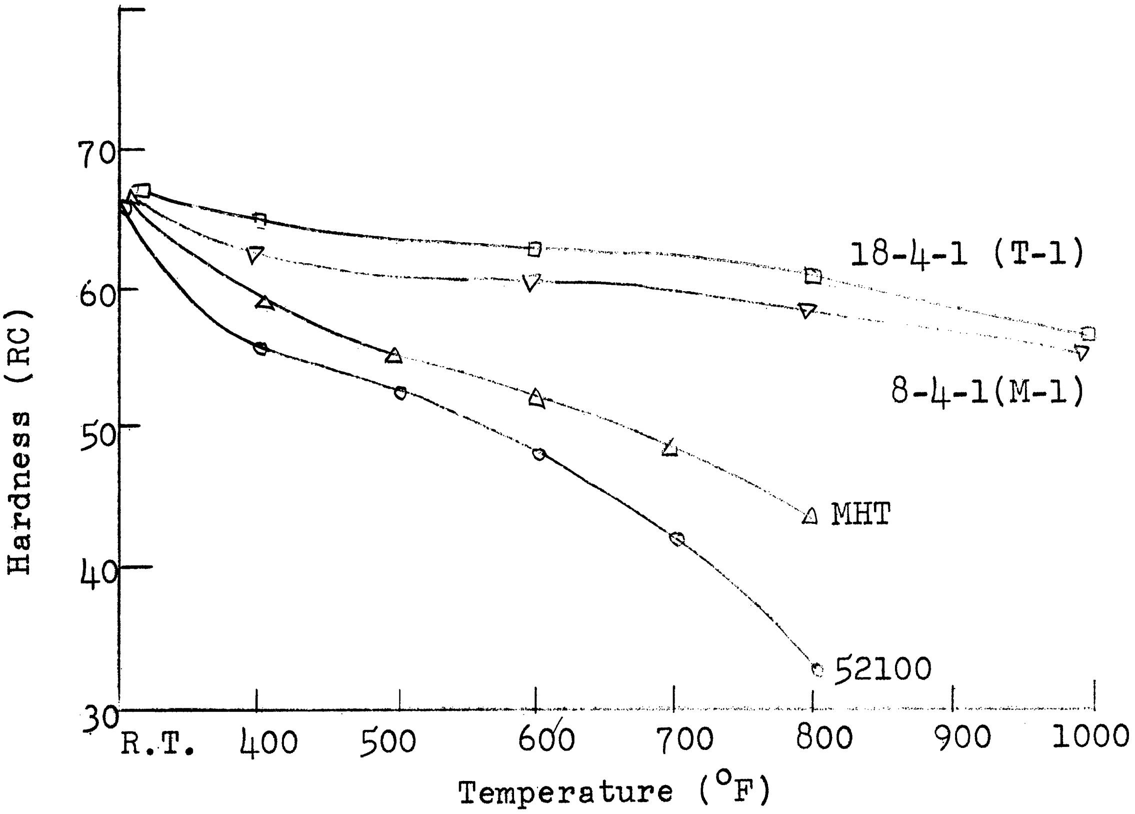


Figure 1

Hot Hardness vs. Temperature of Test

of the material to stress concentrations must be considered when determining the maximum quantity of inclusions permissible.

Vacuum melted steels are much purer than conventional air melted steels. The Weibull plot, of bearing life (to be discussed later) shows that bearings made from these steels are much less susceptible to early failures than conventional bearings. (See Figure 5)[1] A total of three percent of air-melted steel bearings failed in thirty hours of testing; while the vacuum melted steel bearings ran between 100 and 200 hours before a total of three percent of that sample had failed.

Bearing Steels

Bearing steels are classified as either through hardening steels or case hardening steels. Ball and roller bearings are usually made of through hardening steels, and tapered bearings are made of case hardening steels.

The through hardening bearing steels have about .90 to 1.10% carbon, and varying amounts of manganese, silicon and chromium. This steel is hardened at 1525° to 1575° and tempered at 250° to 350°F. The final hardness is 62 to 65 Rockwell "C".

The case carburizing steels have about .10 to .25% carbon, and varying amounts of manganese, silicon, nickel, chromium and molybdenum. This steel is quenched from a carburizing temperature of 1650° to 1725°F, hardened at 1450° to 1525°F, and tempered at 300° to 360°F. The case has about the same carbon content as the through hardened steel; .90 to 1.10%. Case depth, measured to the layer where the carbon content is 50% that of surface, usually ranges from .025 to .255 inches. Surface hardness is 58 to 62 Rockwell "C" and core hardness is 25 to 40.[1]

Some of the steels used in industry for bearings are listed in Figures 2. It will be noted from the figure that one percent carbon steels are used extensively for bearings. This is because of the high elastic and fatigue strength obtainable in these steels. These steels also possess good wear resistance. The heavily alloyed metals are used primarily where chemical corrosion is likely to be a problem or for use at high temperatures.

2. BEARING LIFE

Statistical Life Distribution

The life of a particular bearing even under very carefully controlled operation cannot be reasonably estimated. However, the average life as well as the life distribution of a large group of identical bearings can be accurately predicted.

In the early tests conducted on bearings life ratios of up to 1000 to 1 were not uncommon. But today it is possible to narrow this range to approximately 50 to 1 and in some cases it has been cut down to only 15 to 1. This has been done through more accurate manufacture and testing methods. Also pre-test inspection of ball bearings has been attempted as a measure to cut down the scatter of data. This has not met with much success, however. The relationship of percent of maximum bearing life to percent of bearings tested is given in Figure 3.

Bearing Steels for use to 150°F

- Timken 52100 o
- AISI E-50100
- AISI E-51100
- AISI E-52100
- S.K.F. #1
- S.K.F. #2
- Timken #1
- Timken #2
- SAE 4615 Ni-Mo ø
- Krupp o

o Ball Bearings only
● Roller Bearings only

Bearing Steels for use to 600°F

- U. S. Steel MHT*
- Timken TBS-600
- 440 C
- D-1
- D-2
- 420-F
- Bower 315 φ●
- Timken CBS-600 ø

* 1.35% Al
ø Carburizing Analysis
φ 3.00% Ni

Bearing Steels for use to 1000°F

- M-1 *
- M-10
- M-50
- Halmo

* 1.50 % w

Steels for Tapered Roller Bearings

- AISI 8620
- AISI 4617
- AISI 4320
- AISI 3310
- SAE 465 Ni Mo
- Krupp

Bearing Steels for possible use above 1000°F

- Stellite 19
- Stellite 25
- Sintered Cromium Carbide
- Nickel Bonded Titanium Carbide
- Inconel

Figure 2[1,2]

Bearing Steels

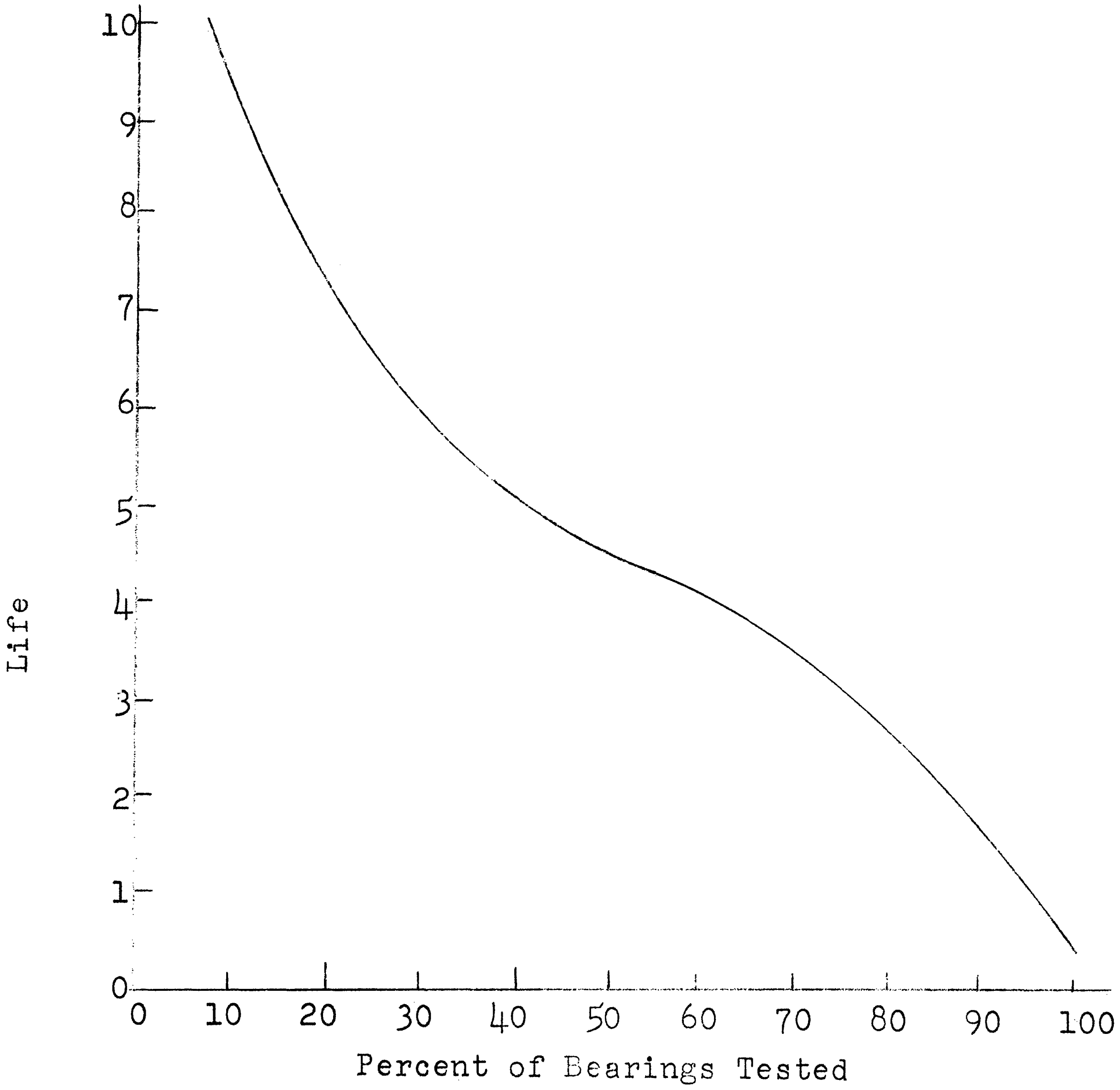


Figure 3 [2]

Life Spread Curve for Bearing Failure

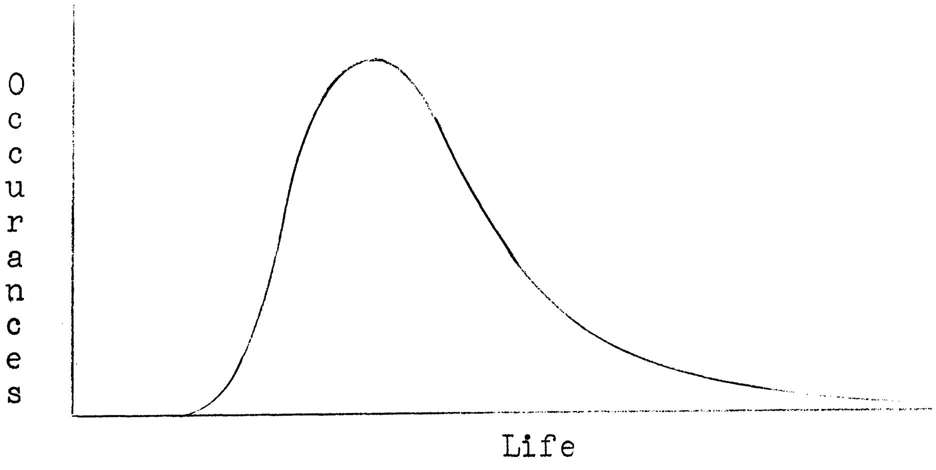


Figure 4 [10]

Life Distribution for Bearing Failure

It can be seen from this curve that more than 90% of the bearings tested reach 10% of the maximum bearing life and that the life for 50% of the bearings is only 45% of the maximum bearing life.[2] This relationship can be seen better on a plot of occurance vs. life as given in Figure 4. This plot shows that the distribution of failures does not follow the normal distribution curve, thus the methods of dealing with the fatigue data for ball bearings cannot be that of normal distribution.[3]

To develop a method of dealing with the fatigue life data it is necessary to make some assumptions. First we assume that the flaws are distributed at random in any body and that their local distribution is also random with respect to their size. The further assumption is made that the probability density of the sizes of flaws is of the exponential type. This assumption is necessary to make the plot follow a Wiebull distribution curve Experimental results have been plotted and found to follow this general pattern.[4] The assumption is made that the number of flaws is very large. With these assumptions the distribution of the largest values. From this it is further assumed that the breaking strength is a linear function of the largest flaw.

The Weibull distribution is useful for plotting cumulative bearing failures versus bearing life (See Figure 5) If a sample of fifty bearings are tested, we simply plot the life of the first failure at two percent failure, the life of the second failure at four percent, etc. Plotted on special statistical paper, these points form almost a straight line. Only a fraction of the sample group need be tested until failure in order to determine

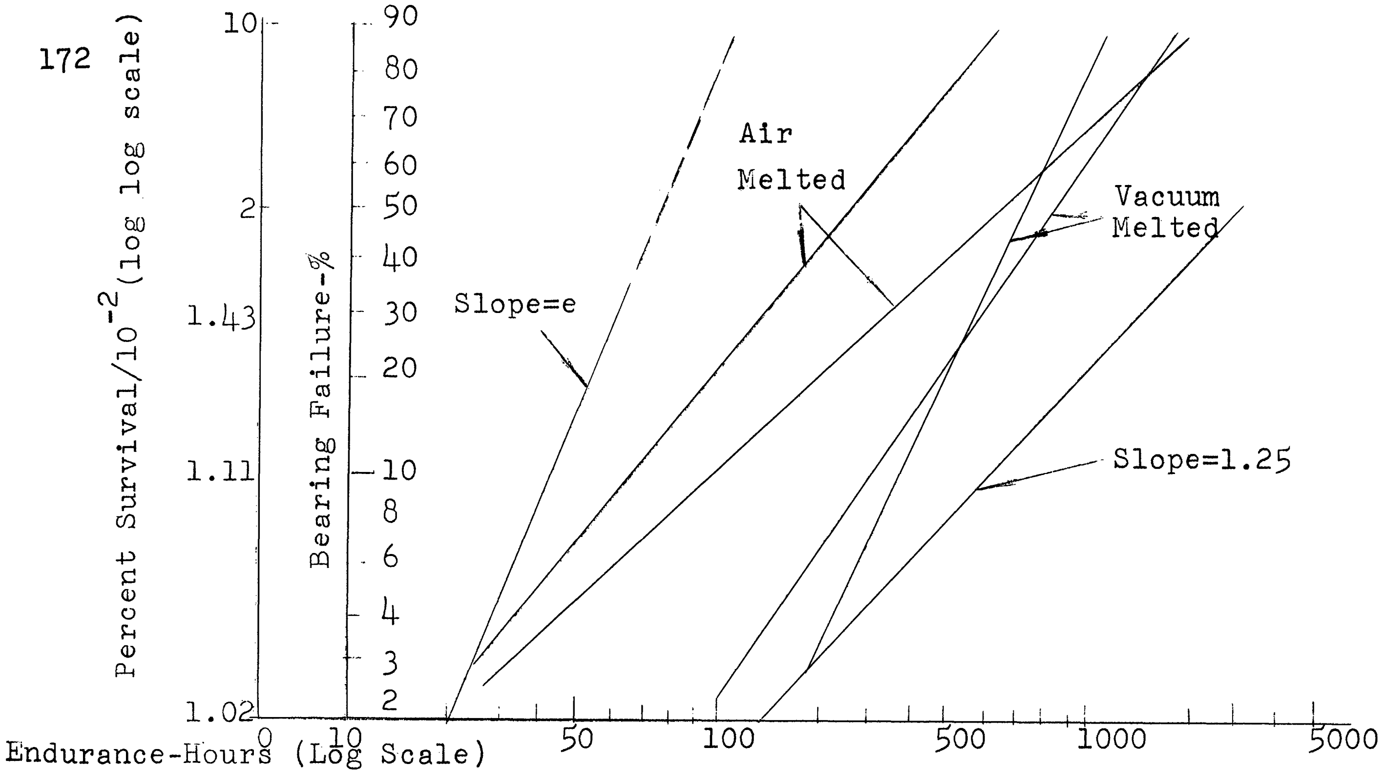


Figure 5

Weibull plot of bearing life for ball bearings made from air melted and vacuum melted steels. (by L. D. Cobb, SAE Preprint, March, 1957.

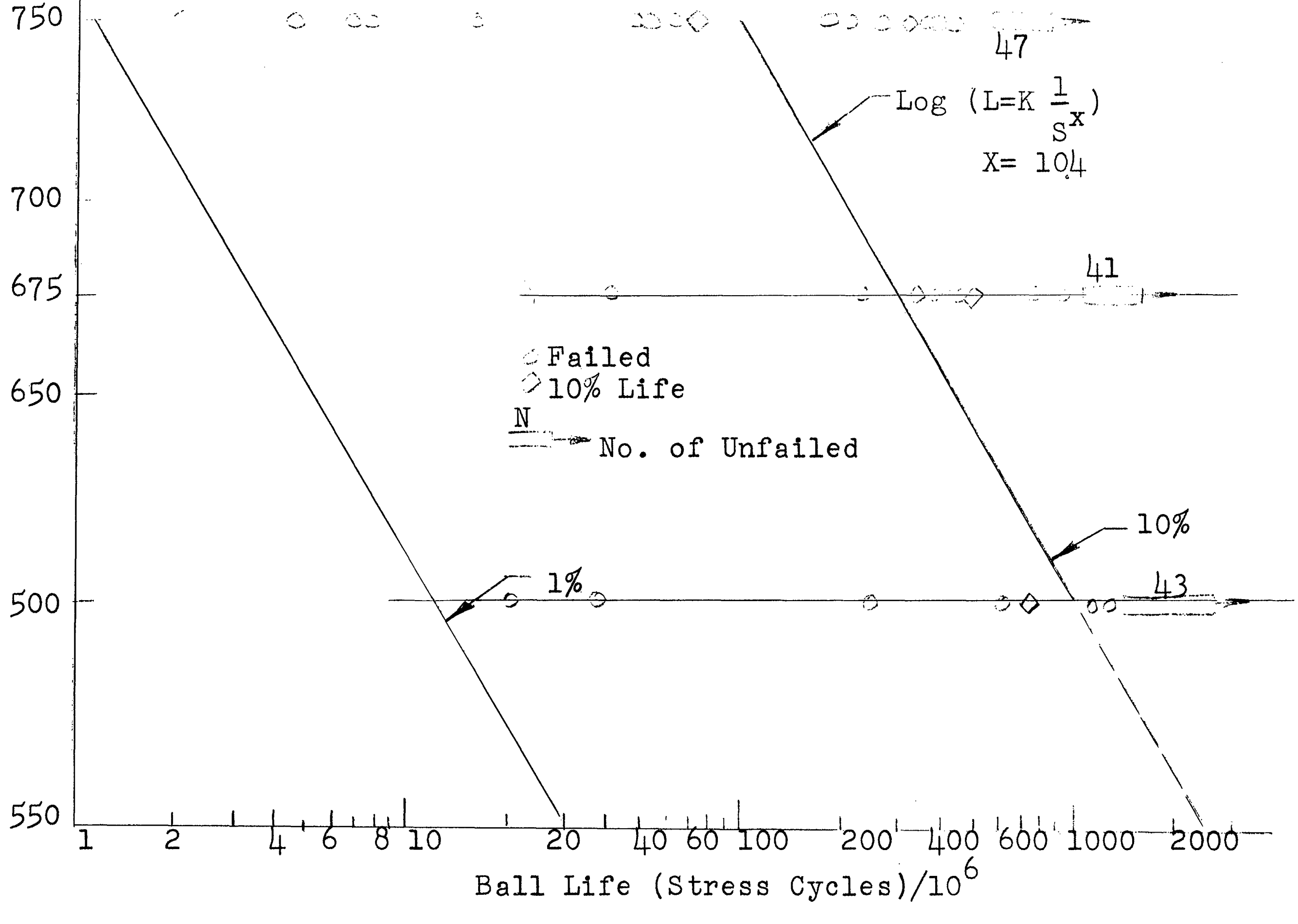


Figure 6

Variation of ball life with applied stress. Balls, grade, 9/16 inch. SAE 52100 steel; room temperature.

From: NACA Tech. Note 3930, p. 19.

the statistical distribution of the group. However, all unfailed bearings must be run at least as long as all the failed bearings.

The Weibull distrubtion equation[3] is:

$$\log \frac{1}{s(L)} = KL^e$$

where:

$s(L)$ = percent survival og group at life L
K = constant
e = natural log base

This equation can be rewritten so that it will form a straight line plot as follows:

$$\log \log \frac{1}{s(L)} = e \log L + \log K$$

This equation is in the common form

$$y = mx + b$$

Therefore, if we plot the equation on log log versus log paper we get our straight line with slope e. Last we change the ordinate scale so that we can read directly the percent bearing failure.

If the slope of the plot was always "e", we would only need a few points to determine the distribution of a group. However, the slope for ball bearings varies as shown in Figure 5, so that enough points must be determined to establish the slope for the life of a particular type of bearing.

A plot of frequency bearing failure versus life of bearing on linear-scaled graph paper gives a distribution which is skewed strongly to the right rather than a normal type distribution.

Stress-Life (S-N) Curve

The S-N curve of ball bearings is approximately a straight line when plotted on log log paper (See Figure 6). Contact stresses in ball bearings, give an S-N curve for flexural stresses. The stress plotted in the figure is the applied maximum Hertzian stress as calculated from modified Hertzian equations. These equations for ball bearing, which are very complex, can be found in Reference 5.

Figure 6 is an S-N curve for a grade 1, 9/16" ball made of SAE 52100 steel. Data was obtained from tests run at room temperature in the NACA Spin Rig. The 10% failure line was drawn at NACA. The 1% failure line has been drawn in Figure 6 . The 1% and 10% failure lines have been extended out to 10" stress cycles in Figure 7. Extension of the plot beyond the limits of actual data gives us only an estimate which should not be relied on too heavily. The 1% failure line was computed from the Weibull distribution equation. The constant was calculated from the 10% failure line of Figure 6 and the slope was estimated from Figure 5. The stress (S in psi)-life (L in stress cycles) equation is:

$$L = \frac{K}{S^{10.4}} \quad ; \qquad \begin{aligned} K &= 1.54 \cdot 10^{67} \text{ for 1\% failure} \\ &= 1.25 \cdot 10^{69} \text{ for 10\% failure} \end{aligned}$$

Relationships of Life, Load, Stress and Size Factors

$$\frac{L_1}{L_2} = \left(\frac{P_2}{P_1}\right)^x , \qquad x \approx 3 \quad \text{(between 2.7 and 3.3)}$$

$$\frac{1}{2} = \left(\frac{P_1}{P_2}\right)^{\frac{1}{3}} , \quad \frac{P_1}{P_2} = \left(\frac{d_1}{d_2}\right)^{2}$$

$$\frac{L_1}{L_2} = \left(\frac{\tau_2}{\tau_1}\right)^{3x}$$

where :

L = life
P = normal load
τ = maximum Hertz stress
d = diameter of ball

If the load on one of two identical bearings is twice that of the second, the statistical life of the first will be only 1/8 that of the second. Similarly, for a 2:1 stress ratio the life ratio becomes about 1:1,000. The tremendous effect of these factors probably cause a major share of the large spread of bearing life distributions.

Bearing Fatigue Tests

Bearings are tested to establish load-life ratings and to determine the effect of variables on life. These tests are very expensive and time consumming. Each variable measured requires the testing of a substantial sample group in order to deterimine the statistical life. Usually, bearings are run at an over-load condition to reduce the time for failure to occur, and then the life for normal load is calculated.

Special test apparatus has been designed to simplify bearing testing. One which has proved very useful is the Contact Fatigue Sping Rig developed and operated by NACA. (See Figure 8) Two test

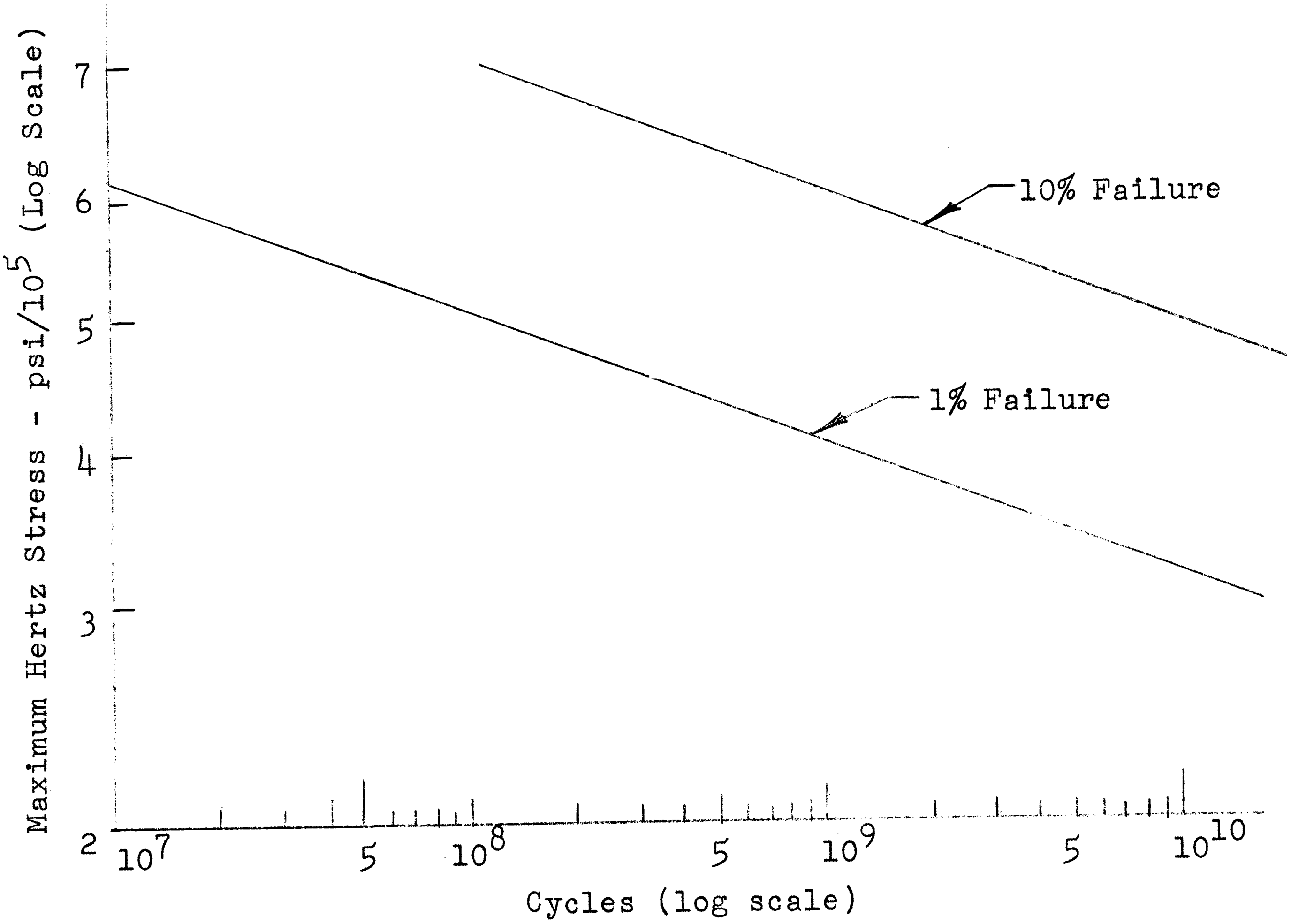


Figure 7
(Extension of Figure 6)

For 52100,
grade 1, balls
air melted ste

$$L = K\left(\frac{1}{S^{10.4}}\right)$$

$K \cong 1.24 \times 10^{69}$ for 10 o/o failure
$K \cong 1.54 \times 10^{67}$ for 1 % failure

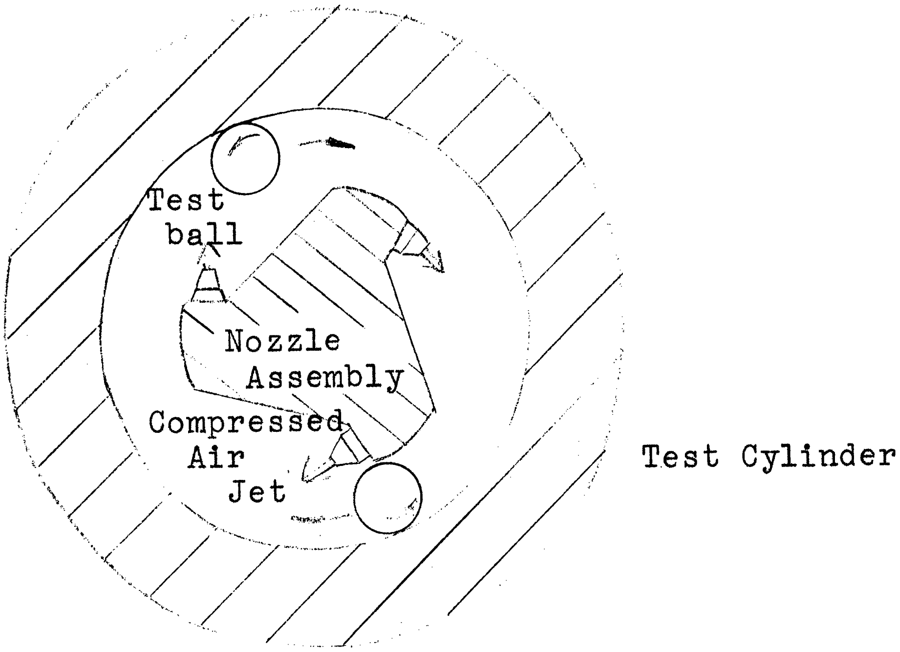


Figure 8
Schematic Diagram
NACA Rolling -Contact Fatigue Spin Rig From NACA Tech. Note 3930

balls are driven by jets of compressed air around a test race cylinder at orbital speeds up to 30,000 rpm producing centrifugal forces greater than 700 pounds and maximum Hertz stresses of about 750,000 psi. A measured amount of mineral oil is injected into the air stream to give uniform lubrication. The orbital speed and number of cycles are measured by an electric tachometer as the balls interrupt a light source to a photocell. When a bearing fails, it causes a vibration which is detected by an amplified vibration pickup which trips a meter relay to shut the test equipment off. The orbital velocity is controlled by adjusting the air pressure, and the system can hold the velocity within 120 rpm at 30,000 rpm.[6]

This test apparatus enables one to control environmental conditions very accurately, measure variables precisely, and reduce test time to a minimum. The results are similar to results of full bearing tests for stress-life and failure-life distributions. The test rig gives acceptable data for comparing bearing materials for ball and race use.

Fatigue Tests on Bearing Steel

Styri[7] made numerous tests upon a bearing steel. (SAE 52100) These tests were made on ring specimens, rotating beam specimens and torsion specimens. The findings of these tests are given in Figure 9.

The results of these tests indicate that the fatigue life is a function of the hardness of the metal. The tests also indicate that for SAE 52100 steel the hardness has little or

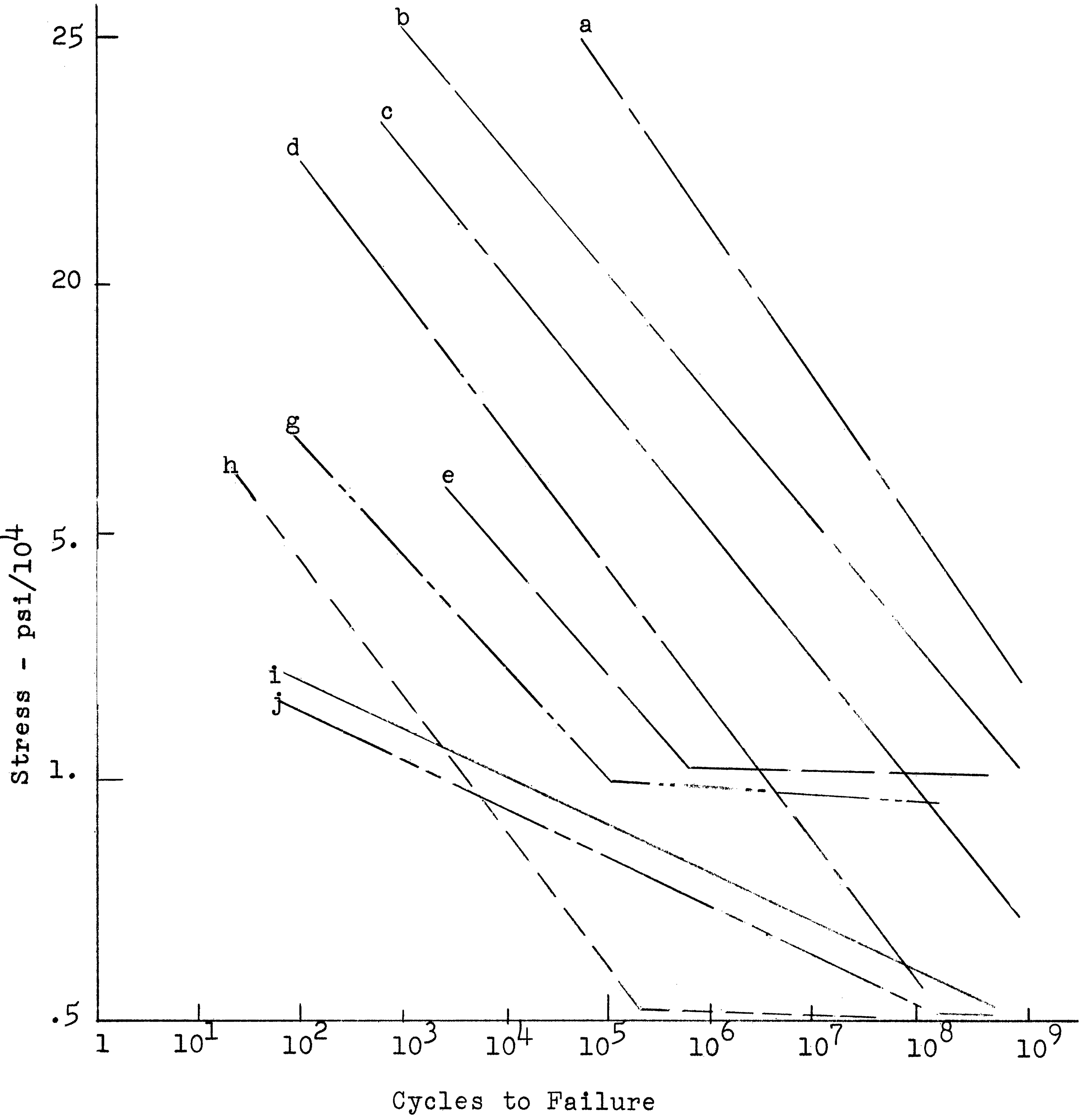


Figure 9 7

Endurance of SAE 52100 Steel

Specimen Type	Rockwell Hardness "C" Scale	Letter on Graph
Rotating Ring	60 and Above Various heat Treatments	c
Rotating Beam	"	b
Rotating Beam Vacuum melted	"	a
Torsion	"	d
"	45.0	i
"	32.4	j
Rotating Beam	38.1	g
Rotating Beam	17.0	h
Rotating Beam Vacuum Melted	42.5	f
" "	30.0	e

no effect for hardnesses around 60 Rockwell "C",but at the hardness decreased from this value the endurance limit decreases rapidly. Figure 10 from Rowland's paper gives the effect of both carbon content and hardness upon endurance limit. The data from Styri's tests upholds this as SAE 52100 steel is of approximately 1% carbon.

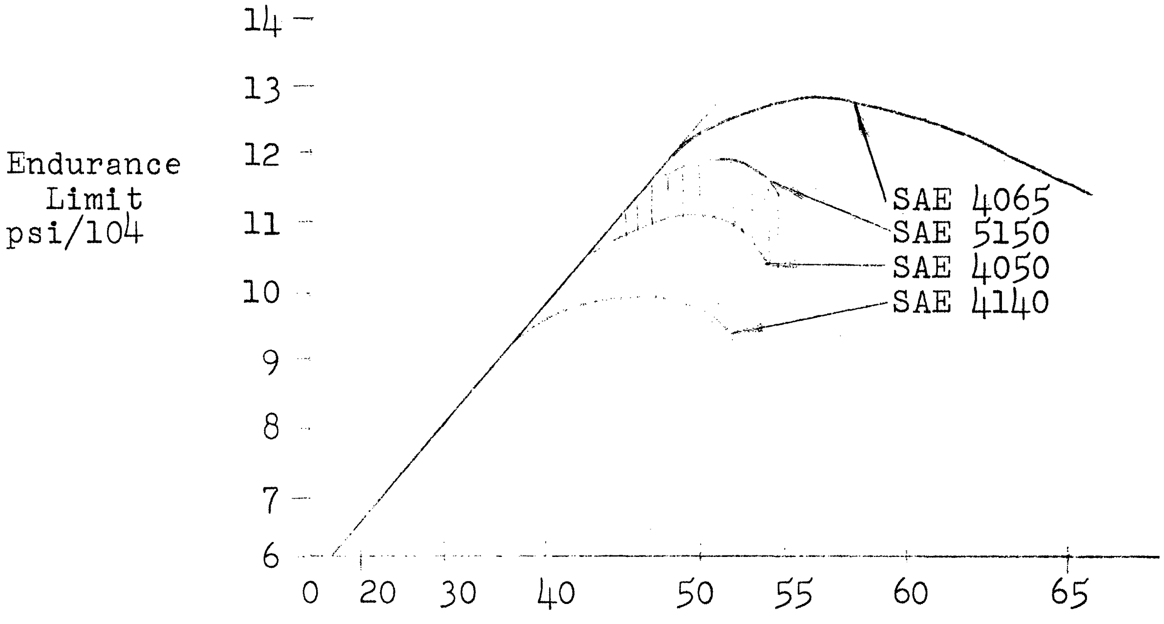
The fatigue life was found to be independant of the procedure used to obtain the given hardness. It also indicated that the scatter of data is inversely proportional to the load. It was also found that vacuum melted steel has a longer fatigue life than standard steel but that the scattering of data was not improved. No endurance limit was found for the 60 Rockwell "C" and higher hardness of the SAE 52100 steel. It was assumed, however, that even though these tests were carried out to 10^9 cycles that an endurance limit exists. This assumption was based on the fact that the same steels of lower hardness were found to possess endurance limits.[7]

Observations of Fatigue Tests on Bearing Steel

The results of the above tests suggest that the cause of failure in ball bearings may be similar to the failure of torsional specimens. This was concluded from the fact that when using mean normal pressures in the equation

$$\frac{L_1}{L_2} = \frac{S_2}{S_1}^{x}$$

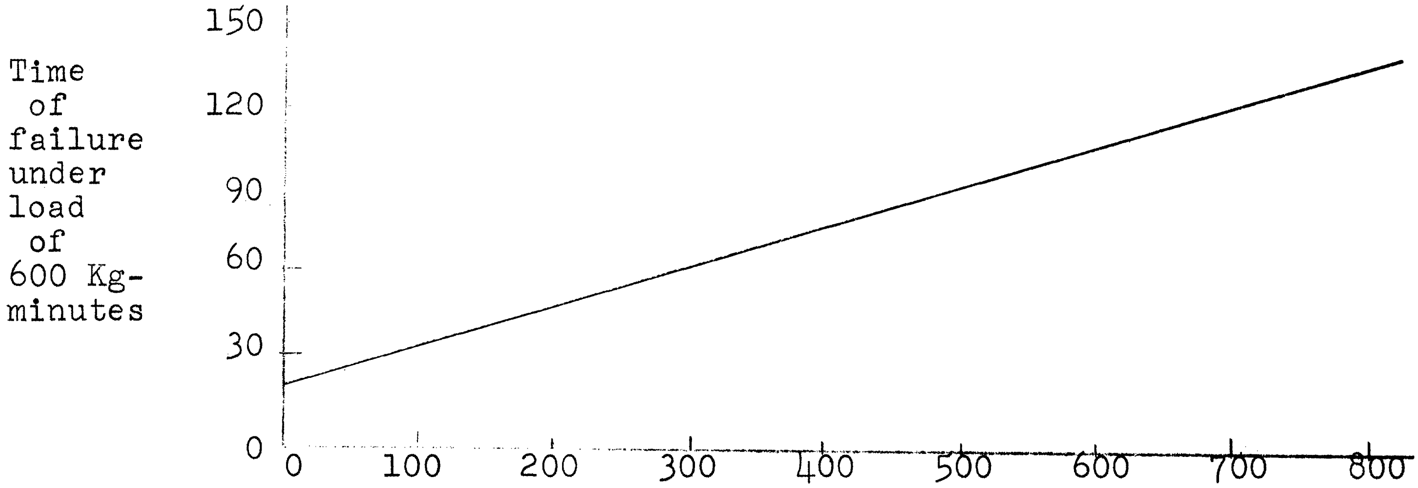
the value of x was found to be ten. This is the same value as



Hardness - Rockwell C

Figure 10

Effect of Carbon Content and Hardness on Endurance Limit in Bending



viscosity at 20°C - Centistokes

Figure 11 [9]

Effect of Viscosity on Life for Hydrocarbon Lubricants obtained in the torsion test specimens. From this it was further concluded that since failures in most of the torsion tests started in the direction of alternating shear stress, that starting points for fatigue flaking might be expected to be found in areas where high shear stresses prevailed. According to the work of Lundberg and Palmgren this area would be at a depth of 1/3 to 1/2 of the minor axis of the contact ellipse. The magnitude of this stress is approximately .43 to .5 of the maximum pressure at the contact point.

In agreement with the above, Styri found that flaking usualy started below the surface and then progressed to the surface. He found that approximately 9 out of 10 of the failures started below the surface. Numerous cracks started at slag inclusions, but as was predicted from the vacuum melted steel tests, they are not the only reason for cracking. It was found that most of the cracks started in approximately the area of maximum shear stress variation. It was found that slag inclusions were important only when they were in this area of maximum shear stress variation. It was found that cracks could start at local weak points due to crystal structure or segregation, but the main source of cracking (other than inclusions) resulted from microstresses created by the austinite to martensite transformations.[7]

The failure of anti-friction bearings under normal conditions is generally due to pitting or spalling from surface fatigue. This process is a progessive breaking away of the surface metal which starts by the development of surface cracks.

These cracks propagate into pits and larger areas of displaced metal called spalled areas. This is a direct contradiction of the experimentally observed results of Styri. The surface crack explanation of failure given above is the result of a theory advanced by Stewart Way of Westinghouse.

Way's theory suggests that cracks start at the surface. The development of surface fatigue or spalling is hastened by the presence of a fluid which enters the cracks. Just before the crack comes into the load zone the size of the crack opening is increased, due to the increasing pressure, and this allows a quanity of fluid to enter the crack. As the crack further advances into the area of the load the crack is forced closed trapping fluid at the root of the crack. When this area is under load the fluid will develop a tremendous hydraulic force which increases the size of the crack and ultimately prys out small sections of the surface metal. Way further predicted the direction of the cracks that would cause failure. This direction is a crack whose root is in the opposite direction as that of rotation. This prediction of crack direction has been given support by experimental evidence. He produced further evidence to support his theory by running tests without the presence of a fluid, and found that no spalling was produced even though cracks developed in the surface.[2]

A possible explanation of this apparent contradiction would be that the type of failure depends upon the material. Harder more brittle materials would be more likely to fail from tensile stresses than shear stress and since the largest tensile stresses are at the surface, cracks would start at the surface and proceed according to Way's theory. But for softer more ductile materials which would be more likely to fail due to shear stresses than tensile stresses, the cracks would start below the surface at the area of maximum shear stress and then work their way to the surface. In the consideration of case hardened materials the type of failure would depend upon the relative distances below the surface of the area of maximum shear stress and case depth. Also the type of failure would depend upon the magnitudes of the shear and tensile stresses, the presence of fluid, and the presence of surface cracks.

2. OTHER FACTORS OF BEARING LIFE

Lubrication

Proper lubrication is essential for efficient and economical use of bearings. The life of a bearing is strongly dependent upon the lubrication the bearing receives. On the modern antifriction bearings the lubricant must be able to dissipate the heat generated in operation. It must also help to reduce the friction of the bearing.

Most lubricants now used are straight petroleum products or are compounded from petroleum products. The viscosity of the lubricant is very important, its selection depends upon the particular operation at hand. However, it may be stated as a general rule that the higher the viscosity the longer the bearing life. See Figure 11 from tests by Barwell and Scott.

A ferrous metal unprotected from corrosion, (which may be due to water, acids or alkalis,) has its endurance limit decreased greatly by the development of surface pits and cracks. These cracks are local points for early fatigue failure. Because of this, anti-friction bearings must at all times be protected by a film of lubricant which will assist in inhibiting corrosion. Corrosion fatigue can occur very rapidly, particularly in summer with high humidity.[8]

Experiments have been conducted to determine the efffect of lubrication of the pitting failure in ball bearings. The most extensive experiments were run by Scott and Barwell and the results are given in Figure 12.

The results of this test showed that viscosity has a very pronounced effect upon the life. But it was not the only contributing factor as some types of lubricants with a high viscosity failed before other types with a lesser viscosity. As was stated before generally the life is directly proportional to viscosity.[9]

Effect of Machining upon Bearing Life

Tests were run on SAE 52100 steel to determine the effects of grinding, polishing, shot peening and abrasive tumbling. These tests were run by Tarasov and Grover. The data from these tests is given in Figure 13. Flat bars were used in the tests.

It can be concluded from these tests that neither hand polishing nor electolytic polishing after grinding affect the

No. of Tests	Lubricant	Load	Hertz stress ton/in.^{2}	Viscosity at 20°C.	Min. time range of failures	Average time to failure Minutes
11	Water Glycol	600	475	44	16-40	31
8	Water Glycol	600	475	95	16-40	28
5	Polyalkalane Glycol	600	475	16	10-12	11
6	Polyalkalane Glycol	600	475	356	17-33	26
6	Dimethyl Siloxane	600	475	1-6	50-65	58
5	Dimethyl Siloxane	600	475	115	48-60	52
6	Silicone lubricant as Hydraulic fluid	600	475	52	59-70	64
7	Naphtlenic Base oil	600	475	101	37-72	58
6	Low V.I. Oil Napthlenic Base oil	600	475	154	59-85	64
6	Straight Mineral oil	600	475	605	97-115	107
7	Straight Mineral oil	600	475	757	105-150	138

Figure 12

Continued Effects of Lubricants

Hardness Rockwell C	Grinding	Additional Processing	Endurance Limit
59	Gentle	None	65,000
59	Severe	None	45,000
59	Gentle	Hand Polishing	59,000
59	Severe	Hand Polishing	47,000
45	Gentle	None	60,000
45	Severe	None	45,000
45	Gentle	Electrolytic Polishing	60,000
45	Gentle	Shot Peening	78,000
45	Severe	Shot Peening	83,000
45	Severe	Abrasive Tumbling	73,000

Figure 13 [10]

Effects of Grinding, Polishing. Shot Peening, and Abrasive Tumbling endurance limit. Gentle grinding parallel to the applied stress left no significant residual stress and did not affect the endurance limit. It was expected that, had the grinding been perpendicualr to the applied stress, it would have had some effect due to the notch effect of the scratches. Severe grinding decreased the endurance limit. Both shot peeing and abrasive tumbling increase the endurance limit. The damage done by severe grinding was eliminated and the endurance limit actually increased due to the shot peening and tumbling.[10]

Orientation of Fibers in Bearing Balls

Steel balls as manufactured for bearings are made by upsetting slugs of steel rod (or wire) between hemispherical dies. The rough ball is heat treated and rough and finished ground. The rod has longitudinal fibers formed by the extrusion at the mill. Upsetting the rod changes the fiber orientation to the form shown in Figure 14. The fiber divides the ball into two polar regions and one equator region.

Tests were run by NACA in the Spin Rig to determine the effect of the fiber orientation on ball life. The axis about which the ball rotated was controlled by drilling a hole through the ball or grinding flats on the ball. Three groups of balls were tested: 1. balls that rolled over polar regions; 2. balls that rolled about the equator; and 3. balls that had random tracks. Results of the tests showed that appoximately twice as many failures occured in the polar region than would be expected

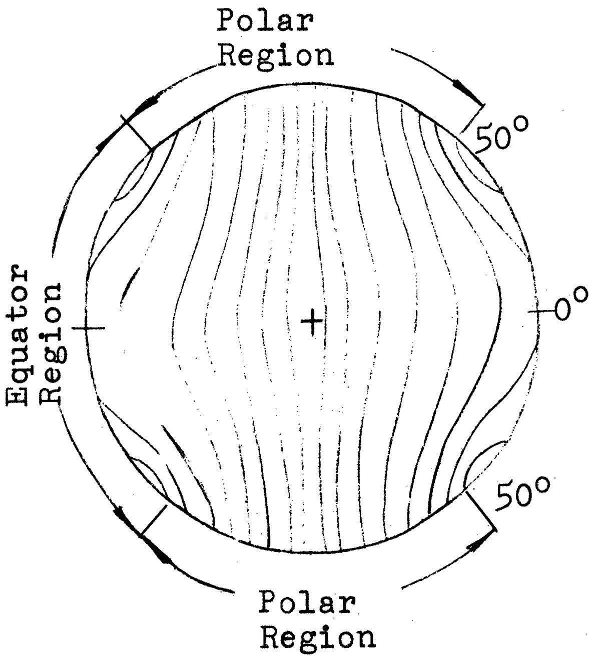


Figure 14

Cross Section of Ball Showing Fibers
From NACA Tech. Note 3933

if the balls were perfectly homogeneous.[11] We may conclude that the polar region of balls are significantly weaker in fatigue strength than the non-polar region.

4. BEARING RATINGS

Method of Rating Bearings

Commonly a bearing is rated for load carrying capacity based on a specified number of hours of life at a certain rpm For example, the rated load for a bearing with a mean life of 500 hours operated at 3000 rpm might 150 pounds radial and 15 pounds thrust force. This means that if a large group of these bearings were tested at the rated loads and rpm for 500 hours, 50% of the bearings would fail and 50% would still be good. Sometimes the bearings are rated for a life when 10% failures occur.

In order to select bearings for a specific application, one must know the load and speed conditons that will be encountered, the life the product is designed for, and the maximum percent of failures permissible.

Calculation of Bearing Life

The following problem is an example of how to compute the life of a bearing from its bearing rating;

Bearing rating:

Mean Life $L_r = 500$ hours at $Nr = 1000$ rpm

with load $P_r = 1500$ pounds

Operating conditions:

% of time	load (P)	rpm
30%	900	1000
15%	2000	300
55%	700	200

Desired failure level = 5%

Calculations:

Equivalent related lives at operating conditons:

$$L = \frac{P_r}{P}^3 \quad \frac{N_r}{N} \, L_r$$

$$L_1 = \frac{1500}{900}^3 \quad 1 \quad 500 = 2315 \text{ hours}$$

$$L_2 = \frac{1500}{2000}^3 \; \frac{1000}{300} \quad 500 = 702 \text{ hours}$$

$$L_3 = \frac{1500}{700}^3 \; \frac{1000}{200} \quad 500 = 24{,}600 \text{ hours}$$

Combined life

$$\frac{1}{L} = \frac{\text{\% of time (decimal) at conditon n}}{L_n}$$

$$\frac{1}{L} = \frac{.30}{2315} + \frac{.15}{702} + \frac{.55}{24{,}600}$$

$$L = 2{,}730 \text{ (for mean life)}$$

To determine Life for 5% failure:

Plot 50% failure life of 2730 hours on Weibull graph, Figure 5 line A, and draw line with a slope of 1.25. Read life for 5% failure.

Answer: 330 hours.

BIBLIOGRAPHY

1. Rowland, E. S., "Resistance of Material to Rolling Loads", Engineering Approach to Surface Damage, edited by C. Lipson, and L. V. Colwell, University of Michigan Press, 1958.
2. Boden, E. G., "Capacity Ratings of Anti-Friction Bearings," Anti-Friction Bearing Symposium, Timken Roller Bearing Co., 1939.
3. Weibull, W. A., "A Statistical Distribution Function of Wide Applicability", Journal of Applied Mechanics, Vol. 18, No.3, September, 1951, p. 293.
4. Gumbel, E. J., "Statistical Theory of Extreme Values and Some Practical Applications," National Bureau of Standards, Applied Mathematics, Series No. 33, 1954.
5. Jones, A. B., "Analysis of Stress and Deflections", New Departure Engineer Data, New Departure Division, General Motors Corp., Bristol, Connecticut, 1946.
6. Butler, R. H., and Carter, T. L., "Stress Life Relation of the Rolling Contact Fatigue Sprin Rig", NACA, Technical Note 3930, March , 1957.
7. Styri, H., "Fatigue Strength of Ball Bearings Races and Heat Treated 52100 Steel Specimens," Proc. ASTM, Vol 61, 1951.
8. Maag, O. L., "Lubrication of Anti-Friction Bearings," Anti-Friction Bearing Symposium, Timekn Roller Bearing Co., 1939.
9. Barwell, F. T., and Scott, D., "Effect of Lubricant on Pitting Failure of Ball Bearings," Engineering, July 6, 1956.
10. Tarasov, L. P., and Grover, H. J., "Effects of Grinding and Other Finishing Processes on the Fatigue of Hardened Steel," Proc. ASTM, Vol. 50, 1950.
11. Butler, R. H., et. al., "Effect of Fiber Orientation on Ball Failures Under Rolling Contact Conditions", NACA Technical Note 3933, February, 1957.

CHAPTER 10

APPLICATION TO GEARS

1. PITTING BY SUBSURFACE FAILURE

Stress Distribution

Most subsurface failure theories of pitting are based on the concept of Hertzian stresses. The theory which will be presented here utilizes modified Hertzian concepts in a more comprehensive form because all stress components are defined by equation not only on the contact plane but everywhere below the surface of the contact area. The contact stress picture is as shown in Figure 1. The principal stresses are X,Y, and Z, and in the region of contact are everywhere compressive. At any point under the contact area the principal stresses are given by the equations:

$$X = \frac{4p'z\theta}{\pi b^2(1+2\theta)}\left[\operatorname{cosec}\left(\tan^{-1}\frac{\sqrt{u}}{b}\right)-1\right]$$

$$Y = \frac{2p'z}{\pi b^2}\left[\sqrt{\frac{u(b^2+u)}{(b^2-y^2-z^2)+4b^2z^2}}+\sqrt{\frac{b^2+u^2}{u}}-2\right]$$

$$Z = \frac{2p'}{\pi}\left[\frac{z^3\sqrt{u+b^2}}{u\sqrt{u}\,(2u+b^2-y^2-z^2)}\right]$$

where:

$$2u = \sqrt{(b^2-y^2-z^2)^2+4b^2z^2}-(b^2-y^2-z^2)$$

x,y,z are the orthogonal coordinate components

p' = Load per unit length of line of contact of cylinders

θ = Coefficinet of elasticity in Kirchoff's notation

In the plane of contact the principal stresses are defined by the following equations;

$$X = \frac{4p'}{\pi b^2}\sqrt{b^2-y^2}\ ,\quad Y = \frac{2p'}{\pi b^2}\sqrt{b^2-y^2}\ ,\quad Z = \frac{2p'}{\pi b^2}\sqrt{b^2-y^2}$$

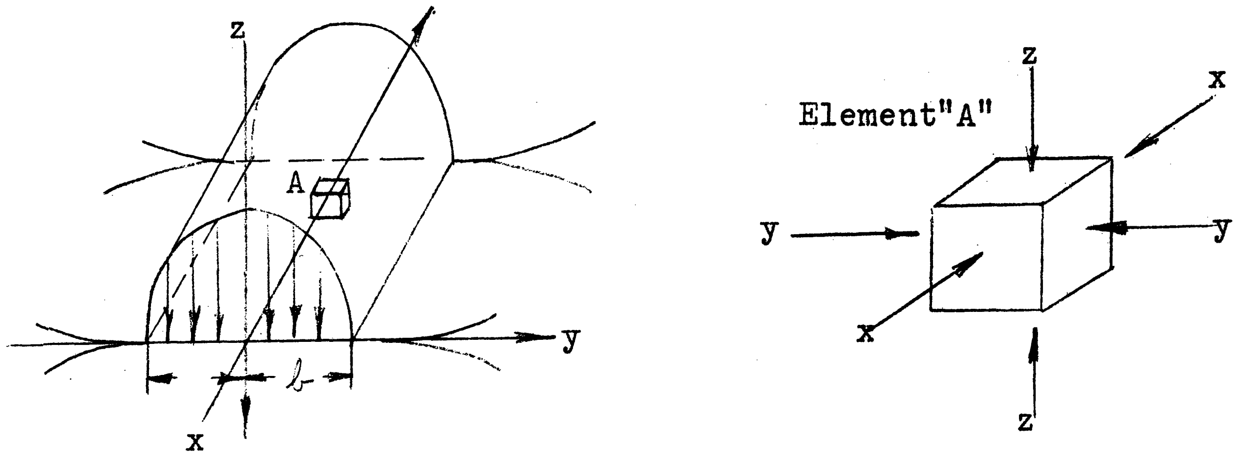
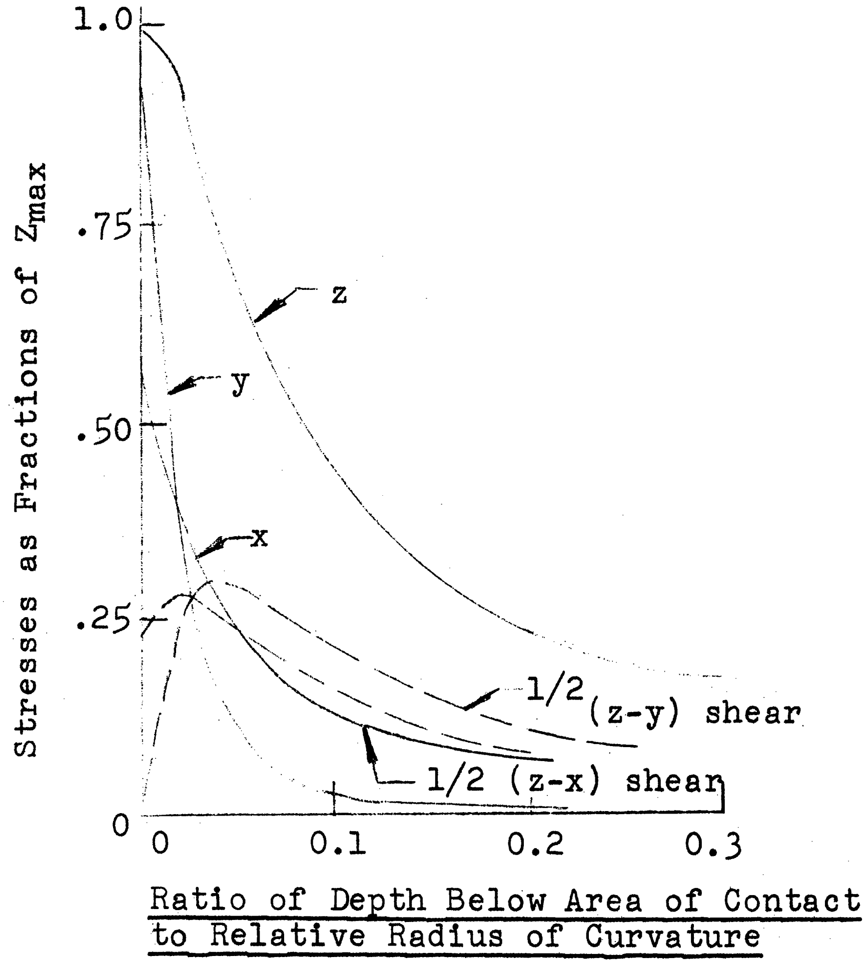


Figure 1

Hertz Contact Stress Picture



$$R = \text{Rel. Radius of Curv.} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}}$$

Figure 2

Beeching-Nicholls (Hertz) Stresses at Plane of Symmetry where it should be noted that at $y = b$, all three principal stresses are equal to zero.

To obtain the familiar Hertzian stress distribution (Figure 2) along the line of symmetry of the contact area, set $y = 0$ in the above equations and get $u = z^2$, and:

$$X = \frac{4p'\sigma}{\pi b^2}\left[\sqrt{b^2 + z^2} - z\right]$$

$$Y = \frac{2p'}{\pi b^2}\left[\frac{(\sqrt{z^2 + b^2} - z^2)^2}{\sqrt{z^2 + b^2}}\right]$$

$$Z = \frac{2p'}{\pi\sqrt{z^2 + b^2}}$$

where σ = Poisson's ratio

The complete derivation of above general equations for the principal stress components is a complex problem utilizing the mathematical tools of advanced calculus. To the majority, presentation of this derivation would be boring and have little physical meaning; hence it is omitted. Those who are interested in following the complete derivation will find it in the paper by Beeching and Nicholls (reference 3).

An important point to note is that Beeching and Nicholls derivation yields a value completely in accord with that of Hertz for the "maximum pressure at the center of the ellipse" (q_o).

Hertz gives $q_o = \frac{2p'}{\pi b}$. Beeching and Nicholls give $Z = \frac{2p'}{\pi\sqrt{z^2 + b^2}}$

where $z = 0$ .... $Z = \frac{2p'}{\pi\sqrt{0 + b^2}} = \frac{2p'}{\pi b} = q_o$

Method of Failure

The Nicholls-Beeching failure theory is based on failure in shear at the point of maximum shear under the area of contact. As seen in Figure 2 the maximum shear is $\frac{1}{2}(Z-Y)_{max.}$ which comes from a direct application of Mohr's circle. For all reasonable values of Poisson's ratio ($\sigma > .25$) the shear stress $\frac{1}{2}(Z-Y)_{max.}$ is less than $\frac{1}{2}(Z-Y)_{max.}$ but as can be seen there is not a great difference between the two. The depth at which failure occurs is at the depth of maximum shear. To find this depth the expression $\frac{1}{2}(Z-Y)$ is differentiated with respect to distance below the gear tooth surface "z".

This process yields: $z_{\substack{max. \\ shear}} \sqrt{\frac{\sqrt{5}-1}{2}}\,(b) = .786\ b$

where once again $b = \frac{1}{2}$ width of contact area. The theory assumes that it is reasonable to suppose a fatigue strength in shear of approximately $0.3S_{f_{compr.}}$ From Figure 2 we see than $\frac{1}{2}(Z-Y)_{max.} = .3Z_{max.}$ Hence when $Z_{max.} = S_{f_{compr.}}$ a local shear failure occurs, and a pit is formed on the gear tooth.

Geometrical Description of Failure

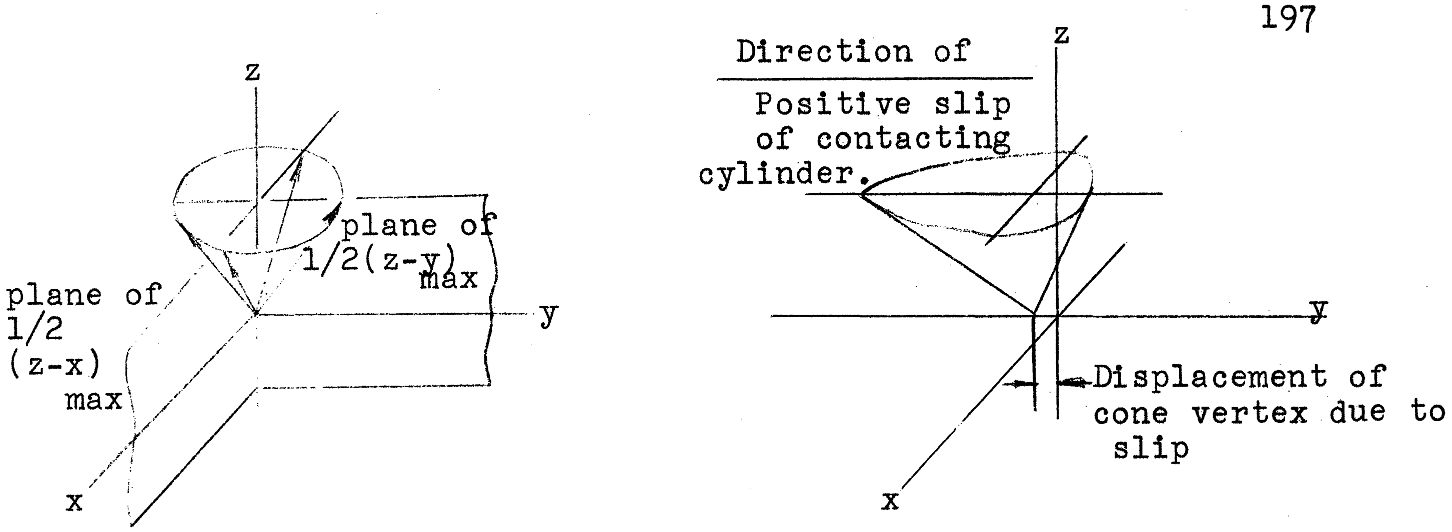
By application of Mohr's circle, it is known that the maximum shear occurs on planes inclined 45^{o} from the planes of the principal stresses. With X, Y, and Z the principal stresses, there are 3 sets of maximum shear stresses which occur:

$$\frac{1}{2}(Z - X)_{max.} \qquad \frac{1}{2}(Z - Y)_{max.} \qquad \frac{1}{2}(Y - X)_{max.}$$

From Figure 2 it can be seen that $\frac{1}{2}(Y-X)$ is small with respect to the other two and can be neglected. At the depth of maximum shear the $\frac{1}{2}(Z-Y)$ and $\frac{1}{2}(Z-X)$ shear stresses are not much different in magnitude. Hence, as is shown in Figure 3, approximately uniform shear stresses exist over a cone of 45° half angle, and in view of the fact that the stress distribution will be disturbed as soon as a crack forms, it is not difficult to reconcile the form of the pits with the stress distribution initialing their formation. The pits are often roughly hemispherical in shape, though they are rather less in depth than their full radius.

Commonly, when there is slip, the surface outline of the pits is in the form of a fan, with the apex pointing in the direction of slip of the contacting surface. The reason for the pits to be fan shaped when formed under conditions of slip is quite readily explained. When slip occurs, the frictional force which acts parallel to the surface adds a horizontal shear component in the direction of slip and thereby tilts and deforms the shear stress cone described above. The situation is shown pictorially in Figure 3.

It should be noted that for rolling cylinders with positive slip (contacting cylinder overriding at a higher speed) the direction of slip of the contacting cylinder is opposite to its direction of approach to a fixed point on the base cylinder. With this thought in mind a correlation can be made between the deductions of Timoshenko and Way with those of Beeching and Nicholls. Timoshenko and Way say that pits always have their apex pointing in the direction of the approaching cylinder, and



Pitting Cone
No - Slip

Pitting Cone
Slip Occurring

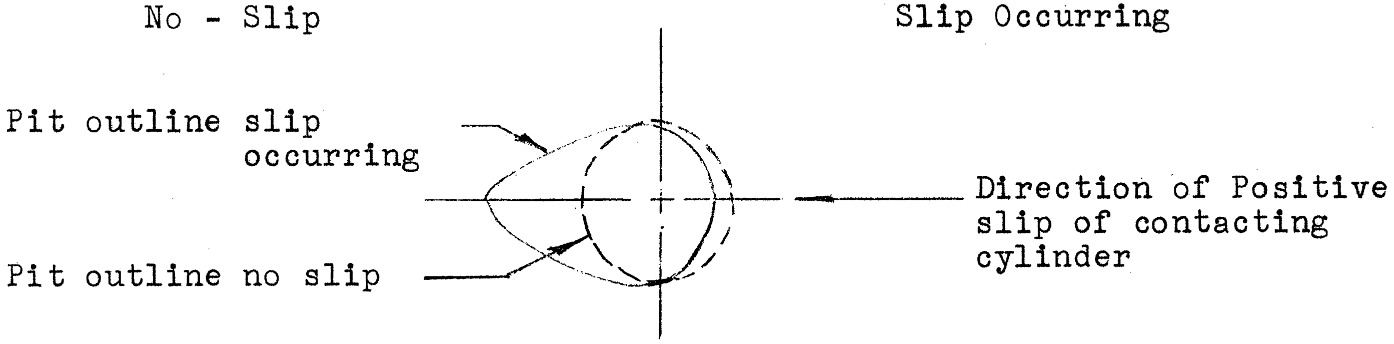
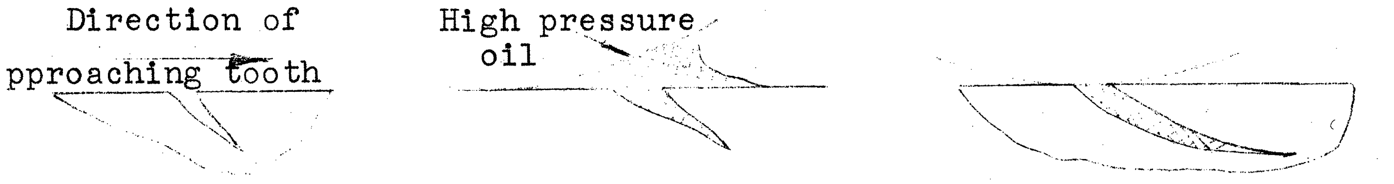


Figure 3

Beeching - Nicholls Shear Stress Cone



Initial Crack

Initial Entry of High
Pressure Oil into Crack

Trapped High Pressure
Oil Causing Further
Crack Propasation and
Pitting.

Figure 4

Magnified Picture of Steps in Pitting Crack Prepagation
(Way's Theory)

by utilizing the above sign convention between slip and direction of cylinder approach, the conclusions are equivalent with those of Beeching and Nicholls. Under similar slip conditions this discussion can be extended to describe the geometry of pitting failures in gear teeth. In all known gear pitting failures where a fan pattern was observable, the apex of the fan pointed in the direction of the approaching tooth. In many actual pitting situations fan shaped failures are not observable but rather irregular pits are formed. Under such conditions we might hypothesize that multiple pits varying in shape from the near hemispherical for non-slip conditions to fan shape under conditions of slip occur adjacent to one another on a gear tooth and destroy any regular geometric configuration of the pit. Such a situation seems feasible because slip and load conditions do vary along the contact length of gear teeth. Most pits are observed near the pitch line which is intuitive because in this area the maximum combination of normal load and sliding exists.

Reasons For a High Strength Case

Many components subject to high contact stress, particulary highly stressed gears such as those for automobile drives, are produced from case hardened steel, or, in some instances, from direct-hardened steels with a thin cyanide-hardened surface skin.

It is clear that to prevent pitting failure of the case due to the collapse of the softer core material, the case must be thick enough to ensure that the shear stresses are reduced within the case to a value below the fatigue limit of the core for fluctuating shear stresses with a zero minimum. A discussion of the actual case depths required which are arrived at through use of Hertzian shear stress profiles and accumulated data is more a topic of "Prevention of Pitting in Gears" and as such will be discussed later in the chapter.

Strengths and Weaknesses of the Beeching-Nicholls Theory

The Beeching-Nicholls theory of pitting is rigorously derived, utilizing precise mathematical application of the elastic properties of metals under load. As such, Beeching and Nicholls were not hampered by the perplexing variances in conditions under which pits occur in actual practice. They did not try to correlate a workable theory from emperical observations, but rather from unresticted pure theory.

In comparison with actual pitting failures, the theory describes pits which closely approximate those found in practice. The loads at which actual pitting occurs (near tensile fatigue strength of metal) are also in fair accordance with theory. However, being a theoretical derivation, it does neglect some practical conditions. Such factors as residual tensile stresses existing at the surface due to grinding operations would alter the Hertzian stress picture, as also would residual compressive stresses existing in a heat treated case, and also the balancing core tensile stresses. In addition compressive stresses due to cold working and deformation of the asperities would also change the theoretical picture which does not consider such factors.

2. SURFACE FAILURE THEORIES

General

Theories which state that pitting failures start on the surface of gear teeth are based on observations of small cracks appearing on the surface (Figure 4) and then growing until a piece of metal dislodges and a pit is formed (Figures 5 and 6). Postulations were made as to why the cracks propagated and then the postulations were varified by experiment.

Even among the supporters of pitting failure starting by cracks at the surface, there is confusion as to a precise mechanism by which the pitting cracks start. However, there is general agreement as to crack propagation by hydrodynamic wedging (Figure 4) which will be covered below.

In discussing the various concepts involved in pitting by surface crack propagation, it is felt that the most benefit can be gained from presenting the ideas of the various supporters of this theory and then summarizing the results.

The Findings of Stewart Way

General Listing

1. Profile sections of test specimens showing cracks in various stages of development indicate that cracks start at the surface.
2. Pitting cracks appear to be fatigue cracks. As a rule 500,000 or more cycles are necessary to produce a crack, and for normal pressure ($P_{max.}$) below a certain value pitting cannot be produced. Ten million cycles was adopted as the duration of a test if no pitting took place.
3. Pitting strength increases with hardness (See Figure 7).

4. Pitting strength can be increased by polishing the surface. This increase may amount to 100% in the P_{max} value for pitting if the polishing is carefully done.
5. Pitting cannot be produced without oil. When rollers are tested dry there is flaking and the formation of iron oxides. The addition of oil to a dry test under way will produce pitting in short order if the load is sufficiently high.
6. For a given material, finish, and load there is a critical oil viscosity such that oils of lower viscosity at the operating temperature will promote pitting and oils of higher viscosity will prohibit pitting.

Shape of Pitting Cracks and Pits

Way also states that the pitting cracks first appear at the surface as a small "V" which then gradually grow to a large V ready to be dislodged. When the particle is dislodged it is seen that the orientation of the cracks is such that the vertex of the V corresponds to the vertex of the fan shaped pit.

In all cases Way found that the pitting cracks are inclined to the surface at an actue angle, and that they dip in the direction of motion of the opposite roller (Figure 6). The shape of the pits described by Way compare favorably with those predicted by Beeching and Nicholls.

Stresses Which Cause Failure

In discussing the stresses which cause pitting failure, Way states that he does not believe that the Hertz theory is applicable to gears because actual cracks first appear minutely close to the surface, if not immediately on the surface, but certainly not at the depth of maximum shear stress. In any case the first minute cracks immediately propagates to the surface.

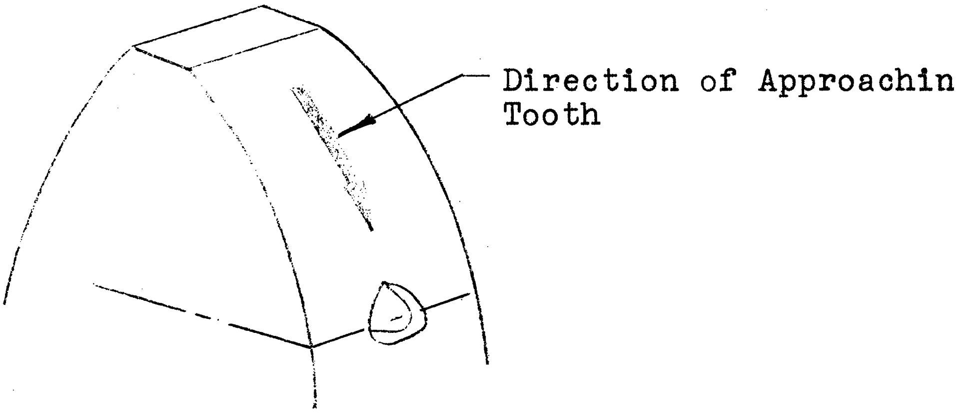


Figure 5

Possible Orientation of a Pit on a Gear Tooth

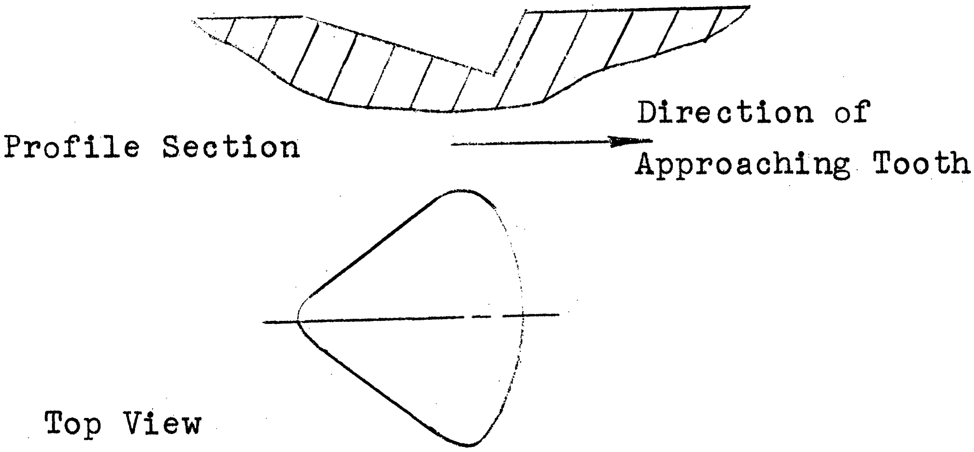


Figure 6

Views of Typical Pit

Also pitting cracks start with loads lower than the values thought necessary for fatigue failure by the Hertz theory. The Hertz theory, expanded by Beeching and Nicholls, utilizes smooth cylinders in the analysis. In the actual situation, the smooth cylinders are not smooth but covered with tiny ridges and valleys (asperities). These irregularities on the surface is probably the cause of the pitting cracks. As has been mentioned above, tests on polished rollers show that surface roughness is an important factor.

Effects of Lubrication

The growth of the pitting cracks to pits appears to be connected with the action of the oil. The generally accepted theory is that the oil enters a crack, is trapped by the oncoming roller which closes the entrance of the crack and is then put under high pressure as the roller above advances. This trapped oil under high pressure may well act as a wedge to drive the crack deeper. This theory is consistent with the observations of the inclination of the cracks and the critical oil viscosity. Probably a very heavy oil would not readily enter a crack and form a wedge.

In order to avoid restating material that has already been discussed, the finding of the other advocates of "surface failure in pitting" will not be presented in their entirety but rather their additional findings and points of contradiction with Stewart Way will be discussed.

The Findings of J. O. Almen

General

Almen agrees with Way insofar as the shape of the pits, propagation by an oil wedge, and effect of hardness. However, he has his own unique explanation of the stresses which cause failure. Almen stands pretty much alone as a supporter of his theory of tensile failure in pitting, but he obviously has uncovered some important factors which have been neglected by others. He also has an interesting discussion on the various stages in pitting.

Stresses That Cause Failure

The fundamental axiom of J. O. Almen is that fatigue failures are tensile failures. He considers the surface of a gear tooth as a sandwich which is being squeezed. The compressive normal loading causing elastic and plastic deformation of the metal. This deformation elongates the metal in a plane parallel to the surface; such elongations, being repeated with each passage of a tooth, cause fatigue fractures due to cohesive failures from tensile stresses induced in the plane parallel to the surface by the compressive load. The cohesive failures occur in a plane perpendicular to the surface and form cracks which intersect the surface. The cracks propagate and form pits when the gear teeth are lubricated and a hydrodynamic wedge is formed. Almen states that the magnitudes of the elongation stresses are not known except that they exceed the elastic limit of the material.

Tensile Residual Stresses

The existence of the tensile elongation stresses are further verified by the following tests results. Stress measurements by dissection disclose residual stresses that could have resulted only from yielding under the applied loads. Since the residual stresses are compressive, the yielding must have resulted from plastic elongation of the compressively loaded metal. In relation to the greater perpendicular compressive stress the yielding may be said to have occurred in tension. The induced "tensile" stresses do not obtain their maximum depth or intensity upon one load application but they continue to grow for a long time even though the applied load does not vary in frequency or magnitude.

Figure 8 shows the extremely high but shallow residual tensile stress which is induced in the surface of gear teeth by grinding operations. This high stress causes grinding cracks near the surface which are normally harmless because the high compressive stress in the bottom portion of the hardened casing does not allow these cracks to propagate. However, if the gears are lubricated, high pressure oil could form a dynamic wedge and cause crack propagation and eventually pitting.

Almens Theory as Applied to Carburized Gears

It is known that carburized gear teeth with a sufficiently thick case resist pitting quite well. The reason is that a carburized, or heat hardened case in general, has residual compressive stresses, a component of which acts in a direction parallel to the tooth surface. This compressive stress resists the induced tensile stress parallel to the surface caused by the normal loading stress, and hence holds the tensile stress to below its critical value necessary to cause failure.

Using the same reasoning it can be seen why shot peening does not reduce the probability of pitting. When a surface is shot-peened, a normal compressive stress is induced which hardens the surface but at the same time this normal compressive stress induces a tensile stress parallel to the surface and hence even makes pitting easier to accomplish.

The Findings of Timoshenko

Timoshenko's conclusions as to the cause and nature of pitting are in complete accord with those of Stewart Way. Any discussion of Timoshenko's analysis would be purely a reiteration of that of Way. One conclusion might be added which further varifies the necessity of oil for pit formation. Timoshenko found that pitting cracks may be stopped by either changing the direction of rotation or reducing the oil pressure in the crack.

The Findings of Buckingham

Buckingham has a new slant on explaining the Hertz stress picture and also on what the cause of progressive (incipient) pitting is. He states that the Hertz triaxial compressive stresses are caused by the reduction of surface area when the crowned surface of a gear tooth is transformed to a plane area of contact under load contact of mating teeth. This seems quite feasible. Buckingham also infers that failure takes place by subsurface shearing in which case he is in accord with Hertz, Beeching and Nicholls. He states that incipient pitting is probably caused by shearing out of the weaker or unsound particles of the material. When all of these have been removed, this pitting ceases unless the dynamic loads create stresses beyond the fatigue limit of the sound material. One should note the "probably" in Mr. Buckingham's last statement, and consider it as representative of the uncertainty as to the exact cause of pitting, even among experts. As with Beeching and Nicholls, Buckingham makes no mention of the necessity of oil to gear pitting.

Plastic Flow in Hard Gear Teeth

This phenomena exists to some extent in all gears under load. It may be physically observed in many rear axle hypoid gears of automobiles. The tooth surfaces have a rippled appearance similar to wind-blown sand. However, whether the plastic flow is or is not apparent, it always occurs and although the tooth form may sometimes be altered by the displacement of the metal, the effect is usually to prevent or postpone pitting fatigue because of the residual compressive stress that is induced. This compressive stress is caused by the plastic packing together effect of combined normal and sliding motion of the teeth relative to one another. The plastic flow takes place approximately parallel to the tooth surface as also does the compressive stress. Do not confuse this stress with the normal compressive loading stress.

Stages of Pitting

In standardized terminology the stage of wear that exists when the first pits appear is called initial pitting.

The stage when the pits are spreading is called progressive pitting.

Corrective pitting refers to the observed fact that the development of pits will cease for no apparent reason. There is, however, a definite reason for the pitting to stop under continuous running at constant load. This reason is directly connected to the above topic of plastic flow in gear teeth. The induced compressive stress caused by plastic flow acts parallel to the surface and increases slowly as plastic flow increases. As this stress increases it eventually reduces the tensile stress acting parallel to the surface, which is the cause of pitting, to below its critical value and hence pitting stops.

Buckingham mentions a final stage of pitting, destructive pitting, which occurs after corrective pitting only if the load on the gears is increased so that the tensile stress is again great enough to cause failure. Under such high load, complete disintegration of the tooth face can be expected.

3. DESIGN OF GEARS

Hertzian and Buckingham Equations-(Limiting Wear Loads)

About 1920, Professor Earle Buckingham of M.I.T. first used the Hertz stress equations for developing his wear formula for gears.[11] Actually, it is surprising that no one had previously done this, since Dr. Hertz first published his equations in 1882.

The equations developed in the first two parts of chapter 5 can be used to determine the limiting wear load for gears. In the following sections the equations for design of several types of gears as developed by Buckingham will be considered.

Limiting Wear Load for Straight Spur Gears

The contact conditions between spur gear-tooth profiles are similar to those between two cylinders, except that on gear-tooth profiles the radius of curvature is constantly changing. In many cases, wear on gear teeth first becomes apparent at or near the pitch line. Hence, we can select the radius of curvature of the gear-tooth profile at the pitch line as the one to use.

For two cylinders in contact, with the axes parallel, we have Equation 15 from chapter 5 giving the magnitude of the maximum pressure. As it is seen from Figure 9, chapter 5 maximum

$$P_1 = .564 \frac{P(1/R_1 + 1/R_2}{a \Delta}$$

pressure is equal to maximum compressive stress at the surface. Calling this stress (s) and rearranging Equation 15 we have

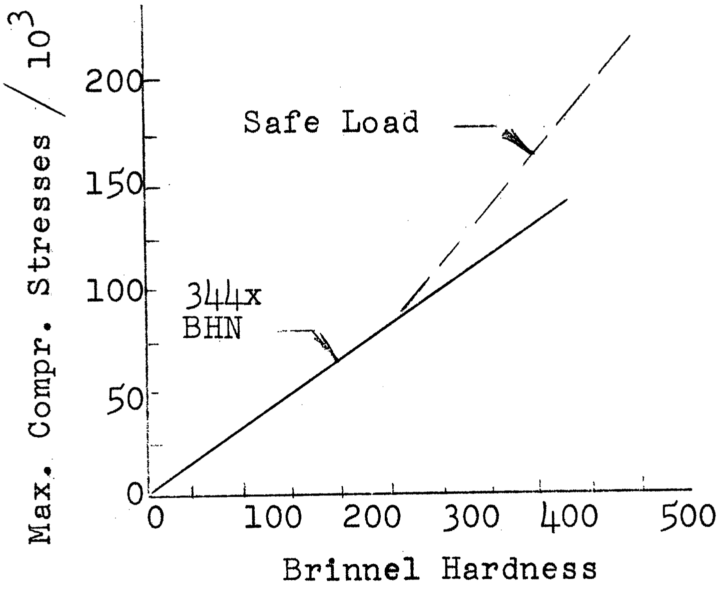


Figure 7

The Load Hardness Relationship For Pitting

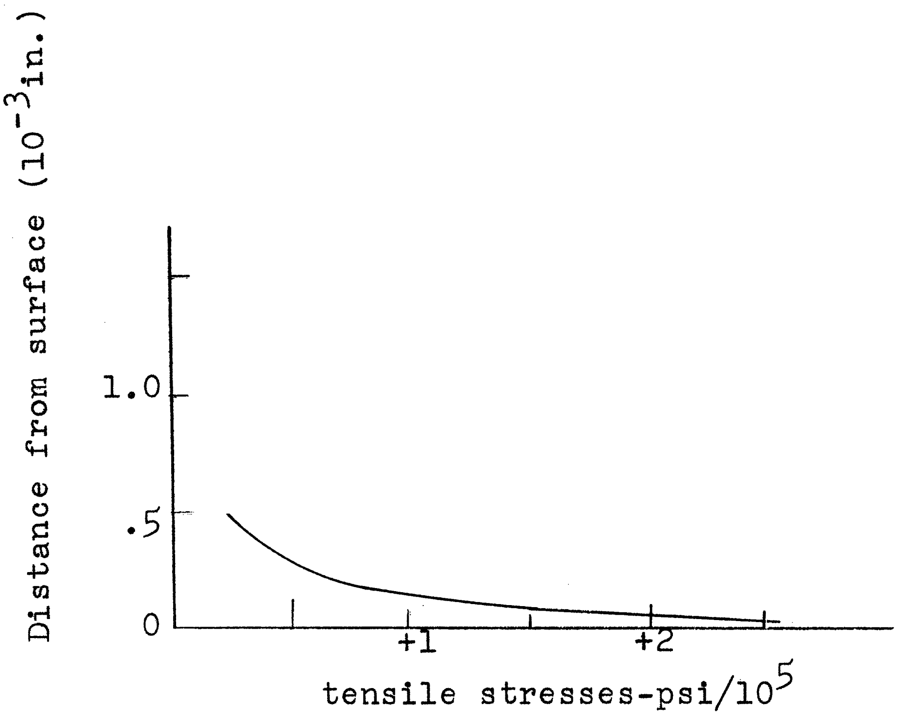


Figure 8

Residual Stress in a High Carbon Steel Speciman Caused by Grinding for limiting wear load for two cylinders in contact

$$F_w = \frac{s^2 b \; \frac{1-\mu_1^2}{E_1} + \frac{1-\mu_2^2}{E_2}}{0.318 \quad (1/R_1 + 1/R_2)} \qquad 1$$

Where:

F_w = Limiting wear load, lb.

s = Maximum compressive stress, psi

R_1, R_2 = Radii of contacting cylinders, in.

μ_1, μ_2 = Poisson's ratio

E_1, E_2 = Modulus of elasticity of materials, psi.

b = Length of the cylinders, in.

When:

ϕ = Pressure angle of the gears

D_p = Pitch diameter of the pinion, in.

D_g = Pitch diameter of the gear, in.

VR = Velocity Ratio = $\frac{D_g}{D_p}$

we have for the radii of curvatures at the pitch line

$$R_1 = \frac{D_p}{2} \sin \phi$$

$$R_2 = \frac{D_g}{2} \sin \phi$$

and

$$\frac{1}{R_1} + \frac{1}{R_2} = \frac{2}{\sin \phi} \left(\frac{1}{D_p} + \frac{1}{D_g} \right) = \frac{2}{\sin \phi} \left(\frac{D_p D_g}{D_p + D_g} \right)$$

Substituting into Equation 1 and rearranging we get

$$F_w = b \, \frac{D_p D_g}{D_p + D_g} \quad \frac{s^2 \sin \phi \; \frac{1-\mu_1^2}{E_1} + \frac{1-\mu_2^2}{E_2}}{0.636}$$

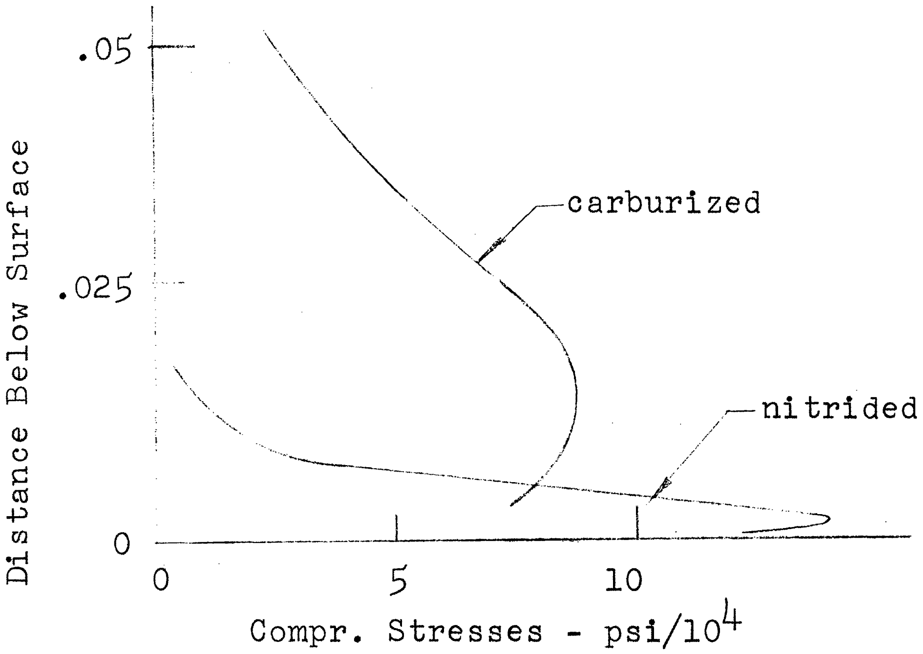


Figure 9

Residual Compressive Stress in Carburized Case of Gear
(tooth not ground)

If we let

$$Q = \text{Ratio factor} = \frac{2D_g}{D_p + D_g} = \frac{2VR}{1+VR}$$

Then

$$\frac{D_p D_g}{D_p + D_g} = \frac{D_p Q}{2}$$

hence

$$F_w = b\ D_p\ Q \left[\frac{s^2 \sin \emptyset \left(\frac{1-\mu_1^2}{E_1} + \frac{1-\mu_2^2}{E_2} \right)}{1.272} \right] \qquad 2$$

The term in the brackets is the load-stres factor for the materials. Denoting it with (B) Equation 2 for the limiting wear load for the straight spur gears becomes

$$F_w = b\ D_p\ Q\ B \qquad 3$$

Graphs of B vs. s for steel, Figure 10; Q vs. VR, Figure 11; and BHN vs. surface endurance strength for steel Figure 12; may be found in the appendix. Values of B for materials other than steel may be found in any number of books on gear design, and there is a table of values of p. 401 of Reference 9. Also, the data given in the appendix of Reference 10 may be used, if K_1 is converted to B by the formula:

$$B = \frac{K_1 \sin \emptyset}{4}$$

Another way to use the load stress factors is through the following formula

$$K_1 = W\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{F_d}{b}\left(\frac{2}{\sin \emptyset}\right)\left(\frac{1}{D_g} + \frac{1}{D_p}\right)$$

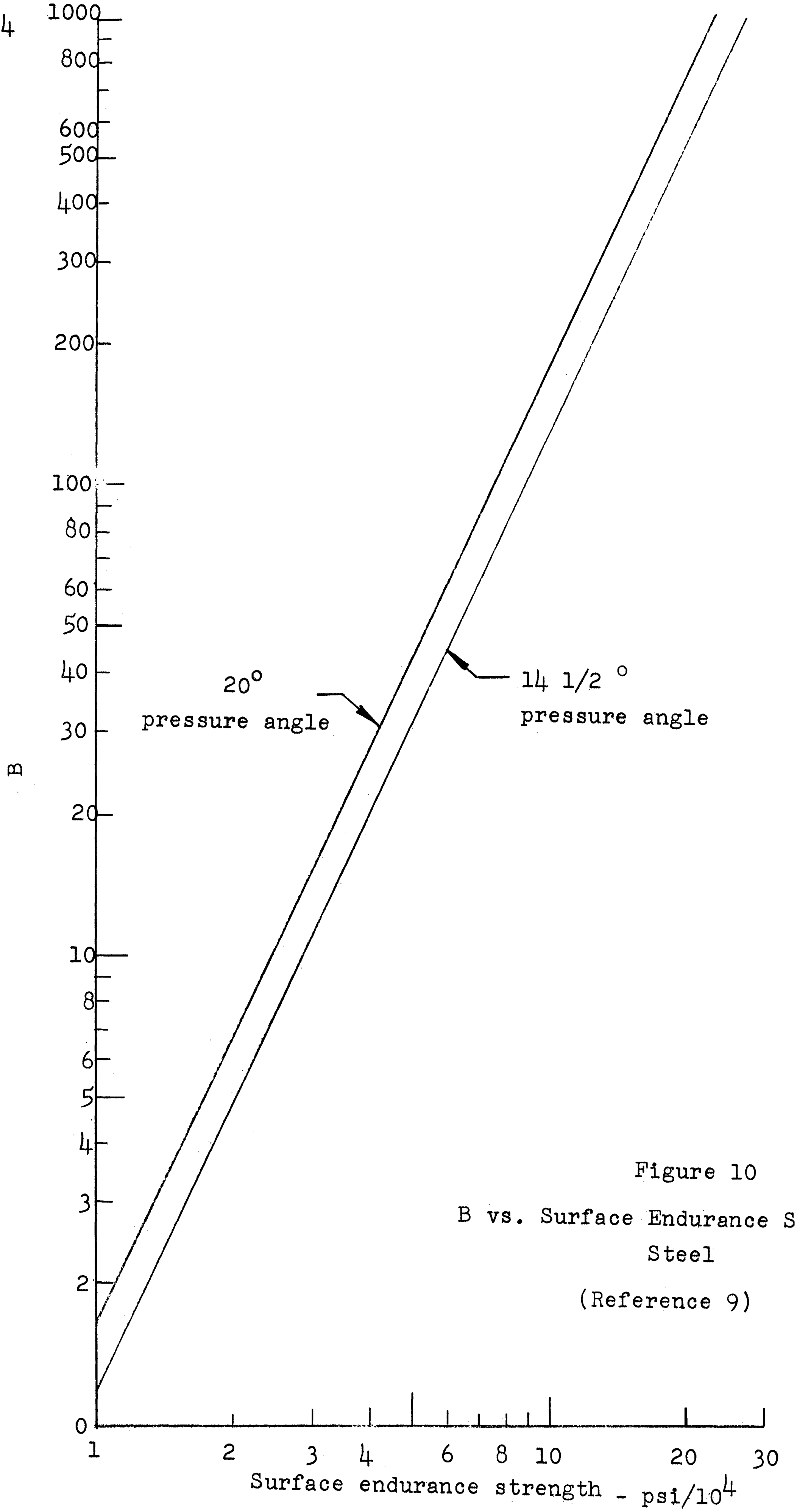


Figure 10

B vs. Surface Endurance Strength of Steel

(Reference 9)

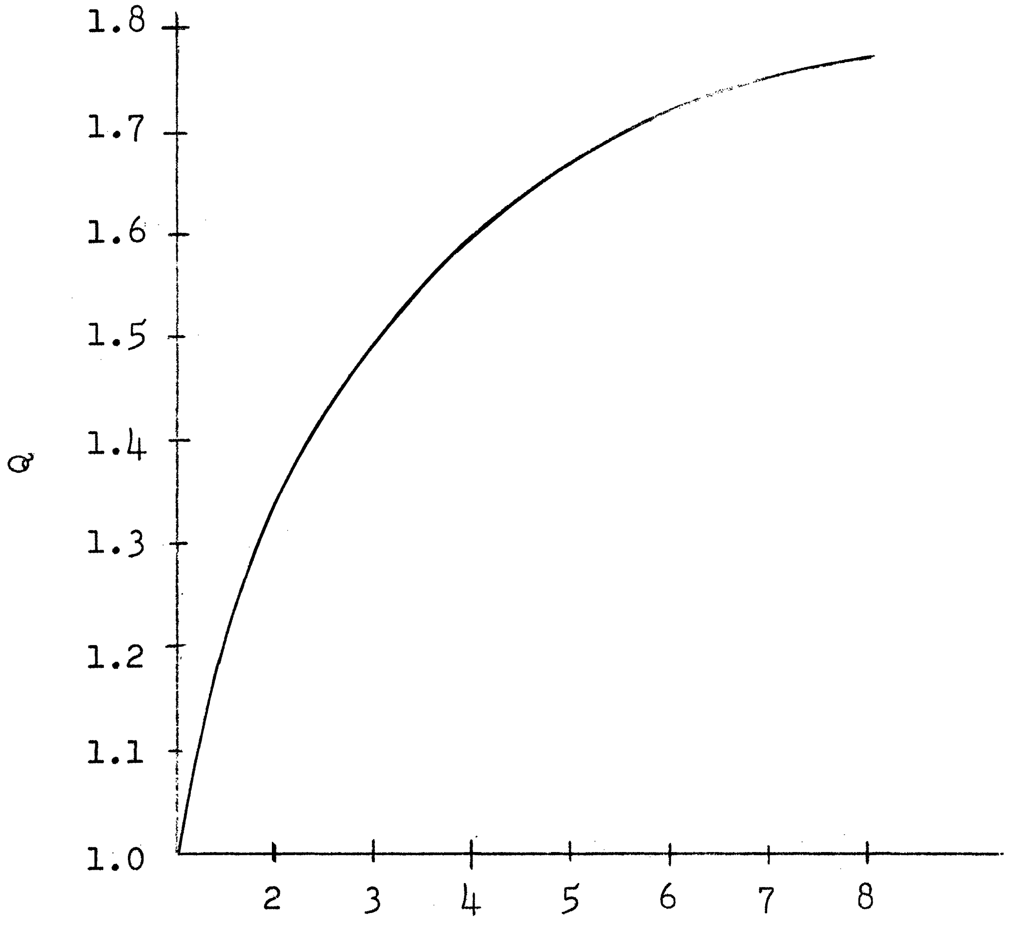


Figure 11

Q vs. Velocity Ratio[9]

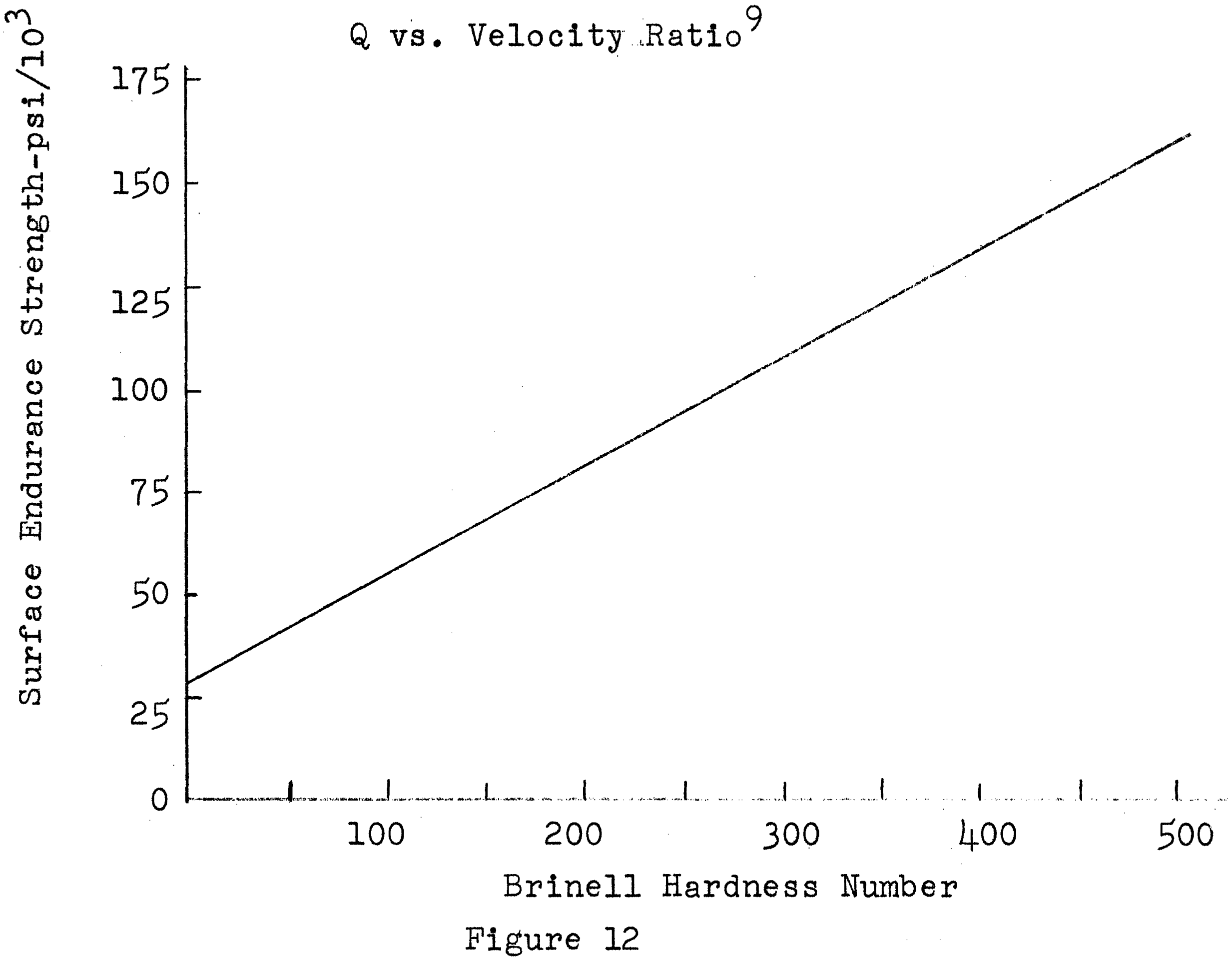


Figure 12

Steel Surface Endurance Strength vs. BHN[9]

A more direct method, and for that reason perhaps a preferable one, is to determine the Hertzian stress directly and compare it with the surface endurance strength of the material. The Hertzian equation, as applied to gears, becomes:

$$s = \sqrt{\frac{\frac{0.70\,F_d}{b \sin \phi}\left(\frac{1}{D_p} + \frac{1}{D_g}\right)}{\frac{1}{E_p} + \frac{1}{E_g}}}$$

For example, this value of stress may then be compared with the value of stress for a particular number of cycles on an S-N curve determined experimentally for a material; with values given in various tables listing surface endurance strengths of materials, or with the stress in gear designs which have proven satisfactory in service.

Limiting Wear Load for Internal Gears

The conditions for the internal gears are the same as those on spur gears except that the form of the internal-gear-tooth is concave instead of convex. Hence, the sign of radius of curvature is minus instead of plus, and Equation 19 applies.

$$P_o = .564\sqrt{\frac{P\,(1/R_1 - 1/R_2)}{a\,\Delta}}$$

Rearranging Equation 19 we have

$$F_w = \frac{s^2\,b\left[\frac{1-\mu_1^2}{E_1} + \frac{1-\mu_2^2}{E_2}\right]}{0.318\left(\frac{1}{R_1} - \frac{1}{R_2}\right)}$$

Using the same reasoning as in straight spur gears and letting

$$Q = \text{Ratio factor} = \frac{2D_g}{D_g - D_p} = \frac{2VR}{1-VR}$$

we get for the limiting wear load

$$F_w = b\ D_p\ Q\ B \qquad 5$$

where the load-stress factor (B) is same as for spur gears.

Limiting Wear Load for Helical and Herringbone Gears

The conditions for helical and herringbone gears are the same as those on spur gears. But here the radius of curvature is different and must be taken as the radius of curvature at the normal plane. From the geometry, the radius of curvature for helical gears is

$$R = \frac{D}{2\cos^2\psi} \sin\phi_n \qquad 6$$

where:

D = Pitch diameter
ψ = Helix angle
ϕ_n = Pressure angle in the normal plane

Using Equation 6 as the radii of contacting cylinders in Equation 15 and 19 and using the same load-stress factor as for the spur gears we have

$$F_w = \frac{b\ D_p\ Q\ B}{\cos^2\psi} \qquad 7$$

where the pressure angle in the expression for the load-stress factor is the pressure angle in the normal plane and (Q) is defined as before for the external and internal gears.

Limiting Wear Load for Worm Gears

The contact on worm gear drives is line contact but changing form of the worm across the fact of the gear makes it impossible to derive a simple mathematical expression for the curvatures of the contacting surfaces. Therefore, for the load-stress factor B, usually experimental values are used.

According to Buckingham, the wear load capacity of the worm gear drive depends upon the diameter of the worm gear. Hence, if

F_w = Limiting load for wear, lb.
D_g = Pitch diameter of worm gear, in.
b = Effective face width of worm gear (Assumed length of the contact line), in.
B = Load-stress factor of the materials

then

$$F_w = D_g\ b\ B \qquad 8$$

where the values for (B) are given in the table below.

Worm	Gear	B
Hardened Steel	Cast Iron	50
Steel, 250 BHN	Phosphor bronze	60
Hardened Steel	Phosphor bronze	80
Hardened Steel	Chilled phosphor bronze	120
Hardened Steel	Antimony bronze	120
Cast Iron	Phosphor bronze	150
Suitable for the values of lead angles up to 10°.		

Procedure of Design

The usual procedure in designing gears is to first determine the transmitted load and then determine the dynamic load, which will be greater than the transmitted load, due to inaccuracies in the manufacture of the gears, the deflections of the mountings and shafts, the vibrations, and accelerations and decelerations of the gears. Then F_s (the strength of the tooth in bending) and F_w are computed from the standard design formulas and compared with F_d. Ideally, both F_s and F_w should be greater than F_d. The lesser of the two forces, F_s or F_w will determine which is the critical factor, strength in bending or wear resistance. Or, as mentioned previously, the Hertzian stress may be solved for directly using F_d, and compared with the surface endurance strength of the material.

The transmitted force is found from the equation:

$$F_t = \frac{33{,}000\ P}{V_m}$$

Where:

F_t = transmitted force in pounds

V_m = pitch-line velocity in ft./min.

P = input horsepower

The dynamic load may then be found from the following equation:

$$F_d = F_t + \frac{0.05\ V_m(bC + F_t)}{0.05\ V_m + (bC + F_t)^{1/2}} \text{ lbs.}$$

Where F_d = dynamic load in pounds

$$C = \frac{k}{\frac{1}{E_g} + \frac{1}{E_p}}$$

$k = 0.107e$ for $14\frac{1}{2}^\circ$ full depth teeth

$= 0.111e$ for 20° full depth teeth

$= 0.115e$ for 20° stub teeth

e = effective or composite tooth error (See

Figure 13 and 14 in the appendix for values.) A nonograph for solving the dynamic load equation appears on page 399 of Design of Machine Elements[9] by Faires.

When using the above formulas, it must be realized that they are based on several assumptions and approximations. The Buckingham equation in itself is valid only at the pitch line, and it assumes that no sliding exists between teeth. In essence, these assumptions introduce only a little error, since pitting usually starts at the pitchline. And,depending upon the accuracy of the manufacturing processes, the true rolling action of involute design is closely approximated by gear teeth.

AGMA Wear Equation

In addition to the Buckingham equation and the Hertzian equations previously mentioned, there is another formula used a great deal to estimate the durability of spur gears. This is the formula approved by the American Gear Manufacturers Association (AGMA):

$$HP = F_i K_r D_o C_f$$

where:

HP = maximum HP which gears will transmit without showing signs of wear

F_i = combined factor for face width and inbuilt factor

D_o = combined factor for pinion diameter, velocity factor, and speed

K_w = combined factor for material, tooth form and ratio

C_f = modification factor to correct for increased stress at the start of single tooth contact for various numbers of teeth.

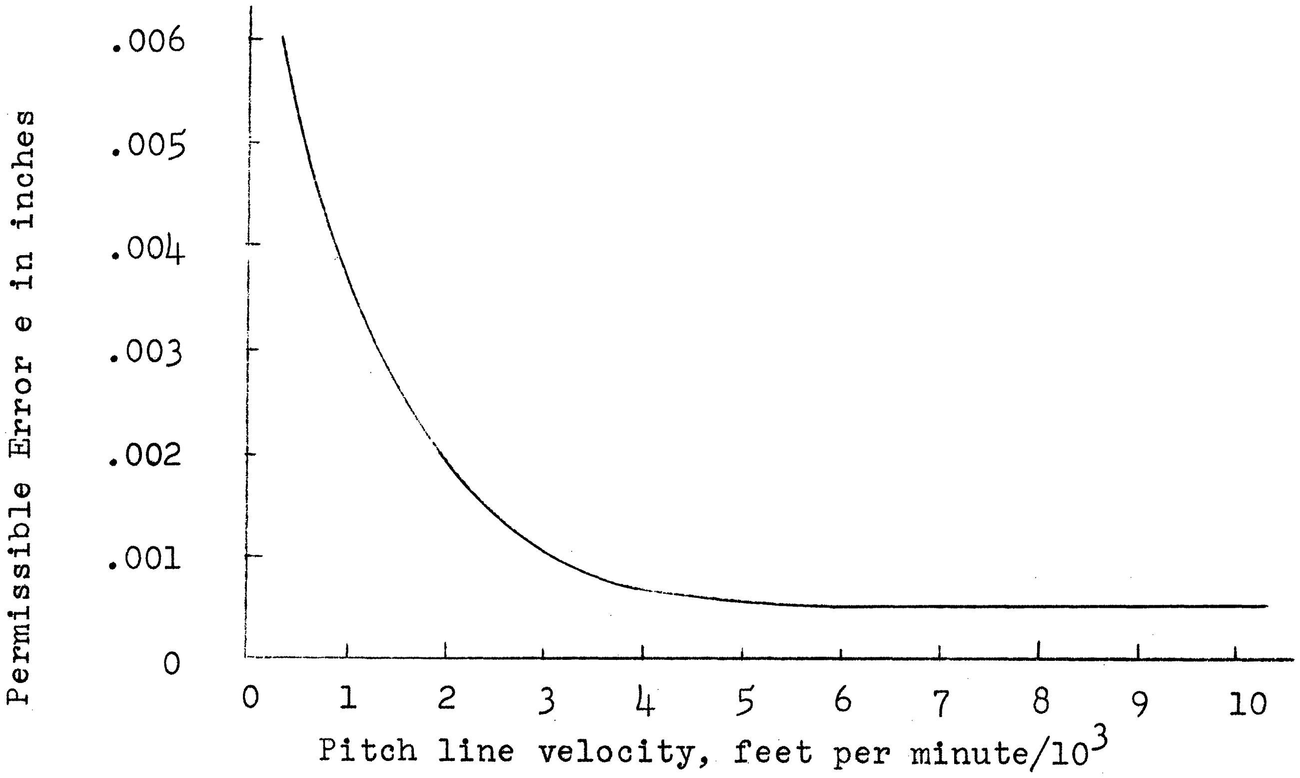


Figure 13

Permissible Tooth Error vs. Pitchline Velocity[9]

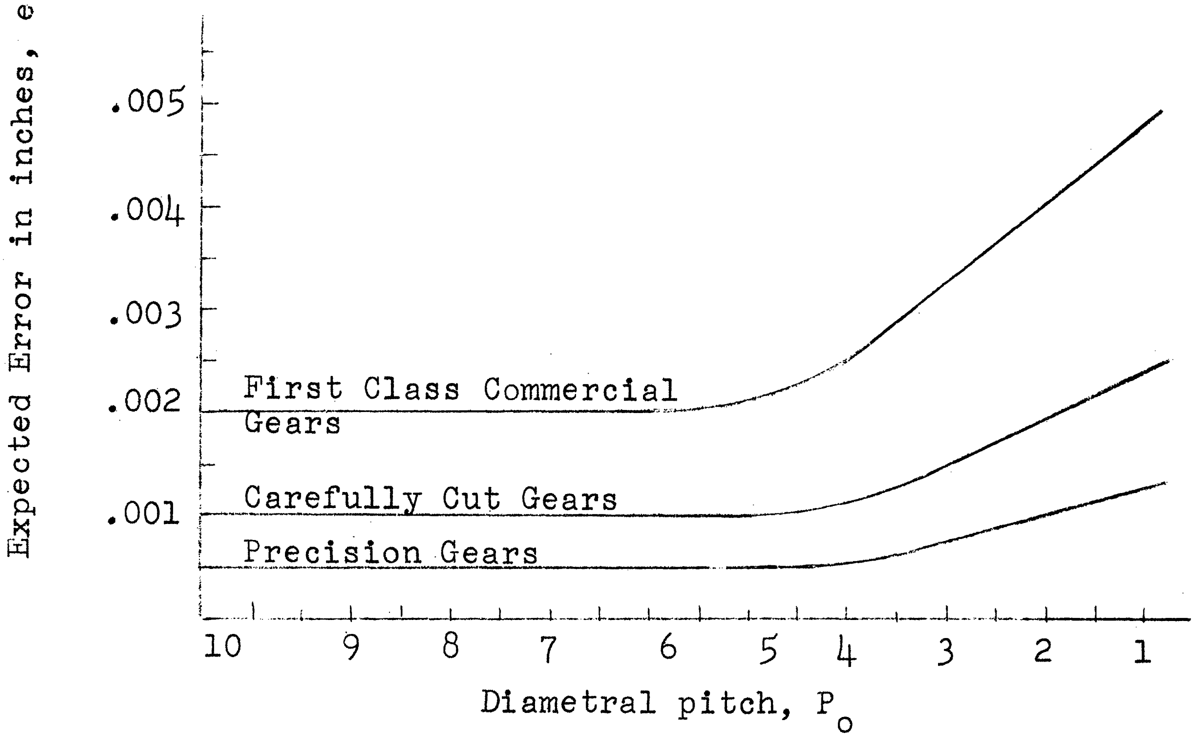


Figure 14

Expected Tooth Profile Error vs. Diametral Pitch[9]

An attempt was made to correlate the AGMA equation to the Hertzian and Buckingham equations, but due to the various factors involved there seemed to be no direct correlation. It appears as though the AGMA equation is one of empirical nature, based on experience, rather than on theoretical considerations. The main purpose of the AGMA equation is to evaluate the "wear H.P." of existing gears rather than to determine the design of a set of gears to transmit a specified HP, due to the fact that the interrelationships between the four constants are not clear, as are the parameters in the Hertzian and Buckingham equat ons.

4. METHODS OF REDUCING PITTING FAILURES

The methods of reducing pitting failures fall roughly into two different categories, and each has its limitations. First, the design of the gear can be altered. In poorly designed teeth, considerable improvement in the pitting strength can be made. However, in well designed gears, little if any improvement can be realized. Let us examine some of the changes that could be made.

Design Changes

One possible suggestion for a change in the design of gear teeth is to decrease the circular pitch of the gear. This results in more teeth on the gear, and so the load on the gear is shared by more teeth and the stress in each tooth is lessened. (See Figure 15) With more teeth on the gear, the problem of

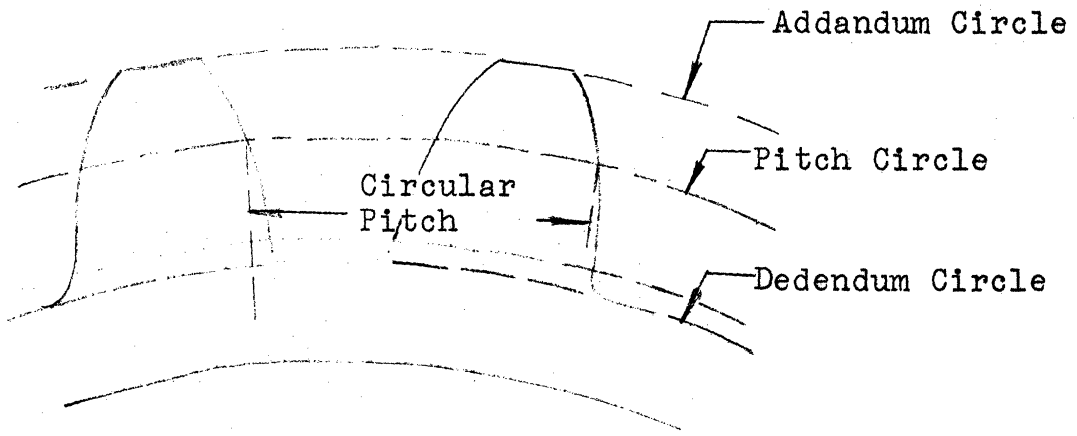


Figure 15

Gear Teeth

Figure 16

Effect of Increase in Pressure Angle

interference of gear teeth is posed. This can be lessened to a degree by increasing the pressure angle (See Figure 16), which results in a broader and stronger base section of the tooth.

Gears mounted on long shafts are particularly subject to elastic deformation, and variations in the mounting tolerances. A possible solution to this problem is to provide a taper or crown on the face of the tooth. (Figure 17) The taper has the disadvantage that it is designed for one particular load only, and other loads will result in misalignment with only a small portion of the tooth making contact. The crowned tooth will operate over a wider range of loading, but its surface of contact is reduced, and so stresses will be greater.

Metallurgical Change

Case Hardening

Some investigators have proposed the idea that pitting is mainly a surface phenomenon, but if this were true, casehardening to any appreciable depth would have little or no effect on the resistance to pitting. However, we know that casehardening does have a substantial effect on pitting resistance. In fact, this is one of the most readily accomplished methods of resisting pitting. The exact reason for casehardening improving pitting resistance is not known precisely, but it is certainly one of two reasons. It may be simply that the increase in hardness improves the fatigue strength of the metal. It may also be that casehardening places the surface, or case in comprensive residual stress, which is known to resist pitting.

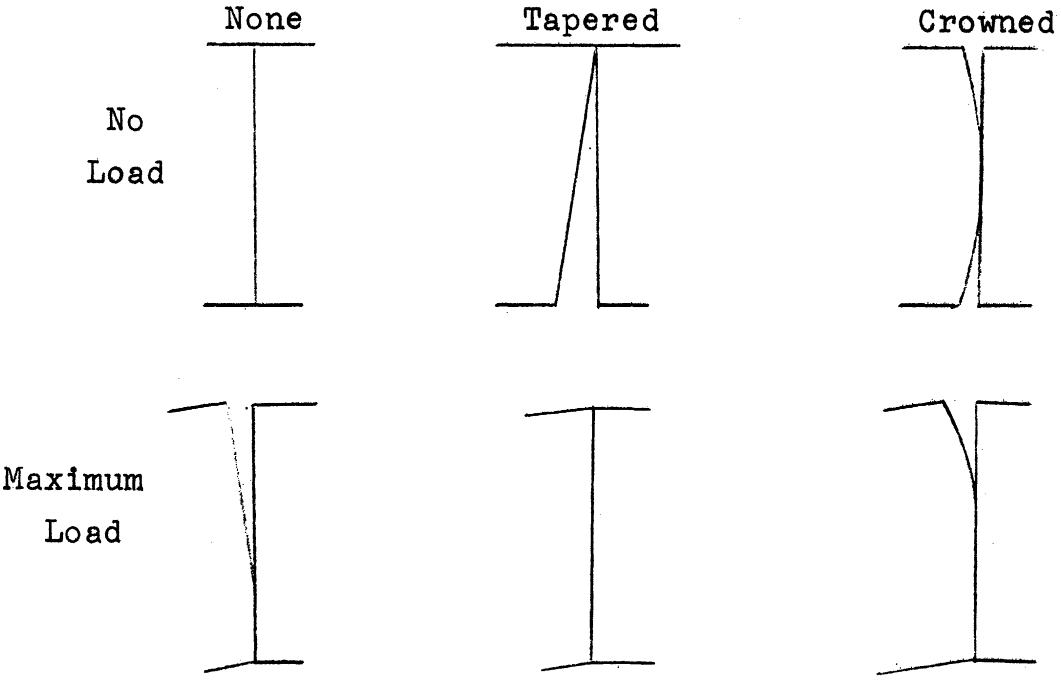


Figure 17

The Effect of Taper and Crown Modification to Overcome End Loads Caused by Misalignment[20]

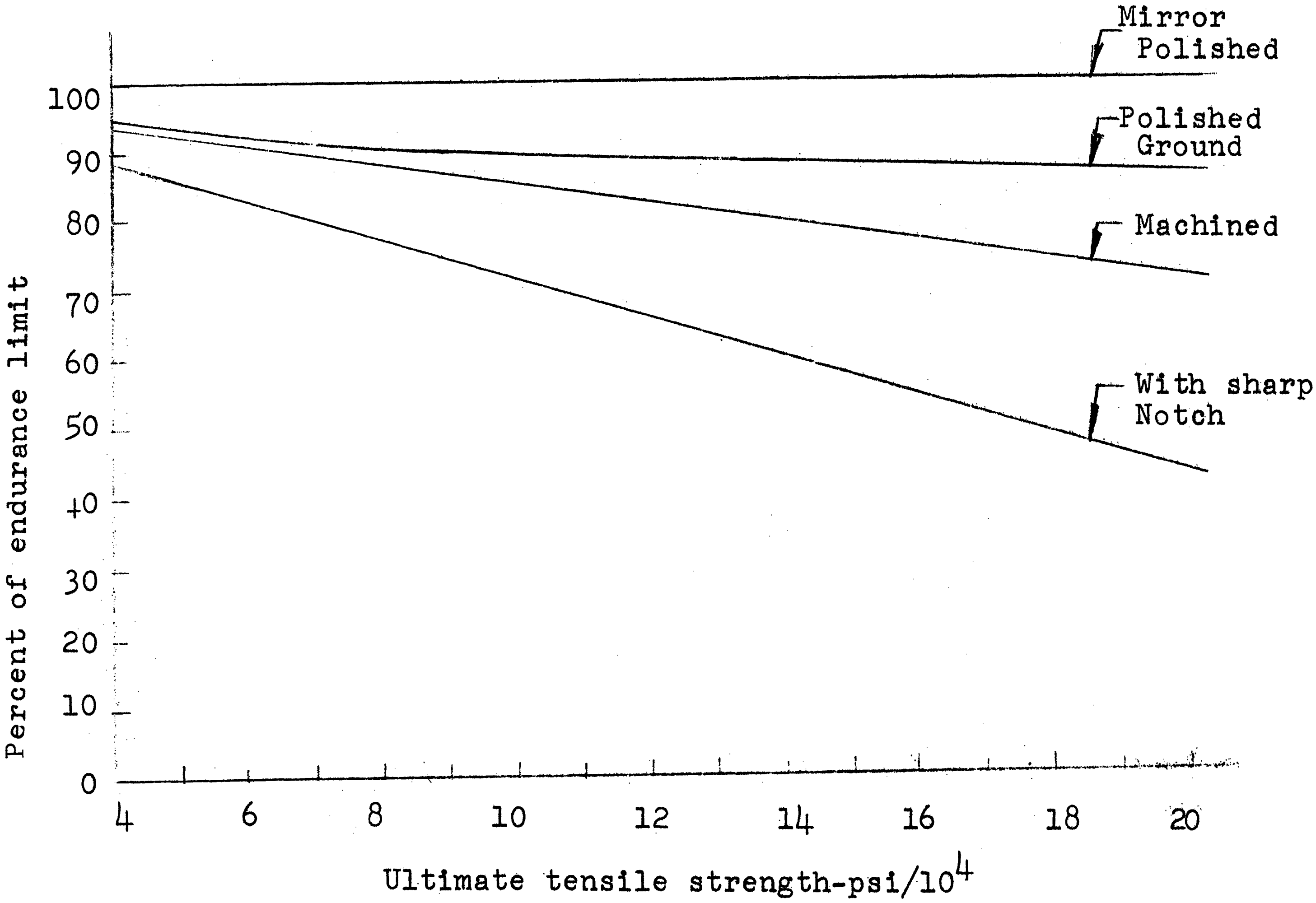


Figure 18

Endurance Limit Variance with Surface Finish[9]

Depth of Case

The depth of the case must be carefully controlled, since the maximum shear stresses occur beneath the surface, and if the case is not thick enough, the failure will start at the junction between the case and the core.[15] To determine the depth of the case, the Hertzian equation will be utilized. From Reference 10, half of the width of contact between two rolling cylinders may be found from the following equation:

$$d = \sqrt{\frac{4P\ (k_1 + k_2)\ R_1R_2}{R_1 + R_2}}$$

where:

P = the load per unit length of the surface of contact

$$k_1 = \frac{1 - \mu_1^2}{\pi E_1} \qquad k_2 = \frac{1 - \mu_2^2}{\pi E_2}$$

$2d$ = contact width

μ = Poisson's ratio

E = modulus of elasticity of cylinder materials

R_1 & R_2 = radii of the cylinders

Assuming that $\mu = 0.3$ and that the gears are made of the same material:

$$d = 1.08 \sqrt{\frac{F_d \sin \phi}{b\ E\left(\frac{1}{D_g} + \frac{1}{D_p}\right)}}$$

Now, the maximum shear stresses occur at 0.78 d,[10] and Buckingham recommends that the case depth be twice this depth. Thus the depth of case h will be 1.56 d:

$$h = 1.69 \sqrt{\frac{F_d \sin \phi}{b\ E \frac{1}{D_g} + \frac{1}{D_p}}}$$

However, since the magnitude of the shear stress may still possibly be greater than the strength of the core material, the actual magnitude of the stress should be computed at various depths, and the case should be deep enough so that the stress at the junction between the case and the core will be well below the strength of the core material.

Also, the effect of the residual stresses in the interior of the tooth must be taken into account when determining the case depth, since they must be added algebraically to the contact stresses in order to determine the true stress distribution. Unfortunately, the residual stresses are rather difficult to ascertain, and often they must be estimated, and the stress distribution will be different for various methods of casehardening.[14]

Any of the usual methods of casehardening may be employed, such as carburizing, cyaniding, nitriding, or flame hardening. Carburizing is preferable when a thick case is desired, and the latter three may be used satisfactorily when only a thin case is needed.[17]

Changes in Fatigue Strength

It is generally agreed that pitting strength of a metal is proportional to the fatigue limit. There are some arguments against use of a high fatigue strength. It is thought that with a more ductile metal, normal operation will wear off slight surface irregularities, and adjust for some misalignment, but design methods have been devised to compensate for a slight loss in ductility. It has become common practice to make the pinion quite hard, since it makes the most revolutions and suffers the greatest wear, and the gear itself is made somewhat softer to "run in" or compensate for slight irregularities. Thus the advantages of both high fatigue strength and ductility can be retained.

The fatigue strength of a metal is defined as being 0.5 times the tensile strength multiplied by a series of factors to account for such effects as size of the part, the surface finish of the part, and the type of loading of the part. Most of these factors are more or less uncontrollable, but we can change the surface finish of the gear teeth, and we can use a material of higher ultimate strength. It is usually more economical to use a lower alloy steel and then heat treat, so changing the ultimate strength by using a different material isn't often done. However, by improving the surface finish of the part, great changes in the endurance limit can be made. Figure 18 is a graph showing the effects of surface finish on the endurance limit. It is readily seen that with steels of high ultimate strength it is easy to increase the fatigue strength by 100%. However, any improvement in surface finish will naturally increase costs.

5. THE PV FACTOR

Throughout this discussion of pitting in gears, many theories have been advanced as to the causes of pitting, and many solutions suggested. Tests by General Motors have indicated that pitting may actually be caused by a number of methods, and possibly no one theory is totally correct.

The tests by General Motors have, however, suggested that there may be certain criteria for wear that are quite general in nature. As a result of test on 17 rear axel gear sets, it was discovered that the value of the Hertz compressive stress multiplied by the sliding velocity on the face of the tooth was a good indication of whether or not wear would occur. This value is called the PV factor, P standing for the Hertz compressive stress, and V for the sliding velocity. It was noticed that wear was greatly accelerated when the PV factor exceeded 1,500,000. The data for these tests is reproduced in Figure 19.

It was generally agreed that for more general applicability another factor must be included. The tests using the PV factor as the criteria for abnormal wear were quite satisfactory for automotive gears, but did not match the data for aircraft gears well. As a result, it was agreed to include a T factor - the length of action of the gear, which is the distance from the pitch point to the end of the action. The new factor was defined as the PVT factor. Its maximum value is still 1,500,000 for normal wear. This new factor was applied to various types of gears, and the correlation between the PVT value and normal wear was excellent. The above discussion was made by J.O. Almen. (See Reference 1)

Included in the book was a discussion by H. Blok of Holland. Mr. Blok criticizes the PV factor, in saying that other factors may fit the data just as well as does the PV factor, and other

Car No.	Scoring	Pinion RPM	Torque	Compress. Stress	Sliding Velocity	PV
1	None	24000	328	71,900	16.70	1,200,000
2		3450	112	73,400	18.42	1.355,000
3		3710	100	84,700	16.40	1,392,000
4		3720	96	94,700	14.75	1,402,000
5		3820	90	87,200	16.05	1,403,000
6		3850	87	87,600	16.20	1,420,000
7		2970	91	80,000	17.85	1,430,000
8		4060	92	88,200	17.05	1,505,000
9	occasional	3600	118	80,000	19.35	1,548,000
10		4000	95	86,600	17.80	1,551,000
11		3800	107	92,500	17.10	1,595,000
12	serious	3600	97	109,200	15.40	1,676,000
13		3800	109	108,000	16.60	1,800,000
14	bad	4266	88	79,300	23.25	1,848,000
15		3810	90	101,200	18.32	1,852,000
16		3800	109	95,800	20.05	1,932,000
17		2400	328	75,000	26.75	2,010,000

Figure 19

Rear Axle Gearsets

factors may be better suited for application to more types of gears. Blok suggests that a more critical criteria for failure is the instantaneous flash temperature occurring on the surface of the gear under operation. He states that for any given combination of gear materials and lubricant, there is a "characteristic scoring temperature" which is independant of the contact pressure and speed. Above this temperature failure will occur rapidly. Blok defines the temperature as;

$$\text{Temp} = Cf\ (P^3V)^{1/2}$$

where:

C is a constant determined by the geometry and the thermal properties of the gears.

f is the coefficient of tooth friction

The feeling of Mr. Blok is that his temperature criteria is applicable to a wider range of gears. However, his criteria does not fit the data for the General Motors tests any better than does the PVT criteria. In fact, his criteria does not seem to work as well with aircraft gears as does the PVT. So perhaps the PVT factor will be the better one since it is much easier to use.

BIBLIOGRAPHY

1. Almen, J. O., "Surface Deterioration of Gear Teeth", Mechanical Wear, (Symposium), ASM, 1950.
2. Almen, J. O., "Fatigue Failures are Tensile Failures", Product Engineering, March, 1951.
3. Beeching, R., and Nicholls, W., "A Theoretical Discussion of Pitting Failures in Gears", The Institution of Mechanical Engineers,1946.
4. Buckingham, E., Dynamic Loads on Gear Teeth, 1931.
5. Gross, M., "Laboratory Evaluation of Materials for Marine Propulsion Gears", ASTM Preprint, 1951.
6. Hashino, J., Some studies on the Pitting of Marine Reduction Gears, The Institution of Mechanical Engineers, 1956.
7. Timoshenko, S., Strength of Materials, Part II, pp. 505-509, 1956.
8. Way, S., "Gear Tooth Pitting", The Electric Journal, April 1936.
9. Faires, V. M., "Design of Machine Elements," The Macmillan Co., New York, 1955, pp. 381-461.
10. Cram, W. D., "Experimental Load-Stress Factors," Engineering Approach to Surface Damage, (Symposium), University of Michigan, Ann Arbor, Michigan, 1958.
11. Buckingham, Earle, "Spur Gears, Design, Operation and Production, McGraw Hill, New York, 1928, pp. 300-310.
12. American Gear Manufactures Assocation, "Surface Durability of Spur Gears," AGMA, Pub. 210.10, December, 1946.
13. Botstiber and Kingston, "High-Capacity Gearing," Machine Design, Vol. 25, No. 12, p. 203.
14. Buckingham, Earle, "Operational Stresses in Automotive Gears," SAE Quarterly Transactions,Vol. 5, January, 1951, pp. 43-55.
15. Buckingham, Earle, and Talbourdet, G. J., "Recent Roll Tests on Endurance Limits of Materials," Mechanical Wear (Symposium) ASM, 1950.
16. Clark,Donald S., and Varney, Wilbur R., "Physical Metallurgy for Engineers," Van Nostrand, Princeton, 1957, pp. 192-221.
17. Coleman, Wells, "An Improved Method for Estimating the Fatigue Life of Bevel Gears and Hypoid Gears," SAE preprint #627.
18. Dudley, D. W., "Modification of Gear Tooth Profiles," Product Engineering, September, 1949, pp. 126-131.
19. Durham, Harold M., "Designing Spur and Helical Gears for Durability," Machine Design, Vol. 23, No. 1, January, 1951, pp. 149-152.

20. Graham, J. D., "Pitting of Gear Teeth", Engineering Approach to Surface Damage (Symposium), University of Michigan, Ann Arbor, Michigan, 1958.

21. Huffaker, G. E., "Predicting Gear Pit Point," SAE Journal, October, 1959, pp. 82-83.

22. Medola, F. C., "How to Caluclate Truck Transmission Gear Life," SAE Journal, October, 1959, pp. 53-54.

23. Oberg, Erick, "Spur and Bevel Gearing," The Industrial Press, New York, 1915, pp. 65, 230.

24. Way, Stewart, "Pitting Due to Rolling Contact," Trans, ASME, Vol. 57, June, 1935.

25. Attia, A. Y., "Dynamic Loading of Spur Gear Teeth," Trans, ASME, Paper No. 58-SA-32.

26. Gross, Mr. R., "Laboratory Evaluation of Material for Marine Propulsion Gears," Preprint, ASME, 1951.

27. Kelley, B. W., "A New Look at the Scoring Phenomena of Gears," Preprint #804, SAE, 1952.

28. Lipson, Charles, "Analysis of Fracture and Pitting of Transmission Gears," 1949.

CHAPTER 11

GALVANIC CORROSION

1. CAUSES OF CORROSION

Under ordinary environmental conditions engineering metals do not combine spontaneously with oxygen or other common gases. The process of galvanic corrosion is much more complicated than a direct oxidation reaction. Before a detailed analysis of some of the more important fundamentals of corrosion is made, a simple laboratory corrosion experiment will be explained.

Corrosion Example

Assume that bars of iron and copper are partly immeresed in a tank of water and connected externally by a wire. (see Figure 1) Immediately a potential differency will exist between the two metals, a tiny electric current will begin to flow through the wire, and ionic reactions in the solution will start the electrochemical destruction of one of the metals. The basis for the corrosive attack lies in the ionic reactions which are occuring in the solution. For most chemical reactions water is not assumed to ionize appreciably, but because of the very slow rate of attack during corrosion, sufficient ions are present, even in plain water. Similarly, there is a slight tendency for ions of iron, Fe^{++}, and copper, Cu^{++}, to enter the solution (Figure 2).

The ferrous irons will tend to combine with the negative hydroxyl ions because, as reference to any chemical electromotive series will indicate, iron is above hydrogen and much above copper in activity. Thus the bar of iron will be attacked,

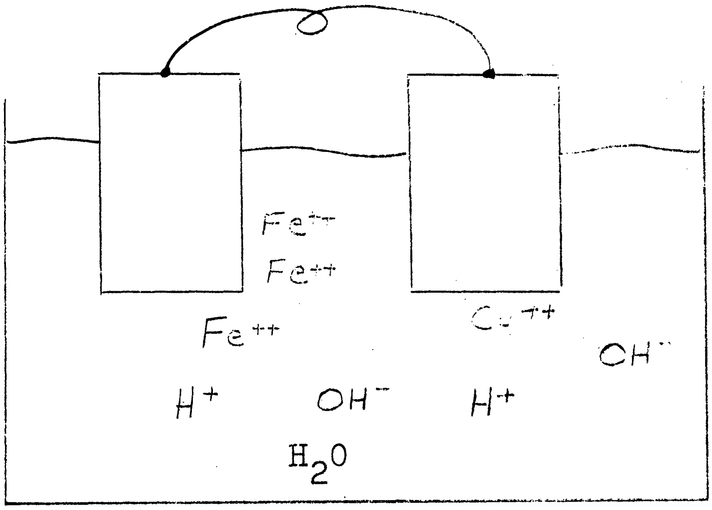


Figure 1

Ions present when iron and copper are connected electrically in water.

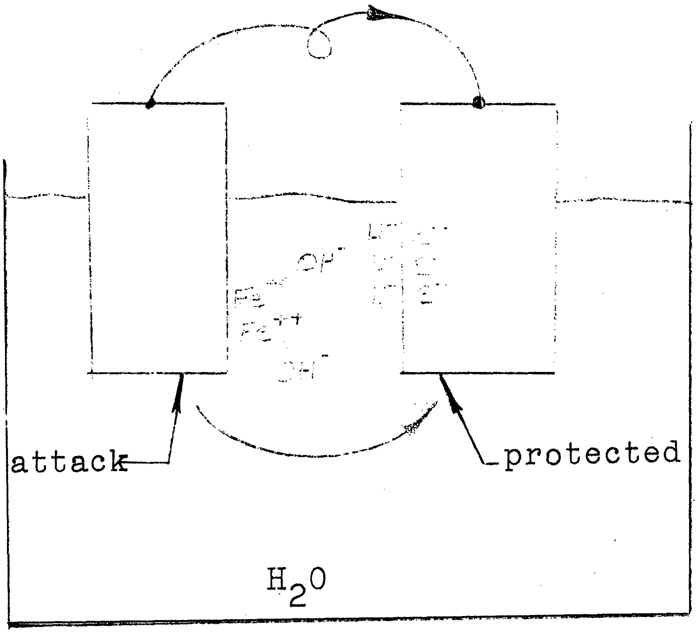


Figure 2

Subsequent corrosive attack of anodic region.

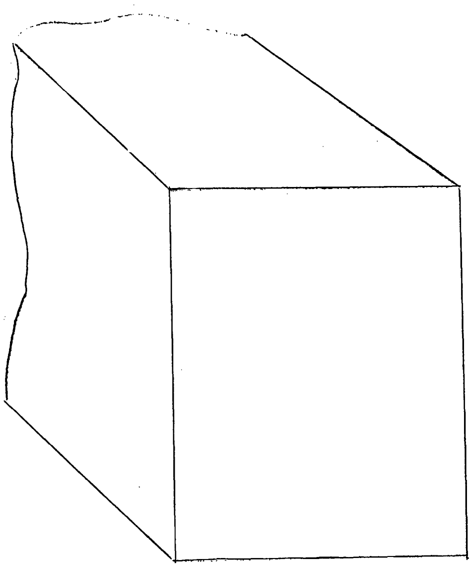


Figure 3

Tiny anodic and cathodic areas on the surface of a metal.

and $Fe(OH)_2$ and eventually $Fe(OH)_3$(iron rust) will be formed at or very near the anode (iron). Simultaneously, the electrons which remain when the iron atoms ionize are free to flow through the wire to the cathode (copper) where they collect in excess. Here the stray H^+ ions combine with the electrons to produce atoms of nascent hydrogen, thus completing the circuit. Assuming that sufficient potential exists to continue the reaction at a steady rate, the iron would soon show signs of galvanic corrosion while, relatively, the copper would be protected from corrosive attack.

It should be emphasized that actual corrosion phenomena are never as simple as the example cited because they rarely occur in the presence of pure water. Dissolved oxygen and other gases as well as countless extraneous ions of impurities enter into a variety of multiple combinations to speed, slow, or even change the direction of reaction. In addition, the problems of concentration gradients, acidic and basic solutions, current density, area effects, inhomogenieties, and the formation of surface films greatly complicate the theoretical treatment of corrosion. These phenomena will be discussed and related later in the chapter.

At present, however, the following generalities can be drawn from the example to be applied to all corrosion reactions: For galvanic corrosion to occur an electric potential must exist between metals across an electrolyte, which is an ion containing medium. Often as just shown, this voltage is the inherent potential difference which occurs between dissimilar metal pairs. Equally as often, however, ionic concentration differences, stresses, and other factors cause potential drops within different areas of the same metal.

Anode and Cathode

Either way, two or more distinct electrodes of different potential must exist, even if on a nearly microscopic scale. The electrode of higher potential is labeled the anode; the other, the cathode. The anode is the region at which metal atoms become ions in solution and form insoluble compounds of the metal. During this process the oxidation of the metal from a lower to a higher valence state occurs. At the cathode positively charged ions of electrolyte are reduced by combination with excess electrons which have flowed through the circuit from the anode.

Expressed more concisely as a rule-of-thumb, the preceding discussion states:

1. For galvanic corrosion to occur.
   a. Regions of different potential must exist in a metal or between metals.
   b. An electric circuit must be completed in the presence of an electrolyte, or ion-carrying medium.
2. The regions are labeled as follows:
   a. The anode is the area of higher potential and is the area where corrosive attack occurs.
   b. The cathode is the lower potential region. During this reaction it is protected from attack.

In the example, it was easy to distinguish the anode from the cathode but in many reactions the above explanation seems to fail. For instance, if iron is immersed in hydrochloric acid, corrosive attack appears to occur simultaneously with the evolution of hydrogen bubbles uniformly across the surface. Actually, variations in potential in the metal and electrolyte cause the existence of tiny anodes, where the attack occurs, in a checkerboard pattern with cathodes, where the hydrogen is evolved. (See Figure 3) These areas move and change so that uniform corrosion is the result. There is, obviously, no distinct line between galvanic corrosion and pure chemical reactions in many cases. This discussion, however, is limited to reactions of engineering metals in which the anodic and cathodic regions can be distinguished.

Electromotive Activity Series

Ions of different metals possess particular potential differences with respect to a standard datum charge level. For example, the familiar activity series of elementary chemistry is constructred with reference to the hydrogen ion. Since many corrosion reactions are based upon the potentials between dissimilar metals, the arrangement of engineering metals into practical tables of decreasing activity is of great importance. Generally, the ionization tendencies of the metals in a certain solution are measured with respect to a standard cell, and the metals then arranged so that the upper metal of any chosen pair is anodic to all metals lower on the list. A typical galvanic series of metals and alloys in sea water is listed for reference in Table 1. The use of such tables will not be dealt with in this paper, but the reader is reminded that in every case the upper metal of any pair will be anodic to the lower, and hence will tend to be attacked while the cathodic metal will be protected. Also, the rate of reaction of a particular pair is determined primarily by the relative distance between them on the chart, because this reflects the magnitude of their potential difference.

Magnesium
Zinc
Alclad 3S
Aluminum 3S
Aluminum 61S
Aluminum 63S
Aluminum 52
Low Steel
Alloy Steel
Cast Iron
Type 410 (Active)
Type 430 (Active)
Type 316 (Active)
Ni-Resist
Muntz Metal
Yellow Brass
Admiralty Brass
Aluminum Brass
Red Brass
Copper
Aluminum Bronze
Red Brass
Copper
Aluminum Bronze
Composition G Bronze
90/10 Copper-Nickel
70/30 Copper-Nickel-Low Iron
70/30 Copper-Nickel-High Iron
Nickel
Inconel
Silver
Type 410 (Passive)
Type 430 (Passive)
Type 316 (Passive)
Monel
Mastelloy C
Titanium

Active End

Table 1

Galvanic Series of Metal and Alloys in Sea Water

EMF Series	Sea Water Series	Air Series
Noble End	Noble End	Noble End
Gold	Platinum	Platinum
Platinum	Gold	Gold
Palladium	Silver	Silver
Silver		Mercury
Mercury	Titanium	Palladium
Copper	Copper	Copper
Hydrogen		Lead
Lead	Nickel	Nickel
Tin	Tin	Tin
Nickel	Lead	Hydrogen
Iron	Iron & Steel	Iron
Chromium		Zinc
		Sodium
		Chromium
	Aluminum	Titanium
Aluminum	Zinc	Aluminum
Beryllium		Beryllium
Magnesium	Magnesium	Magnesium
Sodium	Sodium	
Calcium		Calcium
Active End	Active End	Active End

Table 2

Environmental Galvanic Series

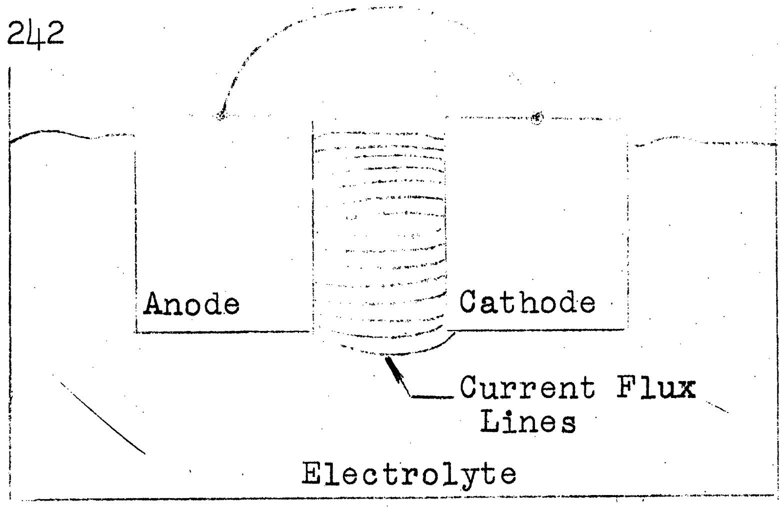
Area Effects

While the subject of anodes and cathodes is still fresh in mind, a mention of the effect of the relative exposed areas of the two electrodes will be made. It is apparent that if the anode and cathode are of nearly equal area, the density of current flowing between them through the electrolyte will be nearly uniform. However, if one electrode is much smaller than the other, the current will be much more highly concentrated near the smaller electrode (See Figure 4). Despite the fact that the current leaving the anode must always equal that entering the cathode, the higher density will cause accelerated reactions at the electrode of smaller area. This condition is usually not serious if the cathode is smaller, but when the anode has the small surface, highly accelerated attack can occur.

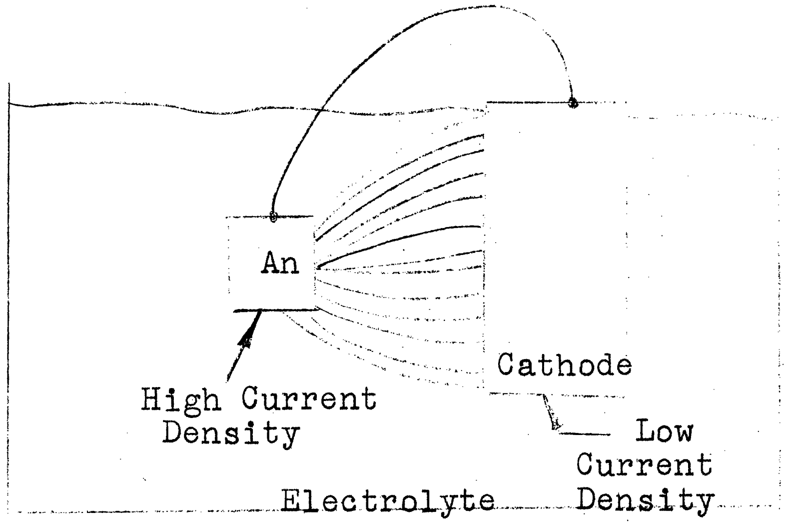
A common technique for the protection of metals is the attachment of small replaceable sacrificial anodes at strategic intervals with the sole purpose of causing high attack of the anode with the subsequent decline of corrosion on the large cathodic areas.

Electrolytic Influence

Until now the treatment of corrosion has dealt primarily with the variables controlled by the metals involved. The electrolyte is equally as important, however, and usually much more difficult to analyze. The impression may have been given thus far that all corrosion is essentially an under-water phenomenon, in which metals are constantly immersed in a liquid



Equal areas, equal current density



Unequal areas, accelerated anodic attack

Figure 4

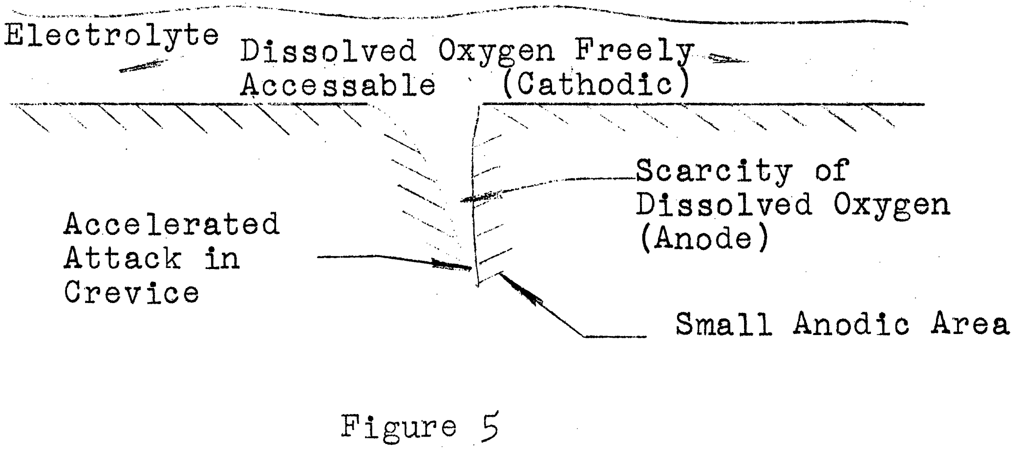


Figure 5
Crevice Corrosion

solution. This generalization cannot be made, although moisture is nearly always present. Corrosion often occurs under conditions of alternate wetting and drying, mists and sprays, and even such conditons as contact with moist earth. With this fact in mind, the analysis will continue with the same simplified assumptions as before.

Generally, the rate and type of reaction is controlled by the number and kind of ions present in the electrolyte. One should expect, therefore, that acids, bases, and dissolved salts would speed the rate of reaction and alter its character. A variety of corrosion products may be formed, and plating, gas formation, and other related activities occur. One must always be careful to consider the electrolyte in corrosion reactions, therefore, because it can have great effect.

Polarization

In the initial iron-copper corrosion example, it was implied that once started, the reaction would continue unchecked until the iron was totally rust. Actually, in a weak electrolyte such as water this would not necessarily occur. The reason lies in the nascent hydrogen formed at the cathode. The hydrogen atoms might combine with each other to form molecules, which in turn might collect in sufficient numbers to go to the surface as bubbles. A certain potential is needed to sustain this reaction, however, and if insufficient driving force exits in the circuit, the reaction might slow or stop completely. If the so-called "hydrogen over-potential" or back emf equals the driving potential, corrosion cannot continue. This is the prime example of a state known as polarization. Polarization may occur at either electrode for a variety of reasons, but this one, the cathodic polarization due to hydrogen, probably is the most common.

Dissolved Oxygen

One thing which defeats some of the benefits of polarization is the presence of dissolved oxygen, for oxygen combines with the molecular film of hydrogen coating the cathode, and destroying the polarizing effect. That is why water is often de-aerated for commercial purposes, so the benefit of cathodic polarization will not be destroyed.

Concentration gradients of oxygen are often responsible for corrosion in regions where no anodic or cathodic areas are apparent. Crevices, for example, or other areas not freely accessible to dissolved oxygen become anodic with respect to large exterior surfaces where oxygen is plentiful. (See Figure 5) . This condition is aggravated by the area effect, because the oxygen-starved anode is also a region of high current density. Similar reactions take place at areas where foreign matter such as lichens attack themselves to the surface, causing oxygen concentration.

Surface Films

If the products of corrosion are resilient enough to block further corrosion, they can be beneficial. In particular, magnesium and aluminum are theoretically very highly reactive and would be of little value if it were not for the invisible protective shield of oxide which forms spontaneously on their surfaces. Films of this nature are not impervious to all forms of attack, however, so one must be cautious. Coupled with certain metals or placed in some electrolytes, for example, both metals may be rapidly corroded.

Other passive films may be applied by the dipping of metals into chemicals, but they are not always permanent. The stainless steels do not usually corrode because alloying elements form a passive film at the surface.

2. REMEDIES FOR CORROSION

The elimination, correction, or arrest of corrosion is based on the elimination of any one of the components of the electrical circuit that makeup the galvanic cell. Any one of the following may be removed:

1. Of course, by various means a state of no potential gradient between the anode and cathode may be attained. With this condition there would be no driving force for the corrosion reaction.
2. The electrical connection between the anodic and cathodic areas could be eliminated.
3. Finally the electrolytic connection between the two poles can be eliminated or hampered by the use of certain insulating materials.

However, more specifically the following mechanisms are available.

Anodic Polarization

Anodic polarization shown schematically in Figure 6 may be obtained by allowing corrosion products to collect on the anode. This reduces the area available for ionization of the anodic material. Therefore, the potential of the anode will approach that of the cathode. This will lead to a reduction of corrosion. However, much care must be taken in the use of this mechanism because there is a very good chance that the accumulation of these by-products will cause what is known as an ion concentration cell. This type of cell often leads to quick and sometimes catastrophic failure.

Cathodic Polarization

Cathodic polarization shown diagrammatically in Figure 6 is basically the accumulation of the reduced materials on the surface of the cathode. This leads to diffusion difficulties. The oxidizing agent has difficulty diffusing through this film of by-products. Without the oxidizing agent to take up the electrons pumped to the cathode by the anode the reaction will slow and eventually come to a halt, if the polarizing film is sufficiently complete. The danger encountered in the use of this mechanism is slight but nevertheless, very real. Recognition of this danger, Anode-Cathode Reversal (discussed later), can be used for judicious application of anodic, and cathodic polarization as well, for a protective mechanism.

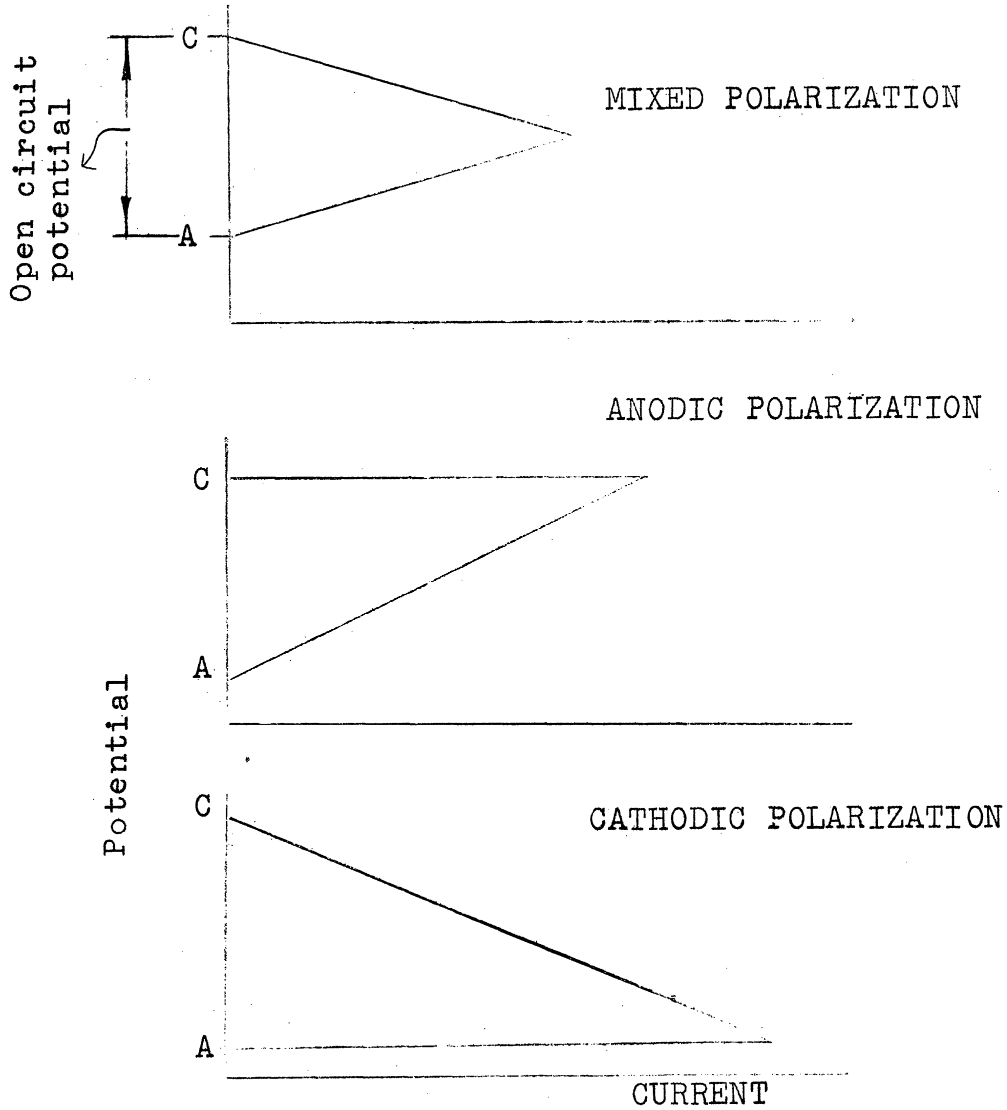


Figure 6

Polarization of Local Cells

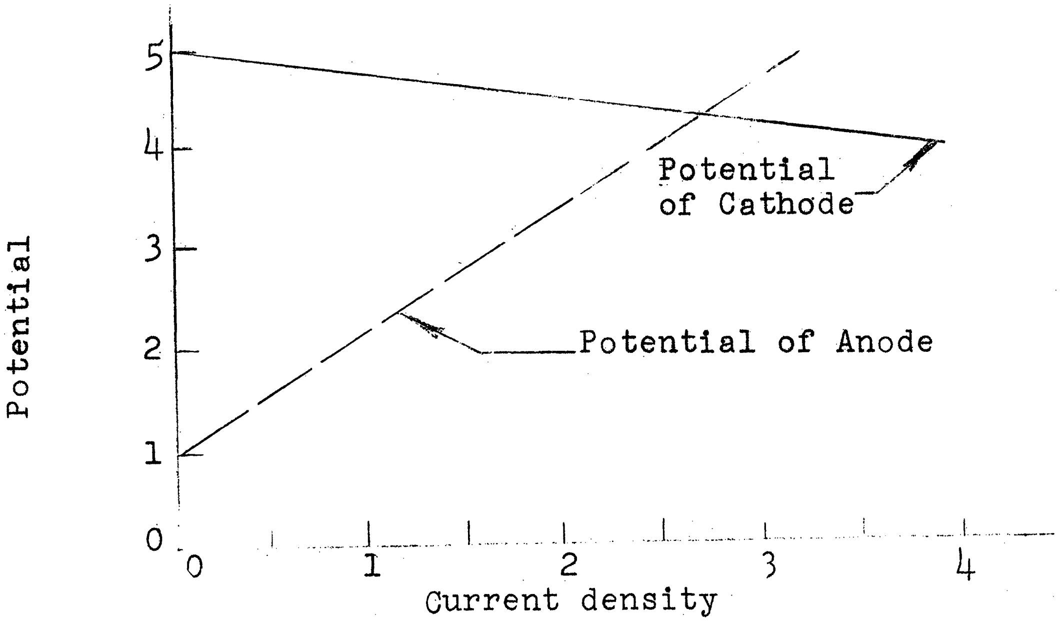


Figure 7

Anode-Cathode Reversal

Area Effect

The area effect in corrosion is very real and of great importance to the practicing engineer. The area effect is the very property of a design that determines the current density at each of the electrodes. In turn the current density controls the rate of corrosion while the magnitude of the potential difference only indicates the tendency for the reaction to proceed to completion. It, therefore, should be required in all designs where there is even a remote possibility of corrosion that the anode should be large with respect to the cathode. This allows the corrosion present at the anode to be spread over a large area and, therefore, yields a slow deterioration of the anodic surface. It follows that the cathodic area should be kept as small as possible so that the area of acceptance of electrons will be minimum. Naturally it requires much less polarization of the small cathode than of the large cathode to provide sufficient corrosion protection.

It is obvious that, with the use of a small anode and a large cathode, corrosion would be extremely concentrated. This is due to the fact that a very concentrated attack is the result of a specific amount of corrosion coming from a small area. In addition long periods of time are needed to provide sufficient cathodic polarization to protect the anode.

Anode-Cathode Reversal

At this time it should be noted that with polarization of both the anode and the cathode a situation can be reached where, as shown in Figure 7, the cathode has a greater solution potential than the anode. This leads to the corrosion of the "cathode". Knowledge of this phenomena is useful in the protection of certain material which normally cannot be protected by an inexpensive anode. The application of this potential reversal can allow sacrifical protection in many cases with an inexpensive metal.

Use of the Environmental Series

Many cases of so called "Cathodic Corrosion" have been found but are a result of an entirely different mechanism. In referring to Table 2 it is seen that the solution potentials of various metals vary in different electrolytes. Zinc, for example, in mineral water at elevated temperatures becomes anodic with respect to steel which is generally thought of as cathodically protected by zinc. It is, therefore, well to investigate the corrosive environment thoroughly before a decision is made as to which material is cathodic and which material is anodic. Of course, the reverse is also true. Now zinc may be used with a protective coating of steel to yield cathodic protection of the zinc in hot mineral water. Therefore, it is advantageous for the engineer to determine the actual solution potential in his specific situation. One of the fastest ways of determining the solution potential of the particular situation encountered is to set up an experimental galvanic cell which very closely approximates the physical setup. The open circuit voltage is then measured to determine the actual position of the anode and the potential difference between them.

Potential Difference Removal

Corrosion may easily be eliminated in an engineering structure if the potential difference is completely removed. This may be done by making the item in question of one material and placing it under a uniform environment. Generally, this is not feasible; therefore, compromise must be made by chosing metals that are as close as possible on the environmental galvanic series. Selecting materials close on the environmental series does not always result in a decrease in corrosion. This is due to the fact that in many cases a change in the metal couple also gives a change in the polarization effects. In turn this may lead to greater corrosion than formerly was encountered. Therefore, it is easily seen that much care must be taken in the selection of metal pairs in a corrosive atmosphere

Environmental Changes

The use of anodic and cathodic inhibitors, such as chromates and silicates, cause an increase in the anodic and/or cathodic polarization. This then decreases or eliminates corrosion.

Purification of the electrolyte, such as the elimination of oxygen from boiler feed water, will often reduce corrosion. This is due to the fact that the removal of these contaminants from the system will either eliminate depolarization of the electrodes or eliminate accumulation of mineral deposits that cause oxygen and ion concentration cells.

The possible elimination of the electrolyte also stops corrosion. This may be accomplished in various ways. For example, the use of free air circulation, natural drainage, elimination of electrolyte traps, the use of an electrolyte-proof coating, or the use of a nonelectrolyte will reduce or eliminate corrosion.

Cathodic Protection

Two major ways of obtaining cathodic protection are available to the engineer that of sacrificial or induced protection.

Sacrificial cathodic protection consists of choosing a metal that is anodic to the one that you wish to protect provided that sufficient electrical connection and electrolyte are present to produce a galvanic cell. This leads to the sacrificial corrosion of the less noble (but less valuable) material. Material in general use as sacrifical anodes are scrap iron, aluminum, cinders, and magnesium. The details of sacrificial corrosion will be eliminated here but are discussed in detail in may reference texts. However, it is important to note that the sacrificial anode should have a chemical depolarizer to prevent the polarization of the anode. If the anode becomes polarized the corrosion protection will become almost nill.

Third Metal Protection

For the protection of a two metal structure many times a sacrificial cathodic cell is constructed by placing a less noble material in electrical contact with metals. This type protection can be made much more effective and longer lasting if the cathode is coated by a protective coating.(See Figure 8)

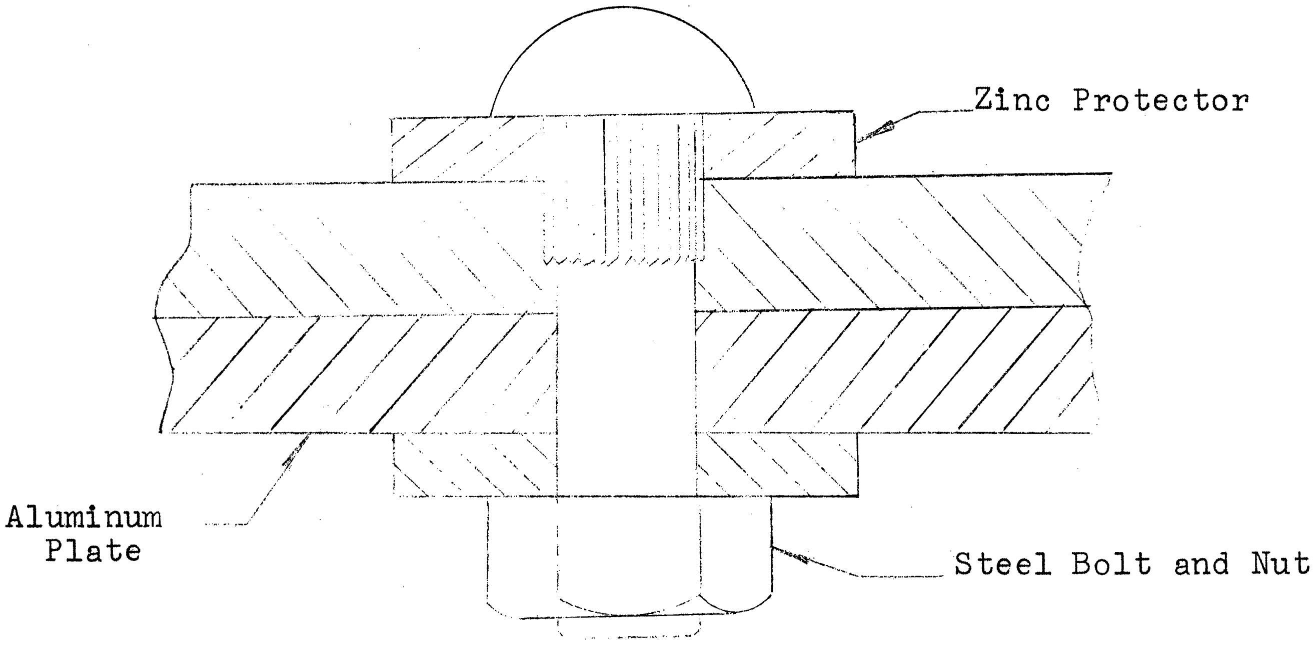
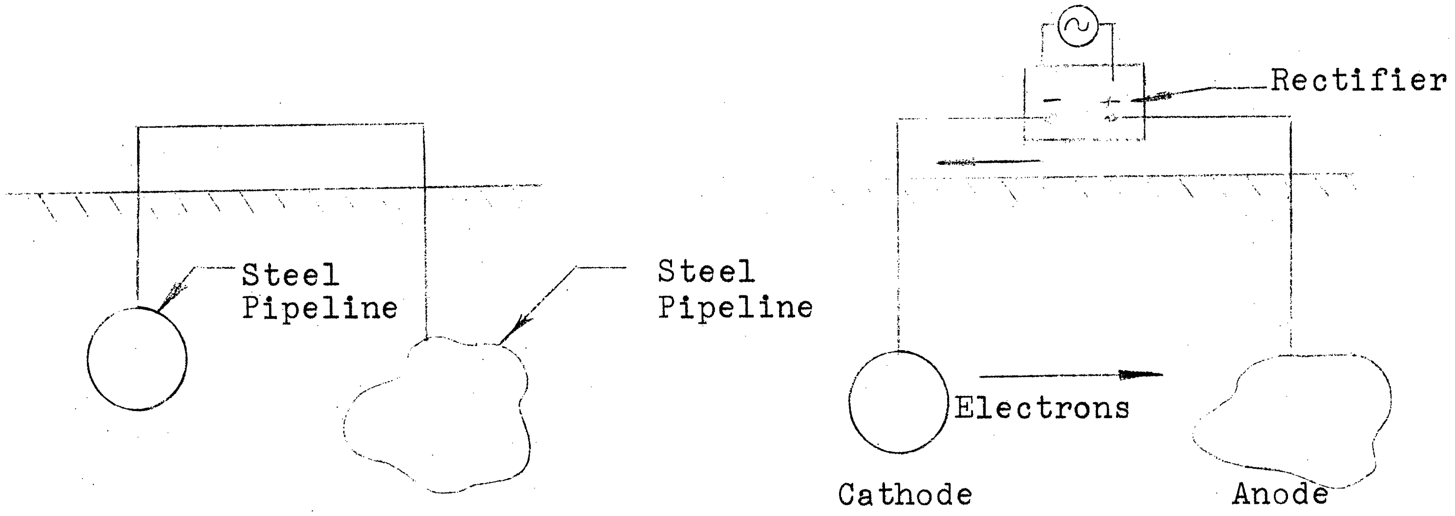


Figure 8
Third Metal Protection



Solution Potential are the same therefore corrosion proceeds on the pipeline

Rectifier drives scrap iron anodic therefore pipeline is protected

Figure 9
Induced Cathodic Protection

Induced Cathodic Protection

Referring to Figure 9 it is seen that induced cathodic protection takes advantage of the ability of a DC source to reverse the direction of current flow thus driving the anode to a cathodic state. The major limitations encountered here are the economic feasibility of this protection. In many cases where depolarizing elements are not present or where initial protective coating applied to the surface of the structure that is to be protected is insufficient, the cost of the power required to protect the structure over a period of years may exceed the cost of replacing the structure. This high power requirement may be reduced by:

1. The application of a sacrificial anodic coating, such as zinc on steel, to supplement the induced cathodic protection,
2. the application of an inert protective coating that has been inspected closely for flaws (90-99% coverage),
3. and using a material that forms a self regenerative oxide or protective film, such as stainless steel.

Anodic Protection

In general, anodic protection is a very dangerous mechanism for the elimination of corrosion. In anodic protection the anodic material is covered by an inert or cathodic material. But having discussed the area effect, it is realized that, if the coating should fail at one point, there will result extreme current densities. This is naturally followed by a high rate of corrosion in a given area. Therefore, unless there is almost an iron clad guarantee of the anodic protector's perfection and retention it is best to leave the anode completely bare and cover the cathode. Covering both the anode and the cathode may in <u>some</u> cases further reduce the corrosion plus stop the local surface cell action.

Elimination of the Electrical Connection

Referring to Figure 10 it is seen that an electrical insulator can be used to completely isolate one metal from the other. Cautious use of this mechanism can yield good results, but careless use can yield mositure retention and crevice corrosion. This crevice corrosion can be eliminated by the use of a pliable, electrolyte-proof, insulating material. If an absorbent material is used the electrolyte will be in continuous contact with the corrodable materials thus leading to the <u>promotion</u> of corrosion.

Crevice Elimination or Isolation

Knowing that corrosion is the result of an electromotive difference between metals or areas of a metal it is recognized that the crevice provides the possibility for the potential difference because of a difference in the oxygen or electrolyte ion concentration in the crevice from the concentration on the exterior of the crevice. There are many means of correcting this crevice situation a few of which follow:

1. Eliminate the crevice.
2. Provide a well vulcanized non-porous rubber cap.
3. A thorough coat of some organic coating or specially designed petroleum product.
4. Provide a coating of a non-brittle paint before assembling mating surfaces.
5. Caulk the crevice with a plastic caulking.

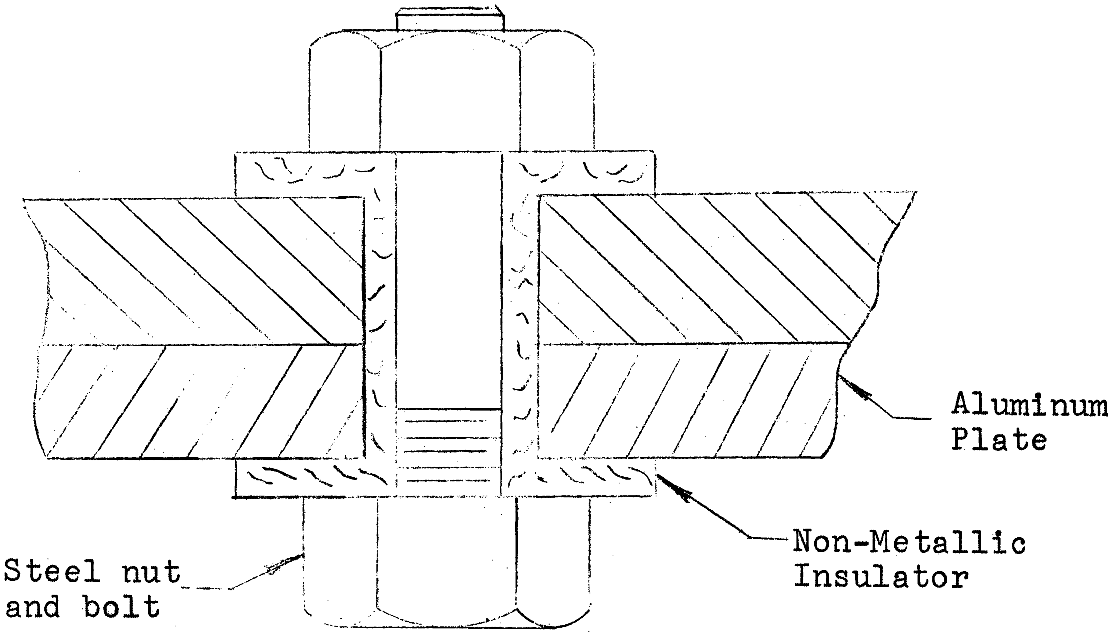


Figure 10
Breaking the Galvanic Circuit

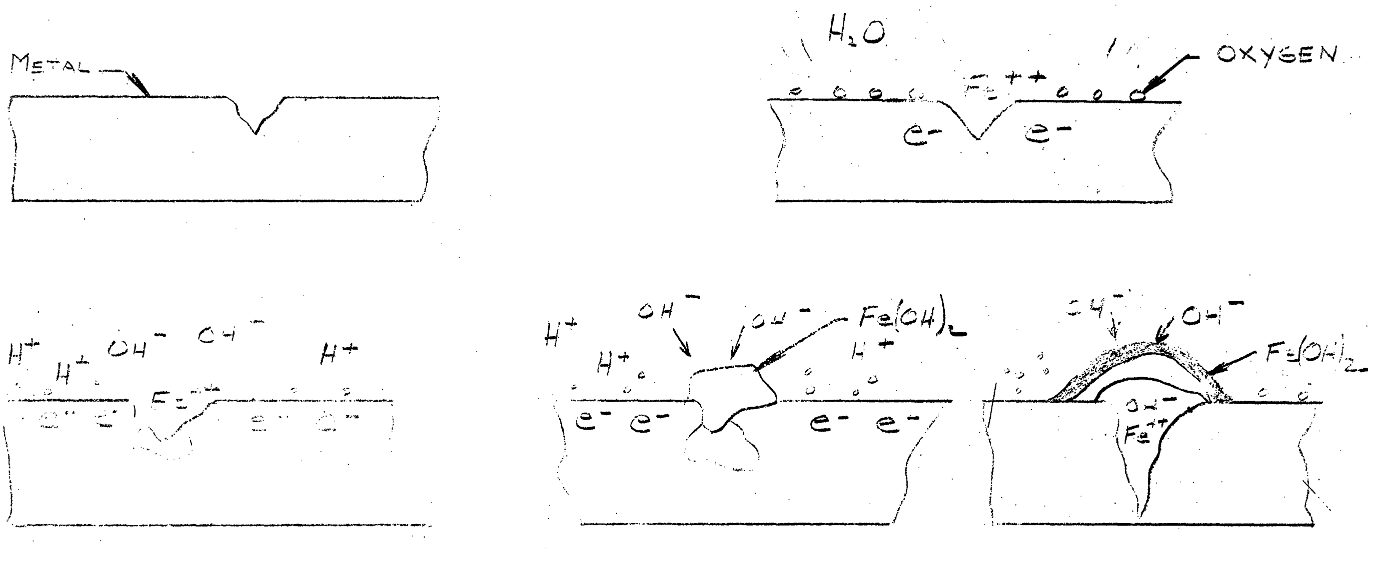


Figure 11
Oxygen Concentration in an Aqueous Electrolyte

The idea behind all these methods is to seal the electrolyte from the crevice and eliminate the corrosion therein.

Cleaning

The life expectancy of many materials under corrosive attack can be extended three to four times as long as would be normally expected, by the use of cleaning. Both manual cleaning and chemical cleaning techniques are very satisfactory. The major attempt in cleaning the surface of the corroding metal is to remove the accumulated minerals and corrosion deposits. These deposits tend to setup concentration cells which results in rapid penetration of the metal in a very localized area. This leads to an early failure. Therefore, if these areas of accumulated products can be eliminated the corrosion can be reduced.

Types of Corrosion

With the basic methods for the arrest of corrosion at hand this chapter will now proceed to site, explain, and present examples of each type of corrosion. Then a solution or solutions to the particular problem will be selected from those developed above.

Corrosion as an element of wear may be thought of as strictly a galvanic action between a metal and its environment. Even zinc being attacked by hydrochloric acid is galvanic action because on close examination it is seen that specific areas are acting cathodic thus forming hydrogen or other cathode product.

Just adjacent to this area is found an anodic area which is dissolved away. However, corrosion can be divided into two major divisions:those of Environmental Differential Cells and Material Differential Cells.

Environmental Differential Cells

Concentration cells are those galvanic corrosion cells setup between two areas on the same material because of the differential in environment between the two areas. This differential causes a change in the solution potential yielding a potential difference. The various types of environmental differences provide the basis upon which to separate the concentration cells into their subgroups.

Oxygen Concentration Cells

Referring to Figure 11 it is noted that the area of the limited oxygen is anodic and that the area of the higher concentration of oxygen is cathodic. This is the result of the fact that at the oxygen rich cathode the solution potential of the metal ions is much less than at the oxygen deficient area. With the migration of the excess OH^- to the anode in the presence of oxygen form a porous cup of $Fe(OH)_3$ over the anode. The OH^- ions easily diffuse through the cup and then upon reacting with the Fe^{++} complete the galvanic circuit. This eliminates polarization of the anode by relieving the number of cations at the anode. The corrosion can therefore proceed. In fact the rate of penetration may well become very large in certain places where the oxygen starvation is great, where the area outside the oxygen starvation area is large, and where the area of the oxygen starvation is small. In these cases very little galvanic current is necessary to yield complete failure. Oxygen concentration cells are generally associated with both pit and crevice corrosion.

Metal Ion and Other Ion Concentration Cells

Metal ion and other ion concentration cells operate on the basis that the solution potentials of a metal are affected by the common ion effect and/or by the change in the concentration of the electrolyte at different points along the structure. For example, corrosion of a pipeline is a common metal ion cell. This cell is due to the differential in environment encountered by the pipe at different positions along it. The pipe, therefore, becomes anodic in places and cathodic in other places. This leads to a corrosion cell using the moist soil as the electrolyte and the pipe itself as the external electrical connection between the anode and the cathode.

To correct this situation the surface of the pipeline should be coated as completely as possible. There is a choice of two methods of protecting against local action where the protective coating fails. The first is the use of a natural anode to give sacrificial cathodic protection. Where heavy corrosion is encountered it is probably best to use induced cathodic protection. Availability of power, size limitations, economic considerations, corrosion distribution, and corrosion load all are important factors in the decision as to which method is best to use in the protection of the pipeline.

Active-Passive Cell

The active-passive cell is a severe manifestation of the oxygen concentration cell. This situation is found in crevice assemblies of stainless steel. Stainless steel depends upon oxygen for the immediate repair of its ruptured protective coating. However, in the crevice situation, such as in a lap weld, the material in the crevice has no ready access to oxygen. This situation leads to a high solution potential in the crevice and a low potential in the area of the external surface. A galvanic cell is then set up between these two areas. To eliminate this condition, eliminate the crevice. If this is not practical, change the materials, remembering that even a small potential difference can cause high rates of corrosion if the areas have a bad relationship. Many nonmetallic materials have found extensive use where the failure of stainless steel is prevalent. For example, the PVC or polyester resin piping have been found very acceptable where stainless steel failed.

Corrosion Induced by Non-metallics

Another manifestation of the oxygen concentration cell is that of non-metallic corrosion. This title is probably a misnomer, but for lack of a more appropriate title it shall be used. Conditions for a galvanic cell are induced when a non-pliable (does not completely seal off the interface of the metal and non-metal), porous or semiporous, non-metallic material is in contact with the metal. When the metal is exposed to alternate environmental conditions, the non-metallic absorbs and retains the electrolyte. This then leads to ion and oxygen concentration cells and corrosion proceeds. The easiest means of correcting this situation is to remove the non-metallic material. If in the process of eliminating the non-metallic a crevice is produced, adequate area and protective coating considerations should be made for the elimination of corrosion in this new situation. However, if the nonmetallic insulation or gasket is necessary in the structure, it should be replaced with a material equivalent in the properties desired, but also it must be of a nonporous or nonsoluble nature. It is realized that this may not be always possible; therefore, as a last resort a coating of a pliable, electrolyte-proof material should be applied to the interface of the metal and non-metal. This will seal the electrolyte from the junction and therefore break the electrical circuit. This coating should be renewed periodically to maintain this type of protection.

Pocket Type Corrosion

The last of the concentration type corrosions to be considered is that of the pocket type corrosion. Pocket corrosion is found on many structures which have traps or pockets in which electrolyte or corrosive may collect. This again causes a difference in environment and a difference of solution potentials. Thus the corrosion proceeds.

Probably one of the most recognizable forms of this type of corrosion is that of rusted rocker panels on automobiles. The holes in, and the crevices and corners around the rocker panel provide excellent opportunity for the collection of salt and other electrolytic agents. This leads to continual corrosion from the underside of the panel. Continuous corrosion of this nature leads to a rapid failure of the sheetmetal. Several things can be done to eliminate the corrosion in these pockets. A break with tradition could result in a car without rocker panels thus eliminating the problem. A complete elimination of the boxed in portion would decrease corrosion by increasing air circulation and by facilitating drainage. In addition an improved undercoat could be applied to the rocker panel to aid in its protection. The present undercoats have a tendency to dry and crack after the period of a year or so. This leads to the possibility of concentrated attack in a given area. Possible improved materials would be some of the newer epoxy resins, if the cost of insulation could be appreciably reduced to the point of economic feasibility.

Velocity Corrosion

Velocity corrosion cells are setup upon the surface of a metal due to the difference in solution potential resulting from different electrolyte velocities at different areas of the metal. If the relative motion of an electrolyte with a rotating circular disc is considered, it is readily recognized that the velocity is greater at the periphery than at the center of the disc. Because of this motion more ions are carried away at the periphery; thus a higher potential exists at the periphery. Therefore, the periphery, being anodic, corrodes. One of the best examples of this type of situation is the corrosion of the extreme portions of a brass propeller. The easiest method of correcting this situation is to change the material to one that has little or no change in the solution potential with velocity. In the case sited, the use of a monel propeller is suggested because of its low change in solution potential.

Erosion Corrosion

Where high temperatures and poorly selected materials for corrosion resistance in a particular electrolyte is encountered, a failure called erosion corrosion exists. Erosion corrosion is generally the result of using an electromotive series instead of the environmental series for the determination of the anodic and cathodic materials. In addition to this situation it is possible to have a situation where the metal dissolves in the electrolyte just as sugar might in water. The spectacular washing away of the metal has led to the use of the term "erosion corrosion".

An outstanding example of the mechanism is the failure of the impeller of a centrifical pump, which having been designed to last close to five years, failed in the period of three weeks. Corrosion of this type must be checked quickly. In this case either the electrolyte or the metal in the impeller had to be changed. It was found that the solution that was being pumped was of an alkaline nature and could easily dissolve the metal. The electrolyte was then made more acidic and the reduction of the metal's solubility was the result. Different materials, lower temperatures and reduced speeds also can solve this problem.

Pure Galvanic Corrosion Cells

All the topics mentioned under "Environmental Differential Cells" are basically a specific manifestation of galvanic corrosion due to the differential concentration of some element in the galvanic circuit. Now it is the task of this chapter to investigate the corrosion resulting from discontinuities in the material. These are chosen to be called pure galvanic cells. This results from the fact that a cell resulting from a differential in metal is more intuitively a galvanic cell than those cells that result from a differential in environment.

Uniform Corrosion

The most common form of pure galvanic corrosion is uniform corrosion or chemical attack. Uniform attack is the result of material inhomogenities on a microscopic level. The summation of these microscopic galvanic cells results in the appearance of uniform corrosion. These microscopic areas of discontinuity result in an anode being formed in one area while the adjacent area become cathodic. Thus corrosion results.

A piece of unprotected steel roofing is an excellent example of uniform corrosion. Correction of this situation may be accomplished by one of the following;

1. Use an inert protective coating such as paint.
2. Apply a natural cathodic protector such as zinc.
3. In rare cases induced cathodic protection is used.
4. Apply an anodic protector that provides complete coverage.
5. Use a corrosion inhibitor to cause polarization of the local cells.

Joint Corrosion

Unexpected corrosion problems appear when the engineer welds, force fits, or contacts two metals intimately. This is referred to as joint corrosion. The solution potential difference on the environmental series plus the additional driving force provided by the stresses envolved result in corrosion.

A common example of this type of corrosion is present in many automobiles. The aluminum door sills which are in close contact with the steel rocker panel corrode badly at the interface of the metals when they come in contact with an electrolyte. However, the aluminum forms a tenacious oxide which must break down before corrosion can proceed. When this happens the penetration of the aluminum is very rapid because of the area effect. The economics of this situation do not warrant a change.

However, in general, joint corrosion can be corrected by area relationship changes, by the proper selection of materials, and by the use of a protective coating.

Cathodic Reversal

Another corrosive mechanism that should be discussed is that of cathodic reversal. In this situation two metals may purposely be set into a corrosive medium with the main purpose of cathodic protection. Then upon progression of the reactions the anode and the cathode may become polarities as shown in Figure 7. Corrosion proceeds on the metal that was thought to be cathodic and is therefore in direct contact with the fluid.

For example, a steel boiler that was coated with a zinc dip developed severe corrosion failures in short periods of time of service in certain alkaline waters. This boiler failed because of the polarization that resulted in the electrolyte used. A few simple tests of the solution potential and what happens to the solution potential as the reaction proceeds in the required environment would have told the design engineer that he would have the problem of cathodic reversal. After repairing or replacing the boiler a chemical inhibitor could be added, the coating metal could be changed, or the base metal could be changed.

Cathode Corrosion

It has been noted formerly that there is, in most cases, some material plated out on the cathode. This causes, in cases where this material is allowed to accumulate, a differential on the surface of the cathode. This leads to the production of local concentration cells. This type corrosion is found in induced cathodic protection. The current used to buck the local cells is generally not evenly distributed. This results in high currents in areas where it is not needed. Excessive accumulations of corrosion products are found and corrosion results on the cathode. Correction of this situation may be accomplished by a better control of the current distribution, more complete coverage of the surface by a protective coating, or by the use of a backfill of chemical inhibitor that will prohibit the formation of these corrosion products.

Alloy Failure

Many instances of inhomogenities of the stainless steels and other corrosion resistant alloys have resulted in the development of a local cell between the passive surroundings and the active point of inhomogenity. This is a very real problem encountered in industry.

Recently one of duPont's cellophane plants encountered a failure in a stainless steel CS_2 line. This failure was only the size of a pin head and resulted from a failure in the alloy. Since the system was under high pressure this pinhole caused the loss of enough flamable CS_2 to cover the roof of the building in a short time. Because of this extremely dangerous situation serious thought is being given to the possibility of the use of PVC or epoxy resin piping system to carry this fluid.

Replacement of the stainless steels and other super alloys seem to be more and more concentrated in the area of plastics, inert or semi-inert materials. This fact is possibly due to the fact that these materials offer the last resort for the elimination of corrosion at a fairly reasonable price.

Intergranular Corrosion

On a microscopic level corrosion is found to exist in the grain boundries of the metal. This is due to the fact that the material in the grain boundries is highly stressed due to the tendency for the grains on either side to pull this material into themselves. Since this is an area of high internal energy and since it is known that a system always tend to its state of lowest energy, it is not out of order to expect that this area will go into solution to a greater extent than the grains in order to reduce the energy level of the system. This gives rise to corrosion. If intergranular corrosion proceeds, a total loss of the grain boundries will result and little or no strength will remain.

Dezincification

Dezincification is a misnomer applied to the preferential removal of any phase of an alloy from its microstructure. A potential cell is then set up between the different area of the microstructure. This is the presence of an electrolyte, promotes corrosion.

Several examples of dezincification can be found but the graphite residual, after a cast iron piping elbow has had the iron preferentially removed, is the most impressive. The cast iron elbow can now be cut apart with a knife. The best methods for the elimination of this type of corrosion are the proper and judicious selection of materials, proper heat treatment, and the use of proper inhibitors in the electrolyte.

BIBLIOGRAPHY

1. Clark, Donald S., and Varney, Wilbur R., *Physical Metallurgy for Engineers*, D. Van Nostrand Co., Inc., Princeton, New Jersey, 1958, pp.241-91.
2. Ellis, O. B., and LaQue, F. L., "Area Effects in Crevice Corrosion", *Corrosion*, Vol. 7, No. 11, November, 1951, pp.362-4.
3. International Nickel Company, *Corrosion in Action*, International Nickel Company, New York, 1955.
4. LaQue, F. L., "Galvanic Corrosion", *Engineering Approach to Surface Damage*, Lipson, Charles, and Colwell, L. V., Ed., University of Michigan, 1958, pp. 221-47.
5. LaQue, F. L., and Cox, G. L., "Some Observations of the Potentials of Metals and Alloys in Sea Water", *Proceedings of the American Society for Testing Materials*, Vol 40, 1940, pp. 670-687.
6. Speller, Frank N., *Corrosion Causes and Prevention*, Third Edition, McGraw-Hill Book Co., Inc., 1959, pp.1-45, 657.
7. *A Short Course in Corrosion*, University of California, February 2-6, 1953.
8. de S. Brasunas, "Nature of Corrosion", *Symposium on Corrosion Corrosion Fundamentals*, University of Tennessee Press, 1955, pp. 13-27.
9. *Corrosion and Temporary Protectives*, Shell Petroleum Co., Ltd., 1954, pp. 24-29.
10. Evens, Ulrich R., "*Metallic Corrosion Passivity and Protection*," Butler and Tanner Ltd., London, England, 1937, pp 6-11.
11. Fontana, M. G., "Discussion," *Corrosion*, March 1947, pp. 118-121.
12. Fontant, M. G., "Some Corrosion Failures In Chemical Plants," *Symposium on Corrosion Fundamentals*, University of Tennessee Press, 1955, pp. 120-126.
13. Gatos, C. A., "Some Inhibitors of Metallic Corrosion in Aqueous Media," *Symposium on Corrosion Fundamentals*, University of Tennessee Press, 1955, pp. 127-132.
14. Heussner, C. E., "Protection of Base Metals By Use of Metallic Coatings," *Symposium on Protecting Metals Against Corrosion*, American Society for Testing Materials, April 19, 1938, pp. 6-15.
15. Hoxeng, R. B., and Prutton, C. F., "Electro-chemical Behavior of Zinc and Steel in Aqueous Media," *Corrosion*, Vol. 5, 1945, pp. 330-339.
16. Means, R. B., and Brown, R. H., "Designing to Prevent Corrosion", *Corrosion*, March, 1947, pp. 97-118.

17. Robald, E., Corrosion Guide, Elsevier Publisher, New York, 1951, p 21.

18. Robinson, H. A., "Cathodic Protection Principle," A Short Course in Corrosion, February 2-6, 1953.

19. Trigg, C. F., An Engineering Approach to Corrosion, Sir Isaac Pitman and Sons Ltd., 1952.

20. Wirshing, R. J., "Pretreatment of Metals," Symposium on Protecting Metals Against Corrosion, American Society for Testing Materials, April 19, 1938, pp. 16-19.

CHAPTER 12

CAVITATION CORROSION

1. CAVITATION THEORIES

Poulter's Theory

"Cavitation erosion is the result of penetration of a liquid into the pores of a metal and its subsequent escape carrying some of the metal with it."

This theory, propounded by T. C. Poulter in 1942 is thought to be substantiated as follows: The rate and amount of penetration of the liquid into the pores of the metal would be governed by molecular size and viscosity. Consequently, if penetration did take place, when two liquids with equal viscosities were tested, the liquid with the smaller molecules would create more damage than the liquid with the larger molecules. Likewise, if the molecular sizes of two test liquids were equal, the liquid with the lower viscosity would cause the most damage. Further, the greater the porosity of the metal, the more severe the attack. According to this theory it should be possible to reduce the severity of the attack if the pores of the metal could be filled or sealed.

Cavitation erosion experiments were conducted with cast iron in water, ether, alcohol, glycerin and paraffin oil. Some of the test results are submitted below:

1. Water, having the smallest molecules and a low viscosity, caused the most extensive damage.
2. Ether (large molecules, low viscosity) caused more damage than glycerin (small molecules, high viscosity). Ether had,however, less effect than water.

3. All other considerations being equal, the test piece with the greatest porosity suffered the most damage.
4. A specimen, treated so that half of its surface pores were sealed, exhibited practically no erosion on the treated half while the untreated half was badly eroded.

Although Poulter's theory seems to be adequately substantiated it is considered to be an early attempt at explaining the mechanism of cavitation erosion and is now generally disregarded by most individuals.

Beeching's Theory

"Cavitation erosion is the mechanical disintegration of surface layers under repeated, highly localized pressure pulses. There pulses are caused by the sudden collapse of cavities in contact or close to the surface."

This is the theory of R. Beeching, who conducted numerous investigations in an attempt to find materials which could resist cavitation erosion and yet be suitable for ships propellers and turbine runners.

In support of his theory, experiments have shown that the collapse of cavitation voids can produce impact pressures by implosion exceeding the yield strength of most metals. Consequently, when the voids collapse, the surface of the metal is subjected to repeated local plastic deformations which result in its ultimate fracture and disintegration. Also, since cavitation erosion resembles progressive fatigue failure of the surface layers, it would be expected that for each metal there would be a limiting intensity of attack below which it would not erode. Beeching's experiments appeared to confirm this.

Beeching's theory is the most widely accepted of any and is considered to be the most plausible. Experiments by other investigators have shown that these impact pressures do exist and can cause damage depending upon their magnitude. This seems to further substantiate this theory.

Crewdson's Theory

"Cavitation corrosion is caused by super-heated steam at extremely high pressure. This steam is the gas or vapor in the cavitation bubbles."

In the presentation of evidence for this theory suggested by E. Crewdson, it is necessary to supply some background material. A majority of the investigators who have studied cavitation have stated that the bubbles form and collapse. Professor R.T. Knapp in his "Laboratory Investigations of the Mechanism of Cavitation," reported on photographs taken at a speed of 20,000 frames per second showing the formation and collapse of cavitation bubbles. Crewdson examined Knapp's film at low speed and stated that it did not reveal the complete collapse and disappearance of the bubbles. He further reasoned that for the bubbles to collapse and reform, a finite time interval would be required. This interval would be longer than the exposure time of a single frame, therefore, the disappearance of the bubble should have been evident in the film. Since it was not, he concluded that the bubble never collapsed.

To further substantiate his claim that the bubbles do not collapse, Crewdson cited the work of Lord Rayleigh. Lord Rayleigh had examined the behavior of a spherical bubble of a perfect gas introduced into a infinite field of water. He proved that if the water pressure was greater than that of the gas in the bubble the bubble would be compressed to a minimum finite size and then would re-expand. He assumed that the compression would be isothermal and showed that if there were no losses, the bubble would re-expand to its original size and thereafter would continue to contract and expand between two limits.

Crewdson developed Lord Rayleigh's equations on the assumption that the compression was adiabatic. He found that this curve resembled the experimental one more closely than the isothermal curve. When the vapor within the bubble was compressed to a minimum, calculation of pressure and temperature yielded values of 5180 psi absolute and 4950°F. Consequently, the mechanism of cavitation erosion was explained as follows:

The super-heated steam in contact with the metal softens it and the high pressure causes it to flow. Also, the momentary temperature difference existing between points that are and are not in contact with the bubbles could generate electric currents leading to electrolytic effects.

Crewdson's theory definitely has some value to it but it is, nevertheless, generally disregarded in favor of Beeching's theory. What Crewdson states may or may not be true but no one has brought forth any evidence to disprove his theory and conversely a lot of the evidence supporting his theory is relatively weak.

2. CAVITATION AND CAVITATION DAMAGE

For the start of this part, let us look at what cavitation is and why it is harmful. Cavitation is the formation of bubbles within a fluid which has relative movement with some solid surface. This relative movement causes the local pressure in the fluid to drop down below the vapor pressure of the fluid, causing boiling or "cavitation". This pressure drop occurs any place in a machine where fluid motion at low velocity might cause a vortex or "swirling" to form. Figure 1 shows several places where cavitation might be found.

These bubbles, formed in low pressure areas, are soon moved into high pressure areas where they become unstable. This means external pressure is greater than the pressure inside the bubble and the walls of the bubble are forced inward as the bubble implodes or collapses. Damage is the result of tiny areas of extremely high pressure produced near the point of final collapse. These small but intense pressure impulses cause mechanical cold working of the guiding solid surface. Repeated blows cause metal fatigue and results in surface cracks and eventual loss of metal. This is essentially the theory as formulated by Beeching.

The classic paper on this subject was written in 1917 by Lord Rayleigh, an Englishman. He considered the collapse of a spheroidal cavity in an incompressable fluid having constant pressure at infinity. He developed expressions for the kinetic energy of the fluid as it rushed toward the center of the bubble. Then, he equated this to the work done at infinity as the pressure acted through the volume of the bubble. From this equation, he got expressions for the velocity and acceleration of the interface, time of collapse, and the pressure in the surrounding fluid. He also found expressions for pressure developed as an empty bubble collapsed on an absolutely rigid sphere at its center.

When a low pressure bubble encounters a high pressure region the walls of a bubble are forced inward until they meet at the center. Much of the fluid must undergo very rapid deceleration, especially at collapse when the surface particles run head on into particles on the opposite side and must stop instantly. Rayleigh found that the maximum pressure occurred as a distance from the center of the sphere equal of 1.587 R where R is the instantaneous radius. He found this pressure to be 1260 times the pressure at infinity when an empty bubble reached 1/20 its original size. When considering the bubble collapsing on a rigid sphere, he found the pressure reached 136,000 psi. This was done by considering the fluid to be elastic at the instant of impact with the sphere. He put his calculated value of velocity into the familiar water hammer equation and found the pressure developed.

Let us now examine at bubble while it forms, grows, collapses. rebounds, recollapses, etc. This section will consider a single bubble flowing by itself and being unaffected either by other bubbles or by solid surfaces. While this is an extremely unlikely situation, it is a place to being study of basic bubble mechanics

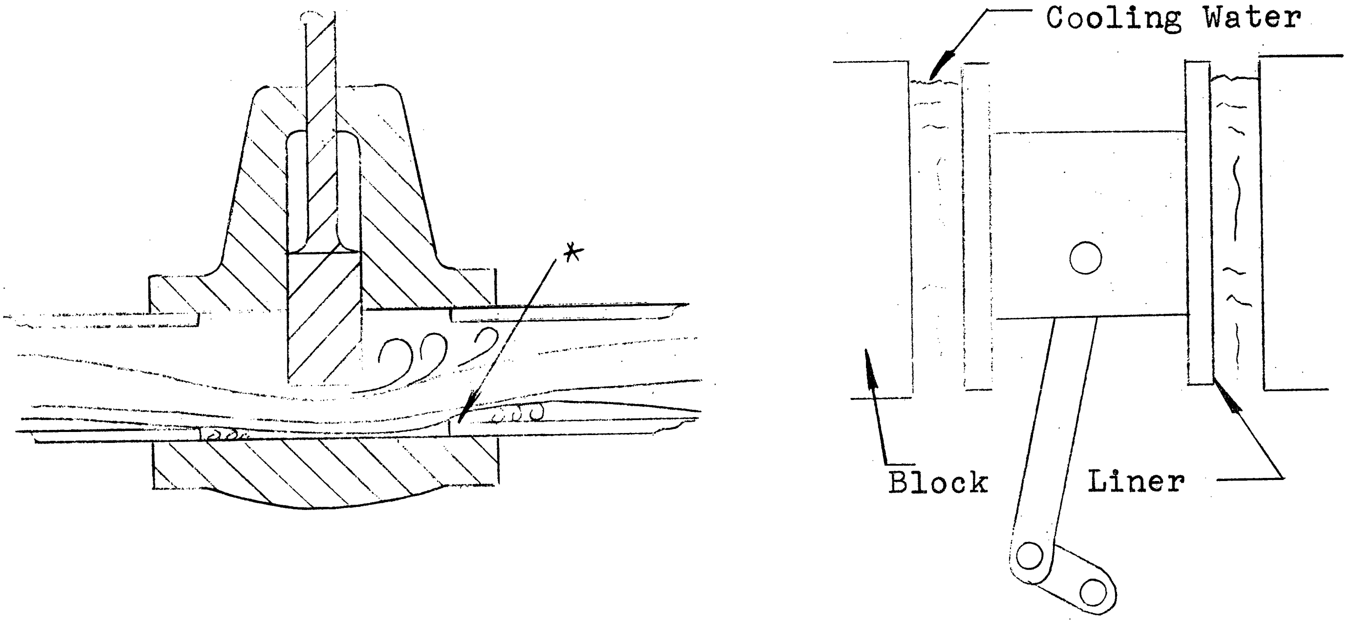


Figure 1

The swirls in the partially closed gate valve show potential cavitation areas.

As the cylinder walls vibrate, alternate high and low pressure areas in the cooling water cause bubble formation and collapse.

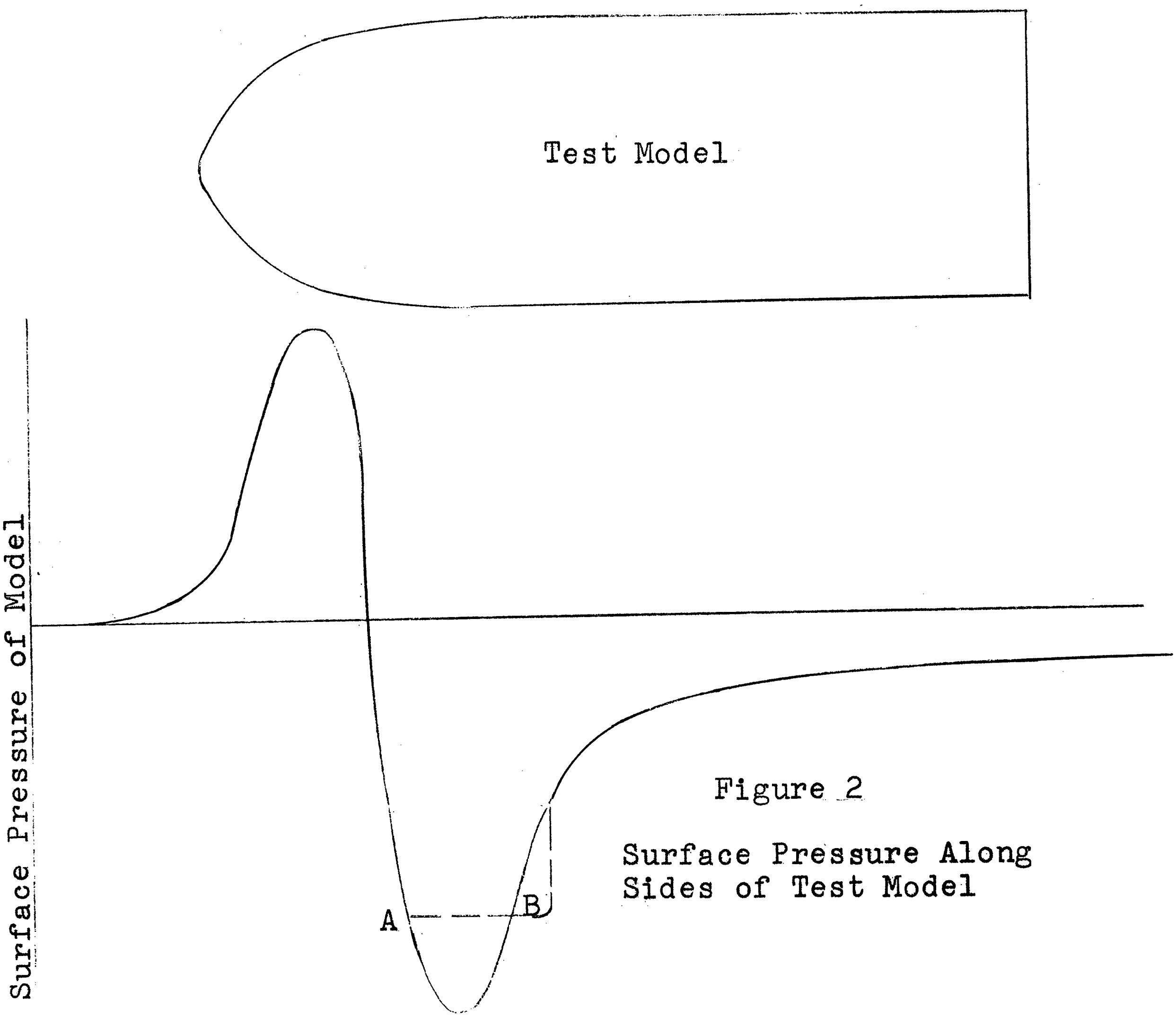


Figure 2

Surface Pressure Along Sides of Test Model without unnecessary complications which can be usefully included only after the fundamentals are understood.

Growth of a Bubble

Consider a water stream impinging on the front end of a bullet-shaped model. As any particle approaches the front end, it is forced to move radially away from the object in order to pass. The force required to change the fluid direction is shown by the increase in the surface pressure as noted in Figure 2. Soon, the particle is moving outward fast enough and the surface pressure drops to the static pressure in the tunnel. Now, as the surface of the model falls away from the flow direction, a radial inward force is required to hold the particle on the surface. This inward force comes from the static fluid pressure as long as this is great enough. Further inward push comes from a reduction in surface pressure. This is also shown in Figure 2 where the the surface pressure of the model drops down to the static fluid pressure. The minimum surface pressure is reached when the vapor pressure of the fluid prevails. At this point, any further reduction in surface pressure causes cavitation bubbles to form. This corresponds to point A in Figure 2.

What is in a Bubble?

At this time, we might ask what is in these bubbles. Two obvious possibilities are water vapor and air. If air filled the cavity, it would have to migrate through the adjacent water a great distance in order to fill the volume of a bubble. Consider an average bubble whose growth time is about .0025 sec. Assuming air concentration to be about 2% by volume, air would have to migrate .001 of an inch in .0025 seconds in order to fill an average bubble with a pressure of .3 mm. Hg. Actual measurements show bubble pressures to about 97 times as high.

Checking the previous assumption of water vapor inside the bubble, we start by assuming three things:

1. The vapor pressure in the bubble is in equilibrium with the temperature of the layer of liquid at the bubble boundary.
2. The vapor to fill the bubble is produced by the evaporation of a uniform thin layer of water from the surface of the bubble.
3. The necessary heat for this evaporation comes from the heat of the liquid of a shell of uniform thickness surrounding the bubble.

Assuming further that the vapor temperature doesn't drop over 10° from that of the surrounding water, a rough calculation shows the thickness of the evaporated shell of water to be 10^{-6} inches and the thickness of the shell of water losing heat to be 10^{-4} inches. These valves are very reasonable even for times as short as are encountered here.

Now we turn our attention back to the bubble. It continues to grow as long as the pressure remains below the vapor pressure. At point B in Figure 2, pressure comes back up to the vapor pressure and bubble growth should stop. However, radial kinetic energy carries the walls a little further out. This puts the bubble well into a region of relatively high pressure where it is very unstable.

Collapse

This leads to the collapse phase which is the damaging part of cavitation. As soon as the bubble feels this high pressure, the water vapor inside begins to condense. The combination of increasingly high external pressure and increasingly rapid water condensation leads to final collapse velocities measured as high as 760 ft./sec.

At this point, pressures in the surrounding water build up and compress the water storing elastic energy. When the inward kinetic enrgy of the bubble is spent, this elastic energy goes into producing outward radial velocity. In this way, another bubble is formed in a high pressure zone. This bubble collapses immediately and possibly continues the rebound and recollapse cycle several times before completely dissipating all energy.

Fixed Cavity Cavitation

All cavitation falls into three general categories.

1. Fixed cavity cavitation.
2. Moving cavity cavitation.
3. Vortex or "tip" cavitation.

Since fixed cavity cavitation is by far the most prevalent and damaging of the three, a closer look at it will be made here. Consider the situation in Figure 3, which might be a blow-up of the part of Figure 1 marked with an asterisk. The fluid leaves the guiding surface at point A and returns at some point downstream. A stagnation point develops dividing the flow into re-entrant flow and that part which continues downstream. At

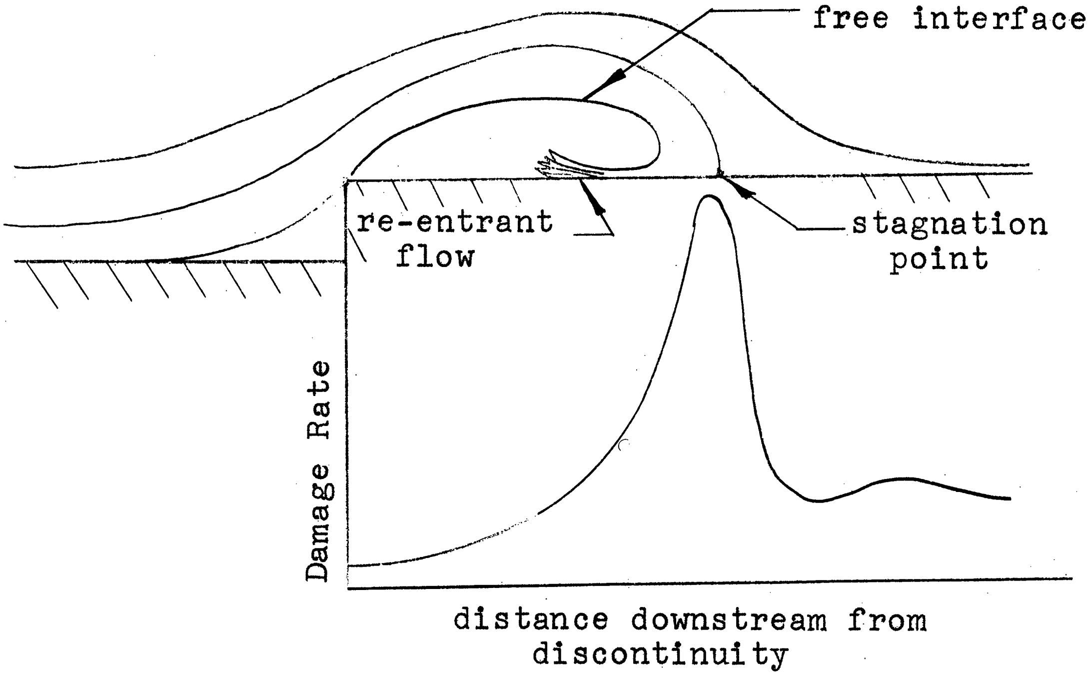


Figure 3

Cavitation Example

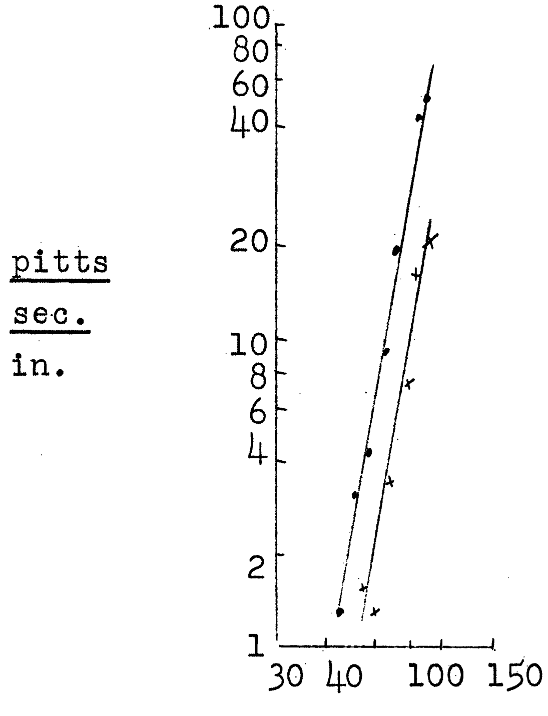


Figure 4

Velocity Test Results

this time, the bubble breaks off from the guiding surface and floats downstream. Another bubble immediately starts to form. Thus, the "fixed cavity", the distance between break-off and re-contact, is not fixed at all but varies from zero to a maximum with the cycle of cavity formation and breakoff. It is considered fixed only in a statistical sense.

A close look at this cavity shows the interface to be covered with tiny individaul cavitation bubbles. These bubbles are swept into the stagnation area and become part of the re-entrant flow. The damage is caused by these bubbles as they are forced onto the guiding surface and are collapsed with great violence by the extra stagnation pressure. As stated earlier, the large number of bubbles causes cold working and eventual surface fatigue and failure.

A plot of damage against distance downstream from cavity beginning is made in Figure 3. It shows maximum damage to be at the stagnation point. Figure 3 also shows a slight damage rise downstream a short way. This is probably caused by the rebound and recollapse of some of the larger bubble.

Tests of this type which were run with different water speeds show the damage to be roughly proportional to the sixth power of velocity. This was found by running soft aluminum specimens at different fluid velocities and for short periods of time. Damage was checked carefully after each run and the results were as shown in Figure 4.

Test Results

There are at least four common types of testing apparatus used today in cavitation studies. One which is gaining in popularity is diagrammed in Figure 5. However, there is now no standard test set-up, no standard procedure or test material. Add to this the fact that the mechanism of damage is only partly known and it becomes easy to see why many tests give conflicting results. Let us now look at some of the variables affecting cavitation damage and see what effect they have on test results.

Hardness is one variable which has a consistant and positive effect. The harder a piece is, the better it is able to stand the pounding it gets from the collapsing bubbles. Relative velocity is another variable which gives consistant results. It was pointed out earlier that damage varies approximately with the 6th power of velocity. High external pressure has the effect of retarding bubble formation. This usually effective, but isn't always practical.

High vapor pressure tends to encourage bubble formation, thereby creating conditions for cavitation damage. A liquid-like castor oil cannot be forced to cavitate because of low vapor pressure. On the other hand, some liquids will boil or cavitate at room temperature and pressure.

Other liquid properties such as surface tension, molecular size and viscosity all effect cavitation for some unknown reason. The lower the molecular size and viscosity and the wetter a liquid is, the more apt it will be to cause surface damage.

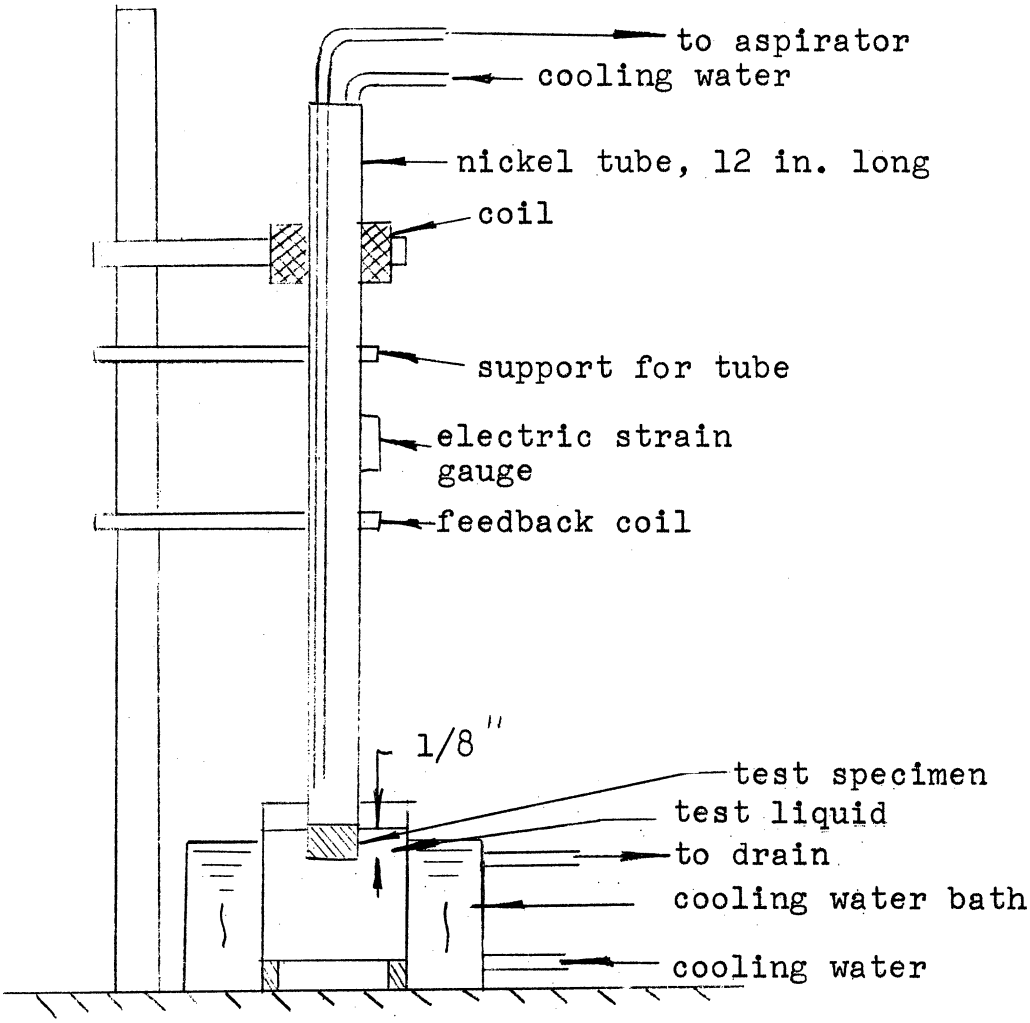


Figure 5

Cavitation Testing Apparatus

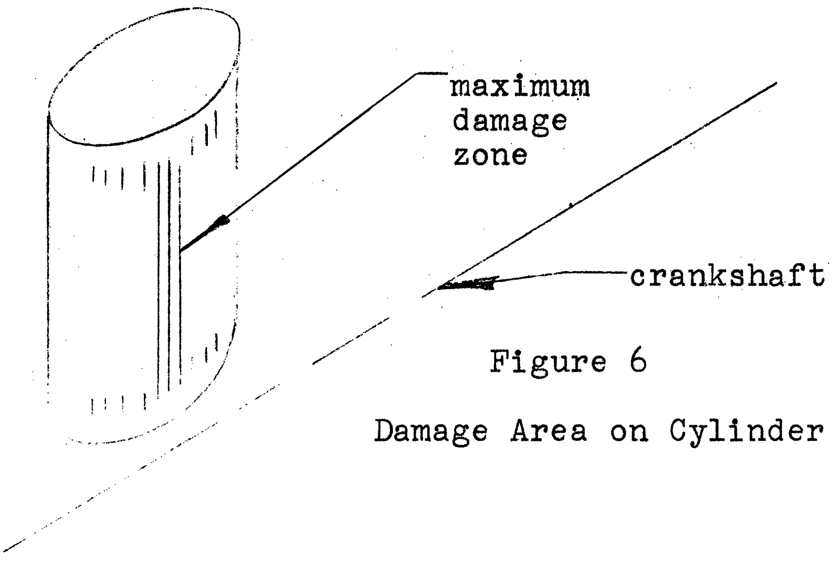


Figure 6

Damage Area on Cylinder Wal

3. REMEDIES

The important question is what can be done to reduce cavitation damage. The first thing to do where cavitation is a problem is to increase the hardness of the troubled part. Generally, no other change is as cheap, easy, or as effective as increasing the hardness of the part. In most cases, the most effective material we have at commerical prices is stainless steel. This is the final answer to many problems. The work hardening steels are especially effective since they are able to resist the pounding given to them.

Since cavitation damage is a surface phenomenon, plating has been used in many cases. Chromium and chrome-nickle compounds are two plating materials which have been rather successful. Care must be taken, however, to get a good bond between the plate and the parent metal.

An interesting exception to this general hardness rule is molybdenum. It is neither particularly hard nor strong. However, when loaded in shock, it is very slow to yield. This slow yielding quality makes this metal able to withstand the intense, but short-lived, pressure impulses of cavitation equally as well as the strongest steels.

Table 1 lists a number of important cavitation resistant materials determined by field experience and laboratory tests.

The second thing to do when cavitation damage is a problem is to reduce the relative motion between the fluid and the guiding surface. This is usually accomplished by reducing the pressure head causing fluid flow or by increasing flow areas to reduce fluid velocity. The amplitude of mechanical vibrations is often reduced by increasing the rigidity of the effected parts.

From this point on, the list of remedies continues, but the results obtained become rather spotty. Chrome compound corrosion inhibitors are often tried with results covering the entire range of good, bad and indifferent. In one case, a few drops in a cooling system may solve the problem. On the other hand, a corrosion handbook says, "addition of insufficient chromate to stifle completely oxygen type corrosion of iron will cause serious intensification of the attack".

Soluble oil can be added to water with small but consistantly good results. Increasing pressure is usually an effective partial cure, but many times, the cure is worse than the disease. This can be verified by many owners of last years' cars. Increased pressure in the cooling systems has caused many radiators to leak rather badly.

Temperature can be varied if it is consistant with good machine operation. Results are not usually remarkable and can be good or bad from one case to the next.

In some cases, air has been added to the system with the result that it "cushioned" the collapsing bubbles and reduced the pressure.

Based upon field experience:

1. Stellite
2. 17-7 Cr-Ni stainless-steel weld
3. 18-8 Cr-Ni stainless-steel weld
4. Ampco No. 10 weld
5. 25-20 Cr-Ni weld
6. Eutectic-Nyron 2-24 weld
7. Ampco bronze castings
8. 18-8 Cr-Ni cast stainless
9. Nickel-aluminum bronze, case
10. 13% cr cast
11. Manganese bronze, cast
12. 18-8 stainless spray metallizing
13. Cast steel
14. Bronze
15. Rubber
16. Cast iron
17. Aluminum

Based upon laboratory tests:

1. Stellite
2. Two layers 17-7 Cr-Ni stainless-steel weld
3. 18-8 Cr-Ni stainless-steel weld
4. Ampco No. 10 weld
5. Cast Ampco No. 18 bronze
6. Nickel--Aluminum bronze
7. 18-8 Cr-Ni cast stainless
8. 13% Cr, cast stainless
9. Manganese bronze, cast
10. Cast steel
11. Bronze
12. Cast iron
13. Sprayed stainless 18-8 Cr-Ni
14. Rubber
15. Aluminum

Table 1

Order of Cavitation Resistance of Metals

4. CAVITATION IN DIESEL ENGINES

The water side of diesel engine cylinder liners is one place where cavitation damage is a big problem. Engine vibration causes cavitation bubbles to form on the water side of the cooling jackets and in particular on the cylinder liners. These bubbles collapse and cause pits which have been known to eat completely through the block and allow the cooling water to leak away.

Several things are common to damage in every diesel engine.

1. Corrosion occurs on a area at 90° to the line of the crank shaft, as shown in Figure 6.
2. Incipient damage occurs on a small band 3/4" to 1 1/2 inches wide. This band is always on the thrust side of the liner. This is the side on which the piston bears during the power stroke (See Figure 6)
3. Corrosion never occurred where any kind of scale or coating had formed. This refers to a scale which might be formed by a corrosion inhibitor such as the chromium compounds.
4. The corroded spot was never the hottest spot, but was where the cooling water stream impinged, increasing relative velocities and preventing scale formation.
5. No cylinder was subject to attack more than any other.

Damage Prevention in Diesels

The first approach to reducing damage was to harden the parts affected. This was done by using new materials or by coating the parts with chrome. This was generally effective if a good bond was obtained and if the base parts were rigid enough to prevent bending which would cause cracking in the plating material.

Redesign of the parts was found beneficial in several ways. Parts were made more rigid to reduce amplitude of vibration. Water passages were made to avoid local areas of high water velocity. The wrist pin was made slightly off center and the piston skirts were made slightly longer. One interesting thing which was found was that if the piston was made to fit better in the cylinder, damage was greatly reduced, as a result of reduced piston slap or vibration. The present day production model of this close-fitting piston is said to be "cam ground". The pistons are made ellipse-shaped rather than round. The major axis of the ellipse is put at 90^{o} to the line of the crankshaft to reduce lateral vibrations and therefore thrust side cavitation damage.

Adding air or soluble oil to the cooling system are also methods which get results consistantly. Again, this nearly exhausts the easily and commonly applicable methods which brings results.

As mentioned before, inhibitors, temperature, pressure, etc. can be experimented with. In some cases, they will bring a complete solution to the problem, while in other cases, they might be detrimental.

BIBLIOGRAPHY

1. Beeching, R., "Selecting Alloys to Resist Cavitation Erosion," Product Engineering, January 1948, pp. 110-113.
2. Crewdson, E., "Cavitation," Engineer, January 23, 1953, pp. 122-3.
3. Eisenberg, P., "On the Mechanism and Prevention of Cavitation," Report No. 712, David W. Taylor Model Basin, Washington, D. C.
4. Euler, Leonhard, "Theories plus complete des machines qui sont mises in mouvement par la reaction de l'eau," (More complete theory of machines driven by Hydraulic Reaction), Historie de L'Academie Royal, Berlin, 1754.
5. Knapp, R. T., and Hollander, A., "Laboratory Investigations of the Mechanism of Cavitation," Trans. ASME, July 1948, Paper No. 47-A-150.
6. Knapp, R. T., "Cavitation Mechanics and Its Relation to the Design of Hydraulic Equipment," James Clayton Lecture, Proc. Of the Institution of Mechanical Engineers, series A, Vol. 166, 1952, pp. 160-163.
7. Poulter, T. C., "Mechanism of Cavitation-Erosion," Journal of Applied Mechanics, March, 1942.
8. Knapp, R. T., "Recent Investigations of the Mechanics of Cavitation and Cavitation Damage," ASME, Paper No., 54-A-106, Presented November-December 3, 1954.
9. Rayleigh, Lord, "On the Pressure Developed in a Liquid During the Collapse of a Spherical Cavity," Philosophical Magazine, series 4, Vol. 34, 1917, pp. 94-98.
10. Sutton, G. W., "A Photoelastic Study of Strain Waves Caused by Cavitation," ASME, Paper No. 57-APM-15, July 24, 1956.
11. Taylor, Irving, "Cavitation-Pitting by Instantaneous Chemical Action from Impacts," ASME, Paper No. 54-A-109, presented November 28-December 3, 1954.
12. Trock, Bernard, "A Study of Cavitation Erosion," SAE, presented June 3-8, 1956.
13. Blackwood, A. K., "Coolant side corrosion Diesel Cylinder Sleeves--Means for reducing", SAE, Paper 761, Presented June 3-8, 1956.
14. Eisenberg, Phillip, "A brief survey of progress on the mechanics of cavitation", Navy Dept., DTMB, Washington, D. C., Report 842, 1953.
15. Joyner, J. A., "Reduction of cavitation pitting of Diesel Engine Cylinder liners", SAE, Paper 760, Presented June 3-8,1956.

16. Margulis, W., McGown, J. A., and Leith, W. C., "Cavitation control through Diesel engine water treatment", SAE, Paper 757, Presented June 3-8, 1956.

17. Plesset, M. S., and Ellis, A.T., "On the mechanism of Cavitation Damage", Trans. ASME., 1955, 77:1055-1064.

18. Schrader, A. R., "Investigation of Cavitation Erosion in Diesel Engine Collant Systems at the U. S. Naval Engineering Experiment Station", SAE, Paper 759, Presented June 3-8, 1956.

CHAPTER 13

FRETTAGE CORROSION

1. INTRODUCTION

Since the first theory of fretting corrosion was published by the Englishman Tomlinson[1] in 1927, there has been an increased amount of interest in this subject. The phenomenon has been described by a variety of descriptive names such as false brinelling, wear oxidation, chafing fatigue, friction oxidation and, more recently, fretting or fretting corrosion. The variety of names is somewhat a measure of the uncertainties surrounding this phenomenon.

Fretting is recognized as a type of wear caused by two surfaces rubbing together under the conditons of reciprocating motion of very small amplitude. Fretting corrosion is a special case of fretting where the surfaces, or particles from the surfaces, react with the environment.

Fretting damage leads to fatigue failure. This was first recorded in 1911 when Eden, Rose, and Cunningham[12] reported that failures of fatigue specimens occured in the grips of their machines and observed the phenomenon of intersurface damage associated with these failures.

Fretting can be placed in a number of different classifications of wear depending on which of the proposed mechanisms is believed to be correct. Fretting can be classified under scoring, abrasion, or corrosion. Depending on the conditions, and one of the above three, or combinations of them, can be occuring during fretting.

2. MECHANISM OF FRETTING CORROSION

There are three main theories on the mechanism of fretting and fretting corrosion.

Adhesion Theory

According to Bailey and Godfrey asperities come into metal-to-metal contact due to the external pressure of the load. Two types of junctions are possible. A strong junction which will lead to metal transfer or a weak junction which will give rise to a loose wear particle.

Both types of junctions are formed due to the presence of a thin oxide film or adsorbed gases. The oxide film will reduce the number of strong junctions because the gas will be adsorbed on the surface and thereby reduce the adhesion of the transfered fragment. If the surfaces were cleaned of any oxide or other surface film and the experiment run under extremely high vacuum conditions (on the order of 10^{-11} mm of Hg.) only strong junctions would be formed.

They concluded from their experiments that most of the junctions were of the "strong" type which resulted in metal transfer. This metal transfer caused the surfaces to be plowed resulting in the removal of finely divided and apparently virgin material. There particles oxidize and are supposed to play no part in causing wear.

Abrasion Theory

Wright[2] and Feng and Rightmire[3] base their theory on the postulate that plastic deformation will cause the asperities to interlock on a microscopic scale. The tangential force necessary to break this junction will result in the shearing off of one of the peaks rather than shearing along the original interface. Their reason for this is that the interface is strain hardened and strengthened which causes the breakage to occur in the weaker section of the asperity further from the interface. There is a fallacy in their reasoning because before fracture can occur on any plane, that plane must be strained resulting in strain hardening and strengthening. It is possible however, that the stresses might be highest at a point back from the interface causing failure at that point.

They continue by saying that small wear particles, after they are sheared off, will accumulate in the spaces between the high spots and become oxidized. The wear process then becomes one of abrasion. Pits will form because of the stronger abrasion action in the central region between the high spots. As these pits fill up, the oxide particles spill into adjoining depressed regions until the entire region becomes covered by the oxidized material. The pits tend to become fewer in number but larger and deeper in size. (See Figure 1)

Corrosion Theory

The third theory, that proposed by Uhlig and his co-workers,[6-8] is one that proposes that fretting corrosion is a combined chemical and mechanical phenomenon. Uhlig has summarized his

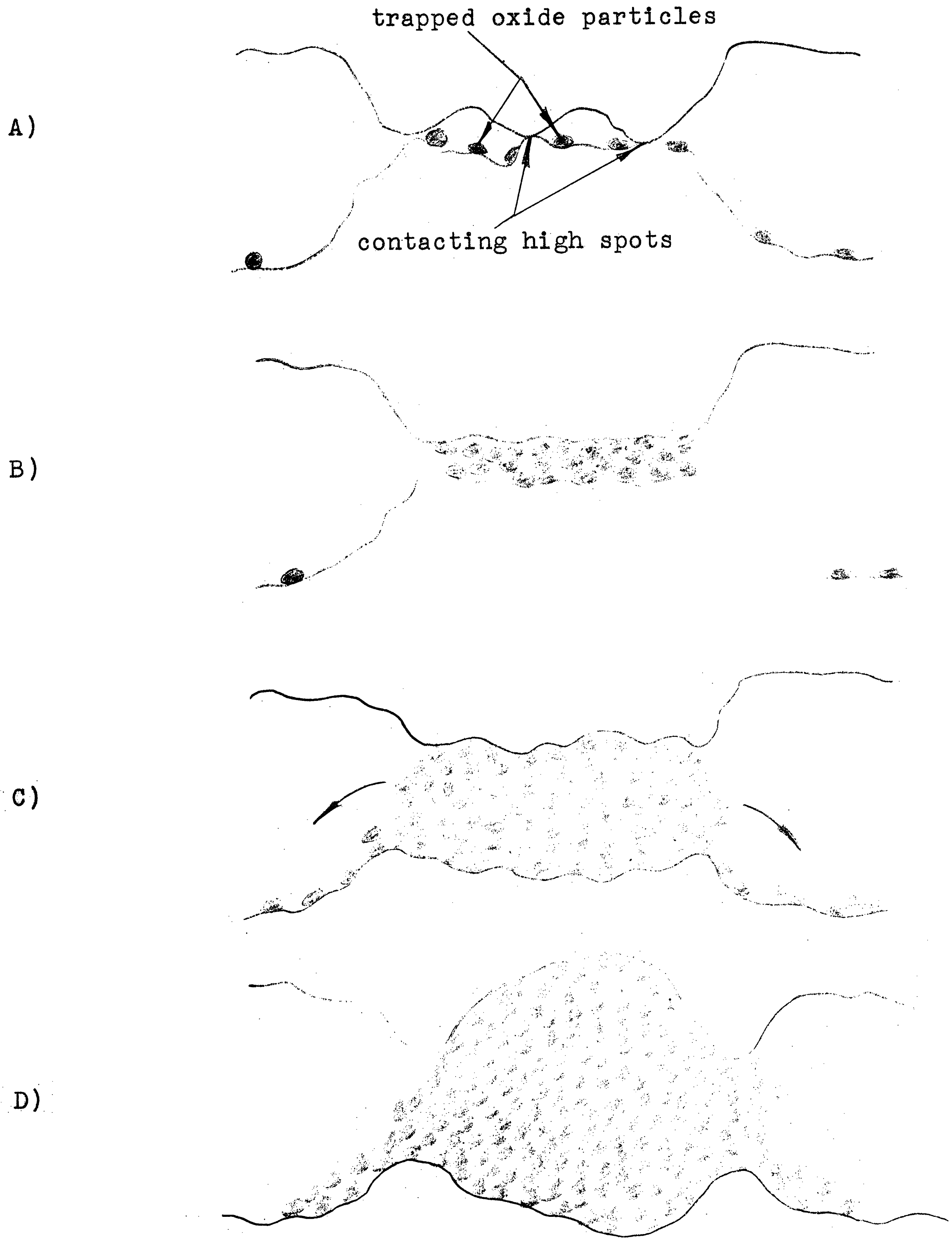


Figure 1

Mechanism of Wear:

A) Loose wear particles oxidize; B) increase in number; C) spill into adjoining depressed regions; D) Form large and deep pits.

theory as follows: (See also Figure 3)

> "An asperity rubbing on a metal surface is considered to produce a track of clean metal which immediately oxidizes, or upon which gas rapidly adsorbs. The next asperity wipes off the oxide or initiates reaction of the metal with adsorbed gas to form oxide. This is the so-called chemical factor of fretting. In addition, asperities dig below the surface to cause a certain amount of wear by welding or shearing action in which metal particles are dislodged. This is the mechanical factor of fretting."

An evaluation of the three theories indicates that there is one main problem that has not been completely solved. The problem is when does the oxidation occur? Does it occur after particles are torn from the surface or does it occur on the surface and is then rubbed off?

Uhlig presents data that shows that the chemical factor in fretting varies from six to 78% of the damage. He also shows that fretting varies with frequency of oscillation. At low values of frequency the chemical factor is more important, however as the frequency increases the mechanical factor increases in importance. (See Figure 2) His data is also consistent with other well-established facts on fretting.

However, many of the other researchers have relegated the chemical factor either to a very minor role or to have no effect at all. Uhlig's data is either not accounted for by the other researchers or represents new contributions to the field. It is also possible that oxidation may occur either way, depending on some of the factors involved such as load, slip, or materials.

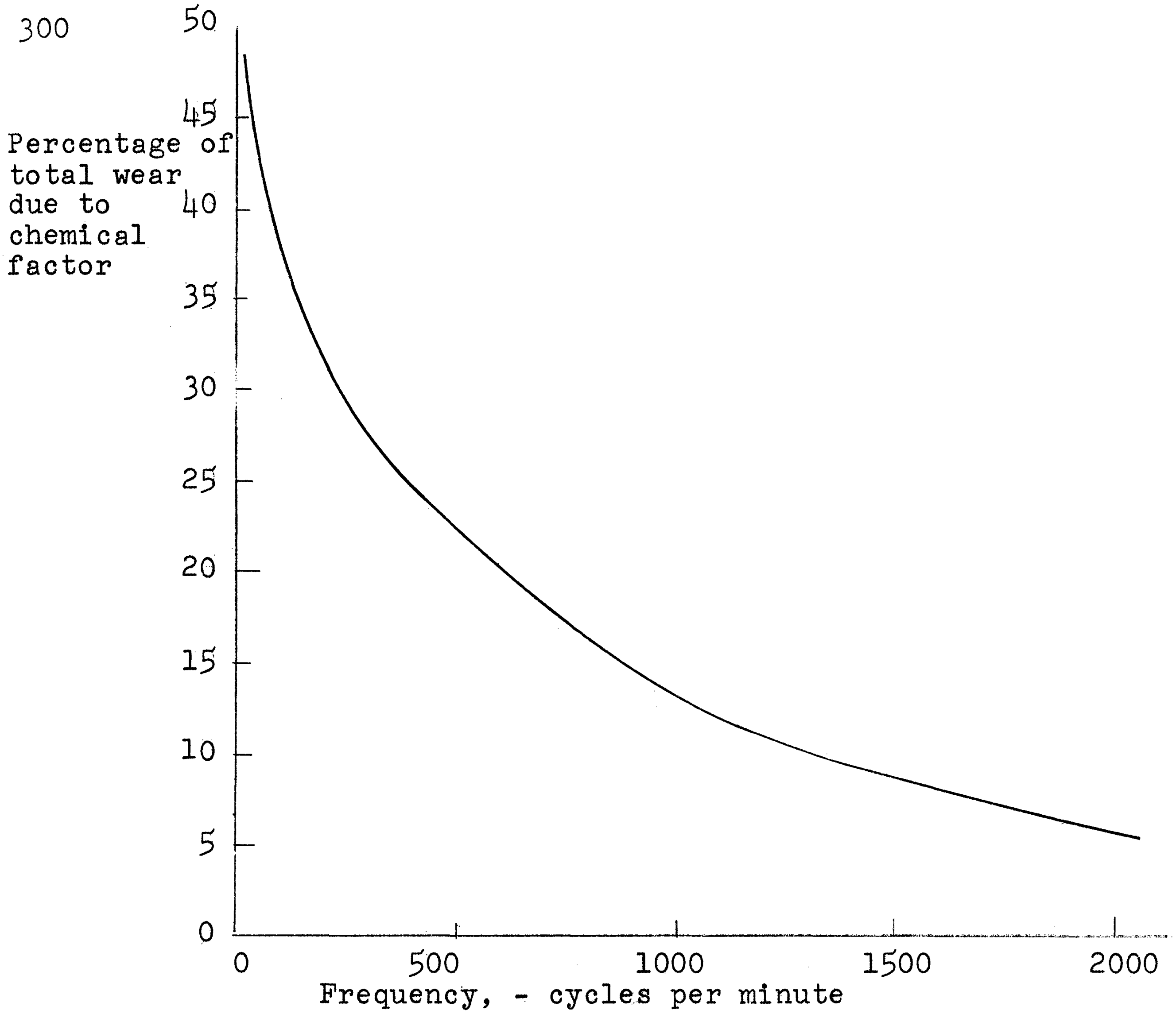


Figure 2

The effect of frequency of fretting damage of mild steel. Total test of 67,800 cycles at a pressure of 5300 psi and a slip of 0.0036 inches.

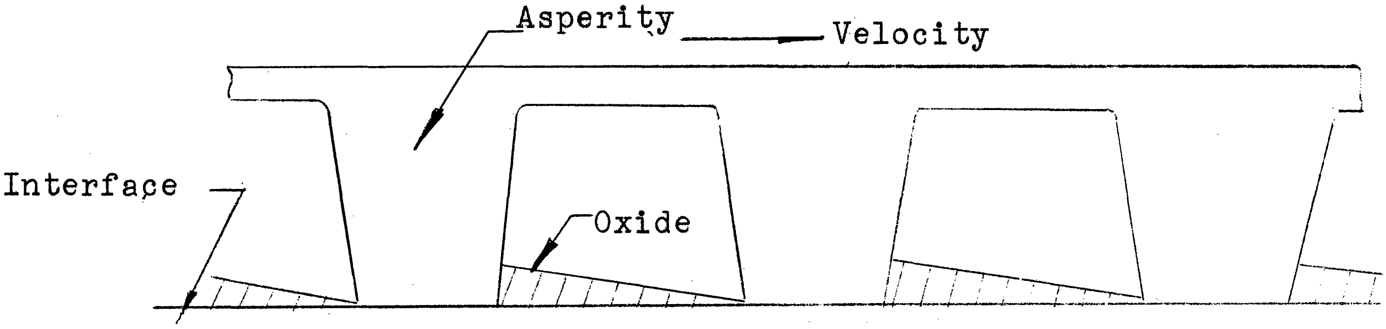


Figure 3

Idealized model of fretting at a metallic surface. Asperities of upper surface move over plane underside removing the oxide layer.

Uhlig states his position and seems to clear up some of the confusion when he says:

> "The fact that fretting corrosion involves a chemical process can hardly be doubted in view of the fact that a chemical-reaction product is formed. The process of fretting, therefore, is truly one of corrosion - a term defined as the chemical or electro-chemical reaction of a metal with its environment. Although the terminology is reasonably clear, there is still ample room for a discussion as to whether the observed chemical process accompanies or follows consumption of metal in fretting damage."

3. FACTORS THAT INFLUENCE FRETTING CORROSION

Slip

Before any fretting will occur there must be relative motion between the two parts. The motion required is in the order of 10^{-7} inches,[2] however this figure is thought to be rather low and .0004 inches is thought to be the value at which fretting will occur.[4] Usually fretting tends to increase with the slip and this can be observed by looking at Figure 4. This graph shows a direct relationship between the two, but at higher amplitudes mechanical wear may start. The amplitude generally encountered in actual service is between .0004 and .009 inches. If the amplitude is any larger grease may be dragged into the fretting area and fretting corrosion will be retarded or stopped.

Frequency

The frequency of oscillation appears to have no effect on small amplitudes of .0004 inches or in this area. However, at higher amplitudes the frequency does have a definite effect upon fretting.

As strange as it may seem the lower the frequency, about 56 cpm, the greater the fretting rate. Figure 5, shows how weight loss is affected by the frequency and at 2000cpm or more the weight loss is seem to remain practically constant. The reason for this is that it takes time for a chemical reaction to occur and when the parts are moving at a fast rate relative to each other, the cleaned surface does not have time to oxidize and the reduced weight loss is due to the plowing effect.[4]

Number of Cycles

It is usually accepted that fretting will increase with the number of cycles. Figure 6, shows the weight loss and number of cycles are directly related for this particular material.[5] Also there is a run in period when very little weight loss occurs and this is during the period of metal transfer. Then when the abrasive action starts the weight loss goes up very rapidly.

Normal Load

This is an area of debate as to what affect the load has on fretting corrosion. Perhaps the main reason for this is that the amplitude of vibration and type of mechanism may give very different results. The reason this is true is because at small loads the vibration is not affected but at higher

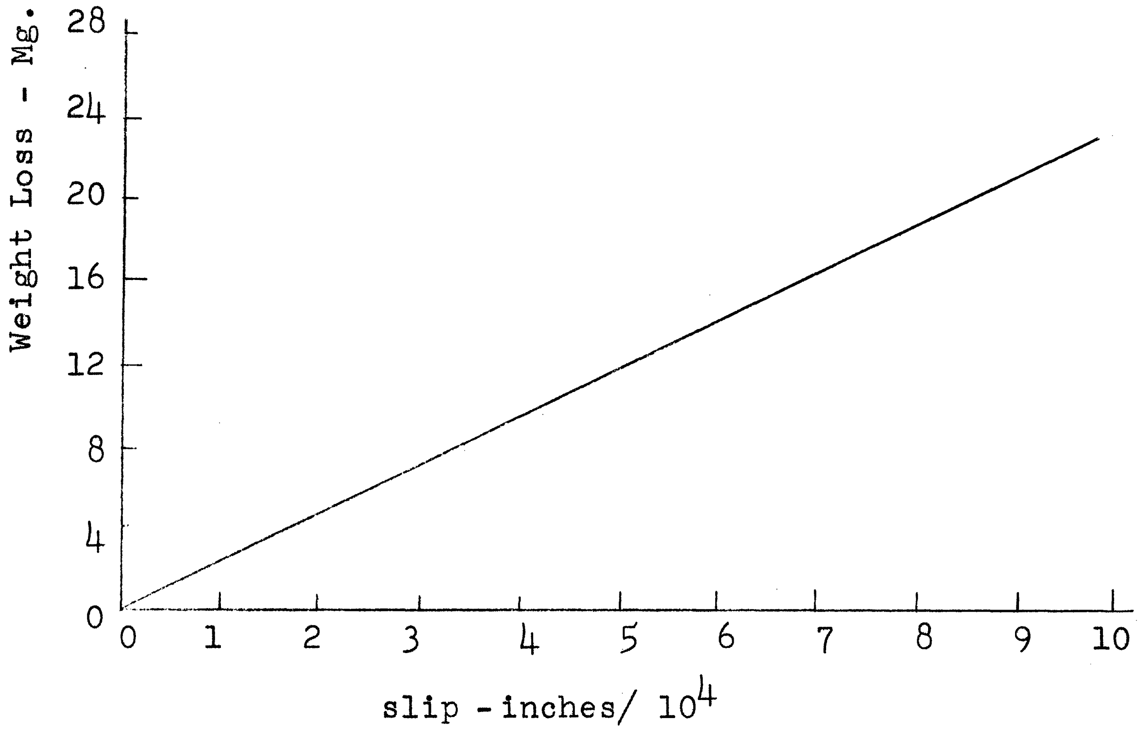


Figure 4

Influence of Slip on Frettage

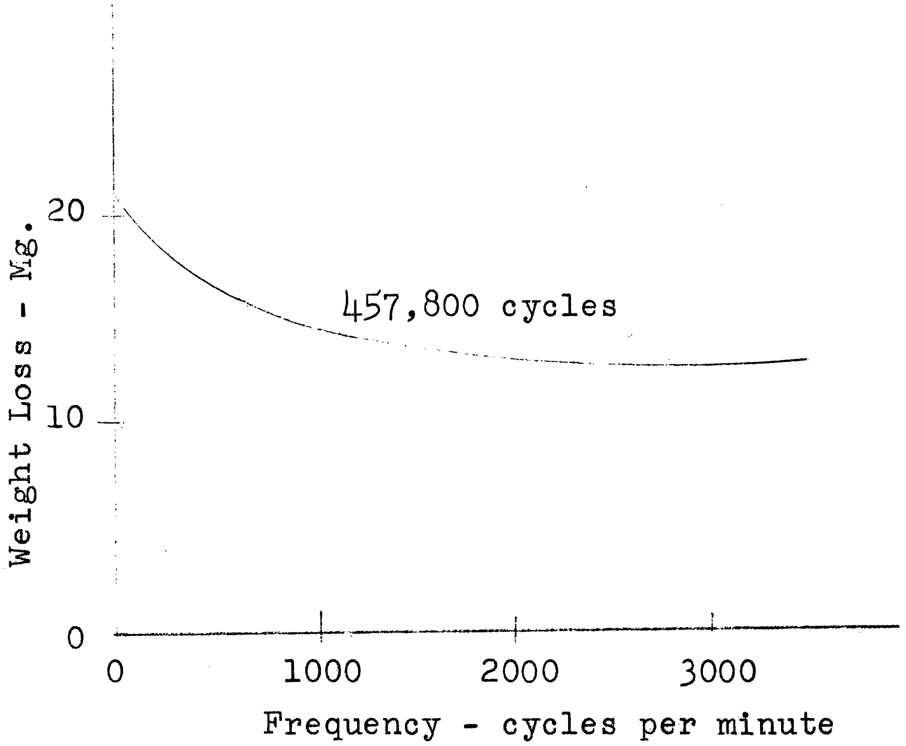


Figure 5

Effect of Frequency on Frettage

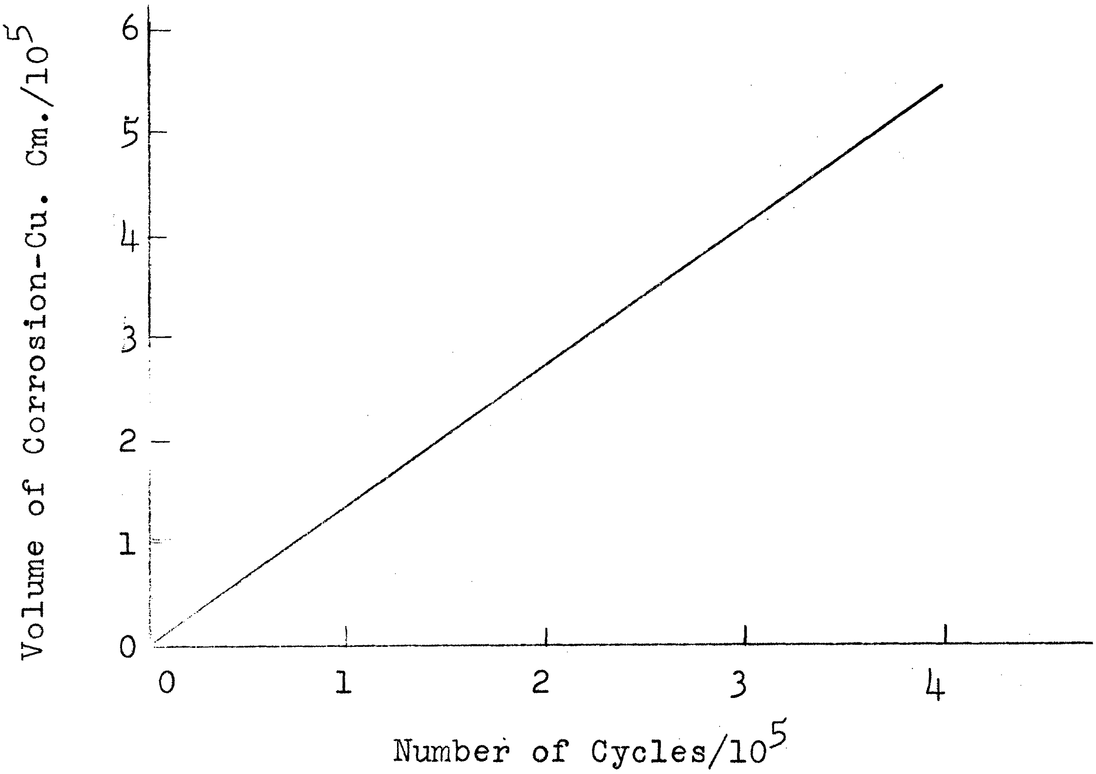


Figure 6

Effect of Number of Cycles on Frettage

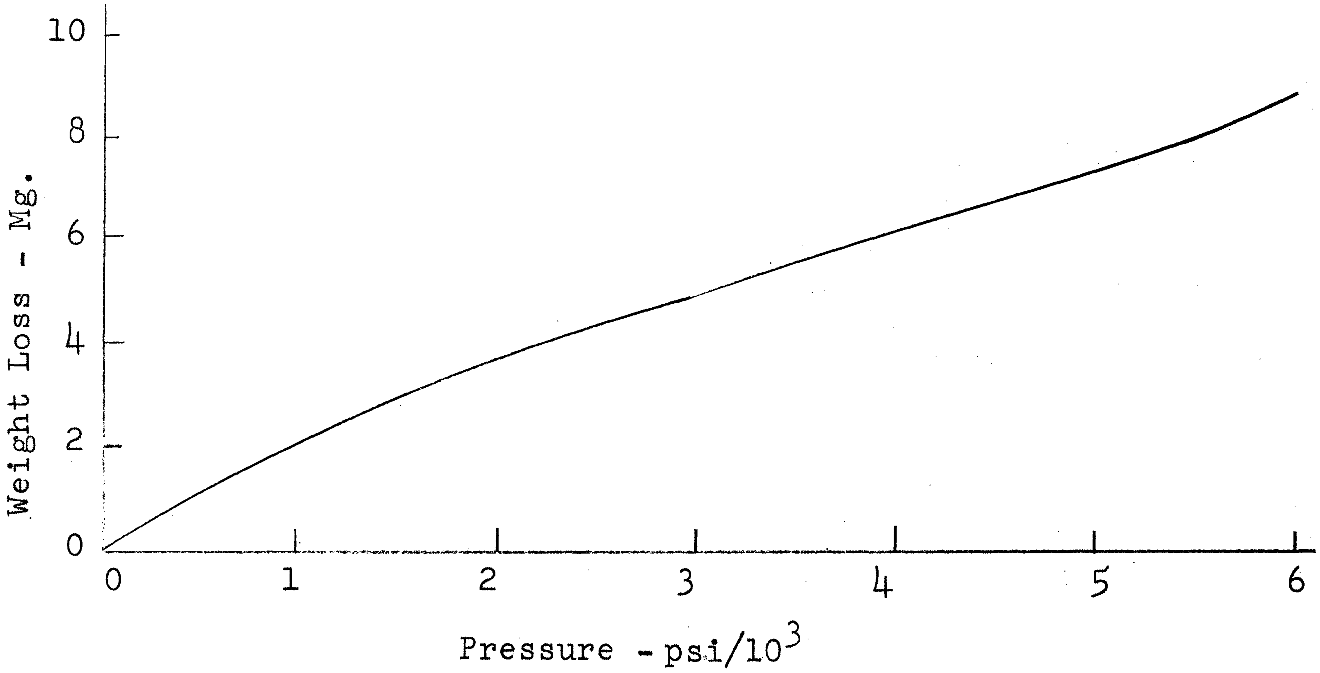


Figure 7
Influence of Pressure on Frettage

loads the vibrations may be stopped or the contact area may be increased and thus more slip is caused resulting in increased fretting.

Figure 7, shows that with an increase in load the weight loss is also increased. This is though to be true if the load applied is kept below a certain value.

Effect of Temperature

It was found the fretting corrosion occurred more in the winter than in the summer and in an experiment done by Uhlig and Feng[6] between -150^{o} and $+ 150^{o}C$. it was found that the greatest damage occurred at $-150^{o}C$. and decreased gradually as the temperature rose to $0^{o}C$. (See Figure 8) At $0^{o}C$. it showed a further sudden decrease. At temperatures above $50^{o}C$. damage remained constant but increased again at higher temperatures.

The sudden decrease at $0^{o}C$. has sometimes been connected to the transition from brittle to ductile fractures, however, the suggestion that water forms a lubricating film after $0^{o}C$. sounds more reasonable.

It should be noted that Uhlig and Feng in their experiments had to use temperature corrections because the value of slip was so high. This makes their experiment less valuable, especially since the temperature compensations were quite large.

Effect of Humidity

The influence of humidity to the amount of damage is import-

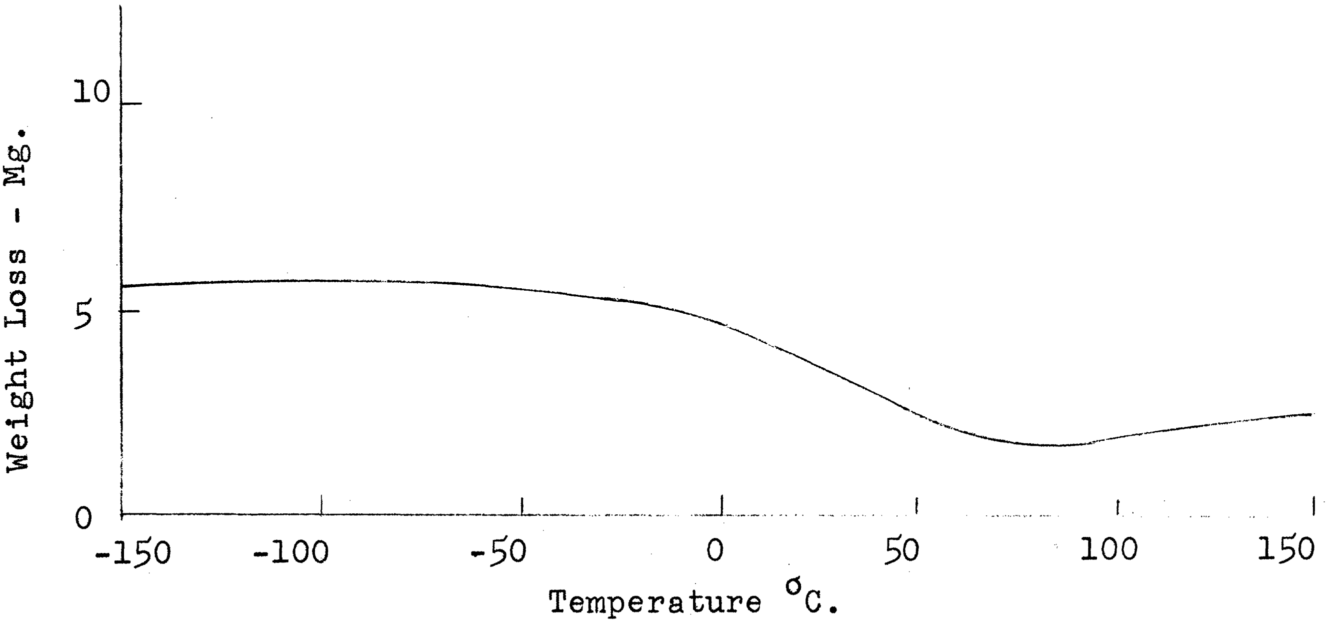


Figure 8

Effect of Temperature on Frettage

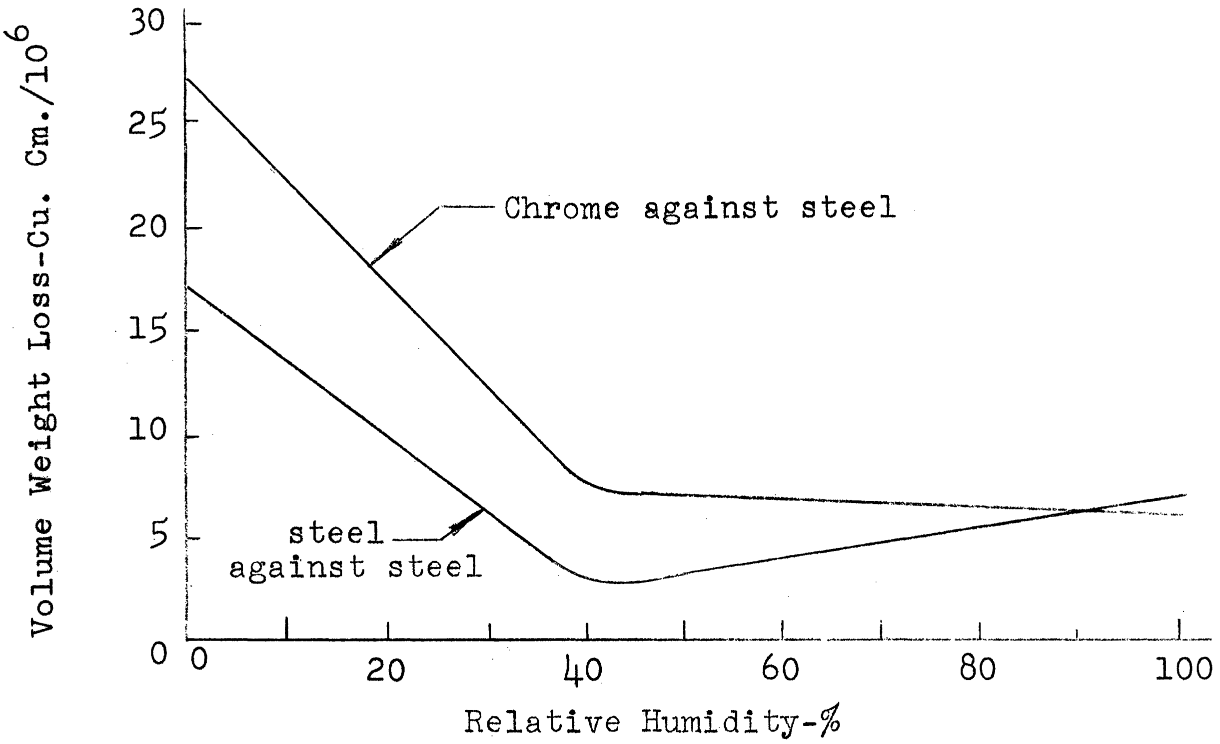


Figure 9

Effect of Humidity on Frettage

tant. The degree to which it influences the damage differs, however, as to the source of information.

All agree as humidity increases for 0% to 50%, wear decreases a substantial amount. From 50% to 100% relative humidity, wear will increase slightly in some cases e.g. steel against steel but decreases more in other cases e.g. chrome against steel. (See Figure 9)

Humidity seems to prove a lubricating film of water as indicated in the previous topic, Temperature Effect. Another possibility is that hydrates $Fe_2O_3 \cdot H_2O$ are formed by the water instead of Fe_2O_3, which is softer, thus less abrasive.

It should be noted that humidity didn't have any effect when using a Nitrogen atmosphere, however, no explanation has been given for this phenomena.

Effect of Lubrication

The role lubricants play in preventing fretting corrosion is mainly to separate the surfaces, to exclude the atmosphere, and to remove the debris.

Since the pressure between the surfaces involved in fretting corrosion usually are very high, it will be difficult to obtain a complete separation of the surfaces by means of oil and grease lubrication. As a rule, only boundary lubrication can be achieved. This does not prevent metal to metal contact completely. Fretting corrosion, by means of welding, will thus till occur even with lubrication but the rate of corrosion will be smaller when a lubricant is used.

Experimental evidence indicates that damage in vacuum or in a protective atmosphere is much less severe than in a space open to the atmosphere. Excluding the atmosphere can partly be obtained by flooding the area with a lubricant in which oxygen has a low solubility and diffusivity. Rolfe prevented fretting corrosion in practice by supplying more channels for the lubricants to the fretted surfaces.

The same results are obtained by using a rubber gasket. This will mould to the surface and prevent the access of air.

After fretting has started, the lubricant might act as a dispersing agent for the debris. This can be either harmful, when the debris is abrasive, or beneficial in the case where debris can provide a cushion between the surfaces.

Of the solid-film lubricants, MoS_2 is the most important. It possesses the following desirable properties: high load-carrying capacity and capability of minimizing metallic adhesion; tenacity for steel and low coefficient of friction; and thermal stability.

As investigation at the NACA Lewis Laboratory on the effect of MoS_2 in preventing fretting corrosion showed very good results when using a coating of dry MoS_2. This was bounded to steel by rubbing a mixture of MoS_2 and syrup into intimate contact with the clean steel at elevated temperatures. This delayed the start of fretting by about 10,000,000 cycles. Dry MoS_2 and mixtures of MoS_2 with carriers were less preventive. (See Table 1)

Method	Surface Condition	Average number of cycles for first evidence of fretting corrosion	Observation of contact area during and after experiments
		Steel ball against glass flat	
-	Clean	1-30	Instantaneous fretting corrosion
1	MoS_2 dusted	72,000	MoS_2 formed smooth bearing surface then rubbed away
2	MoS_2 rubbed	21,600	MoS_2 smeared thinly then rubbed away
3	MoS_2 water, and aerosol	21,600	MoS_2 visible only in surrounding area
4	MoS_2 and oil	86,400	MoS_2 visible only in surrounding area
5	MoS_2 and grease	216,000	MoS_2 visilbe only in surrounding area
6	Bonded MoS_2	28,000,000	Coating formed smooth bearing surface then brown stain appeared
		Steel flat against steel flat	
-	Clean	100	Immediate fretting corrosion
1	MoS_2 dusted	100,000-160,000	Dry red Oxide and stain
4	MoS_2 and oil	700,000-760,000	Brown-black debris and spotty surface failure and stain
5	MoS_2 and grease	1,560,000-1,620,000	Brown-black debris and spotty surface failure and red stain
6	Bonded MoS_2	9,823,000-9,883,000	Smooth shiny bearing areas formed then brown powder appeared.

Table 1

Results of Fretting Corrosion Investigation Using Molybdenum Disulfide MoS_2 as Inhibitor

Effect of Oxygen

While oxygen is favorable in prevention of common wear by forming a hard wear resistant oxide layer, its role in fretting corrosion is not yet clearly understood. However, it has been found that oxide formation does not form a protection but instead it increases wear in fretting. In most cases of fretting corrosion, wear between vibrating surfaces is of such a severe nature that any oxide films are worn off even with those metals on which, under less severe conditions of wear, stable oxide films would be formed. The worn off debris which are trapped between the vibrating surfaces will act as an abrasive since most oxides are extremely hard. The severity depends of course on the hardness of the oxides and the parent materials. The rate of wear might be reduced if one of the parent materials is soft and the oxides are embedded in this soft material.

It has been found that fretting corrosion also occurs on non-metallic, non-oxidizing surfaces. This relegates oxidation as a cause to that of a secondary factor and indicates that finely divided and apparently virgin material is first produced. This is confirmed by the fact that the first symptoms of wear showed up much later when a vacuum or N_2 atmosphere was used instead of air.

Surface Finish and Closeness of Fit

Generally speaking fretting appears to progress more rapidly in parts that have high surface finish and closely shrunk fits.[1] Perhaps the main reason for this is due to the fact that the lubricant can't penetrate into the area of wear and prevent fretting. It is well established that machine parts with a rough finish retain the lubricant in the small pockets of the rubbing surfaces, thus retarding fretting corrosion. However, if the surface is made too rough and the closeness of fit is too loose then wear will occur in the parts rather than fretting corrosion.

Hardness

In most cases it appears that materials with a higher hardness are less susceptible to attack by fretting. Using materials with a high hardness is beneficial in two ways.[4]

1. The welds formed from metal-to-metal contact will break at the joint, because the weld is weaker than either of the two metals and less debris is formed.
2. The particles formed will be of a small size and not cause a great deal of damage to the hard surface.

At present there is not a great deal of information on this subject but one thing that is known is that stainless steel, a hard metal, will fret very rapidly. None of the literature seems to give any information because there appears to be very little quantitative work in this area.

4. THE RELATIONSHIP BETWEEN FRETTING CORROSION AND FATIGUE STRENGTH

We now shift our attention to investigating the effect of the reduction of fatigue strength due to fretting corrosion. Most of the experiments involve the fatigue testing of press fitted assemblies.

Horger[8,9] has done a considerable amount of work in this field. He has found that shallow cracks occurred in the press fitted region at stresses far below the expected endurance limit of the material. He states that the cracks were initiated at stress levels as low as 5000 psi. The question does arise if actually the stress level is much higher due to some other unknown factors. The cracks that were formed in this manner did not propagate to failure if the environment producing the fretting was removed. Only if relatively high stresses (the value depending on the strength of the material) are imposed on the specimen will the cracks propagate to complete failure. This contradicts the popular notion that the stress necessary to produce a carck will cause the crack to propagate to failure.

It is quite well agreed that there is a definite correlation between the types of fit (i.e. shrink or press fits) and the resulting reduction in the fatigue strength. The fatigue strength decreases as the pressure of the fit increases. (See Figures 10 & 11) Figure 10 shows a decrease in the endurance limit of 55% due to fretting. However there seems to be no correlation as to how much the fatigue strength is reduced in fretting in a general case because as seen in Figure 11 fretting is also a function of the clamping pressure.

The problem in determining the effect that fretting corrosion has on the fatiuge strength is difficult because of four factors which operate to cause failure by fatigue. They are:

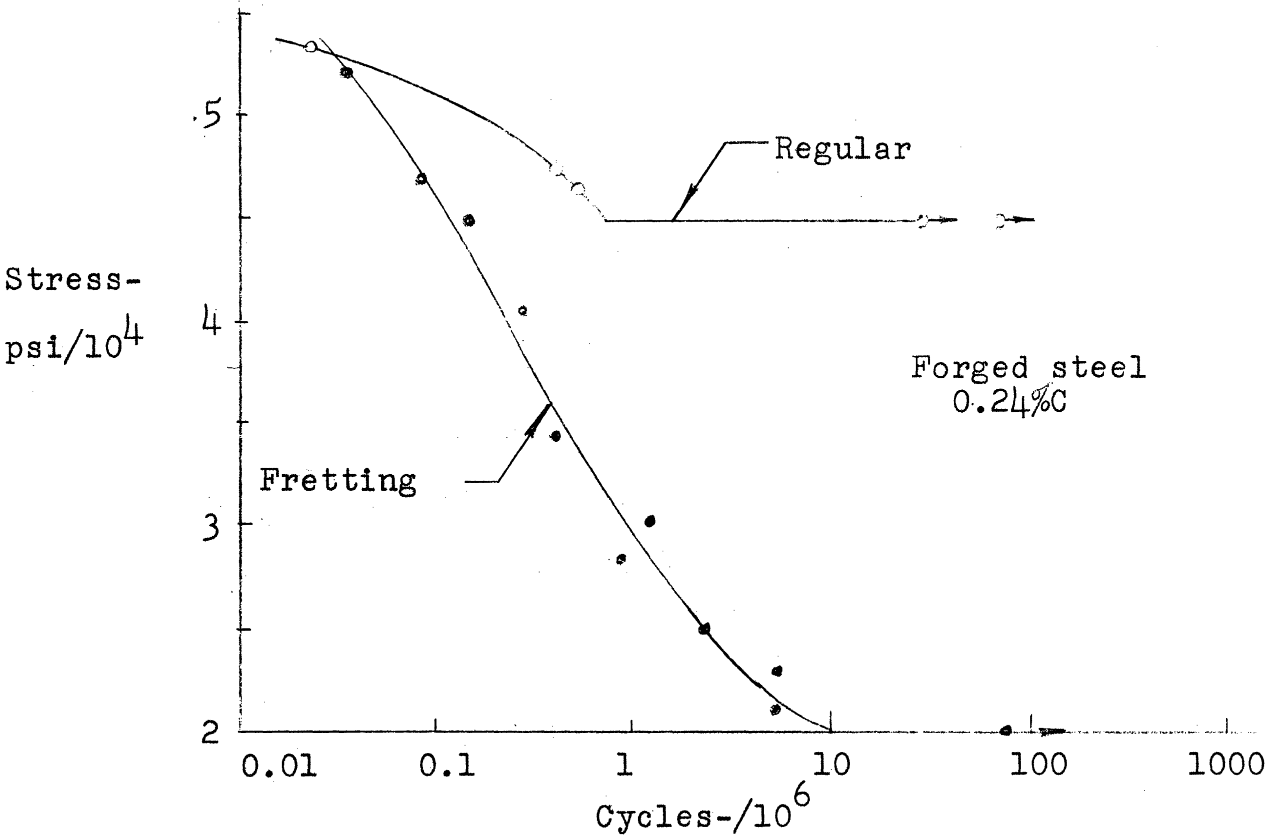


Figure 10

S-N Curves for Tests on Fretting Fatigue and Regular Fatigue of a Forged Steel.

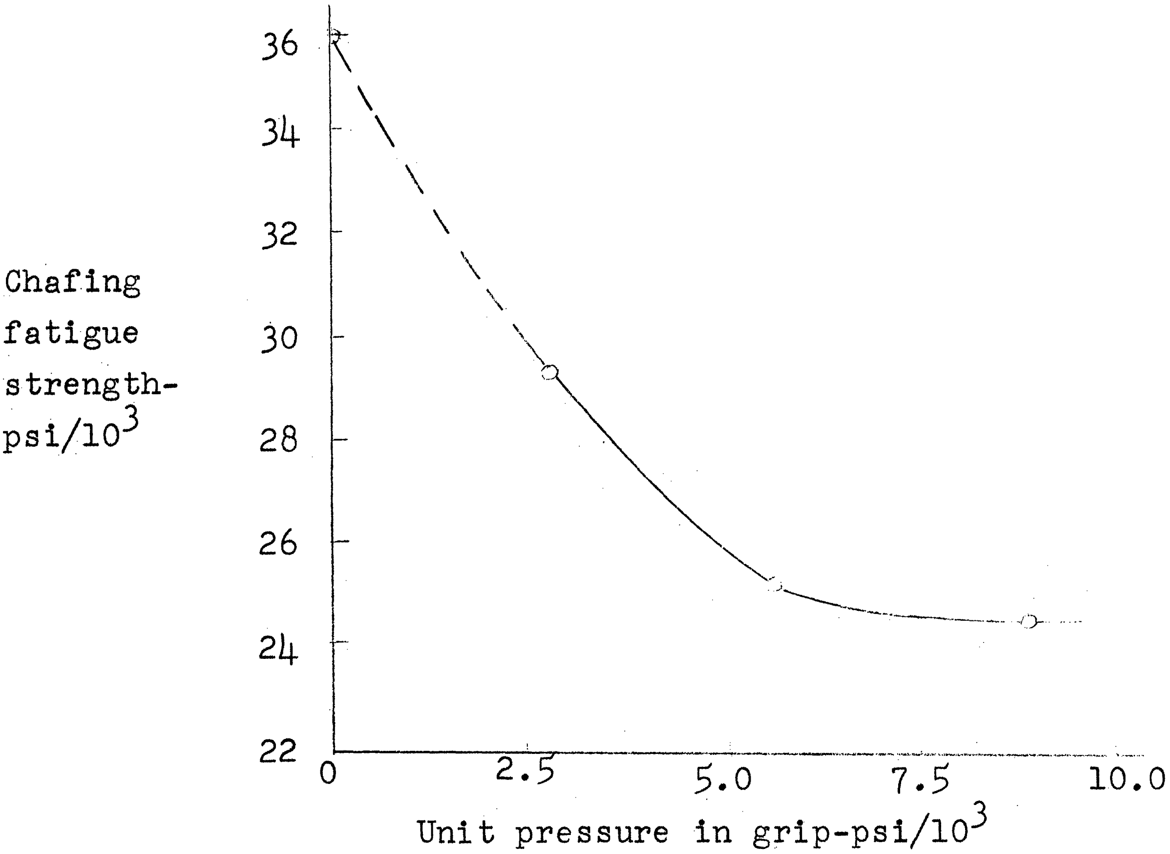


Figure 11

The Effect of Clamping Pressure on the Fatigue Strength of Cantilever-Type Steel Specimens.

the normal fluctuating stress; the unknown disturbance of the fluctuating stress due to fixed clamping stresses; the unknown alternating tangential stress due to the fitted parts being subjected to vibration; and the increasing surface roughness due to fretting.

A novel experiment was preformed in England by E.J. Warlow-Davies[10] in which he tried to determine the effect of fretting corrosion on fatigue strength. His experiment involved the separation of the fretting and fatigue tests. The first step was to produce frettage on specimens in a special machine, followed by the transfer of the specimens to a standard fatigue testing machine where a bending moment was introduced in the fretted region. By this means the damaging effect of the fretting corrosion of the surface could be measured by the percentage loss in fatigue strength as compared with unfretted specimens. On other specimens the fretting damage was ground off and the fatigue strength determined of these samples.

Two materials were tested; a medium carbon steel and an alloy steel. Scatter of data was to be expected in this experiment due to the impossibility of defining the degree of fretting corrosion on the surface of the specimens.

The results of the test were as follows: The endurance limit of the medium carbon steel was reduced from 39,000 psi to 32,200 psi, ie. by 13%. However, when surfaces of the fretted specimens were ground off until no evidence of fretting remained, the resulting endurance limit was 34,000 psi indicating that the weakening was not due entirely to the fretting damage or that the fretting damage extended further than the depth of grinding.

The endurance limit of the alloy steel was reduced from 61,000 psi to 50,000 psi or approximately 18% due to the fretting corrosion. The endurance limit of the specimens on which the fretting damage had been ground off was 60,500 psi.

A dilemma arises: Is the fatigue strength reduction due only to fretting corrosion or is it due to a combination of fretting and a factor introduced by the stress involved in obtaining the fretting? The plain carbon steel seemed to validate the later while the alloy steel supported the first idea. A satisfactory answer to the dilemma was not found in the literature. Most researchers feel that specific results on the amount of reduction in fatigue strength determined in this manner can not be applied directly to applications in the fied.

An investigation was made in England by Fenner and others[11] to determine what effect fretting corrosion has on the fatigue strength. Their procedure involved determining the effect on the endurance limit of clamping pressure (which was an indirect measure of the fretting damage as the fretting damage was proportional to the load.

Their results, using an Aluminum alloy as clamps, indicate very definitely that an increase in the clamping pressure resulted in a decrease in the endurance limit. It was also noted that all of the failures originated in the region where the fretting had taken place. (See Figure 12)

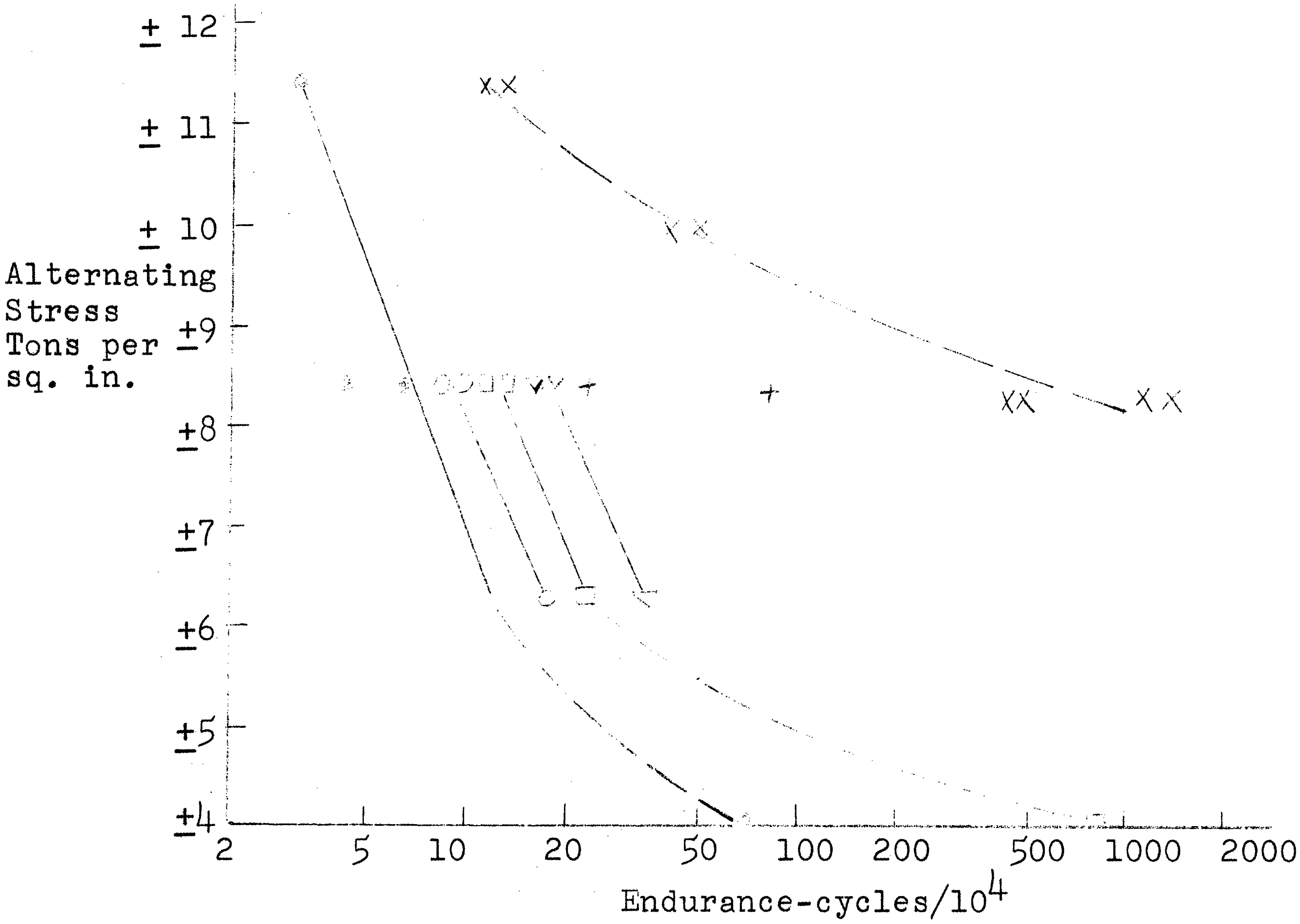


Figure 12

Fatigue tests of Aluminum alloy (L65) Test pieces with and without clamps.

× No clamping

● Clamping pressure 8 tons per sq. in.

○ Clamping pressure 8 tons per sq. in.

□ Clamping pressure 0.25 ton per sq. in.

▽ Clamping pressure 0.085 ton per sq. in.

\+ Clamping pressure 0.01 ton per sq. in.

In Figure 13 an S-N curve is plotted for the two limits of clamping pressure i.e., zero and infinite pressure. The material and size were the same as used by Horger in Ref. 9. With infinite pressure the press fit acts as a stress concentration. An approximate fatigue stress concentration factor (K_f) of 2.5 was used in the calculations.

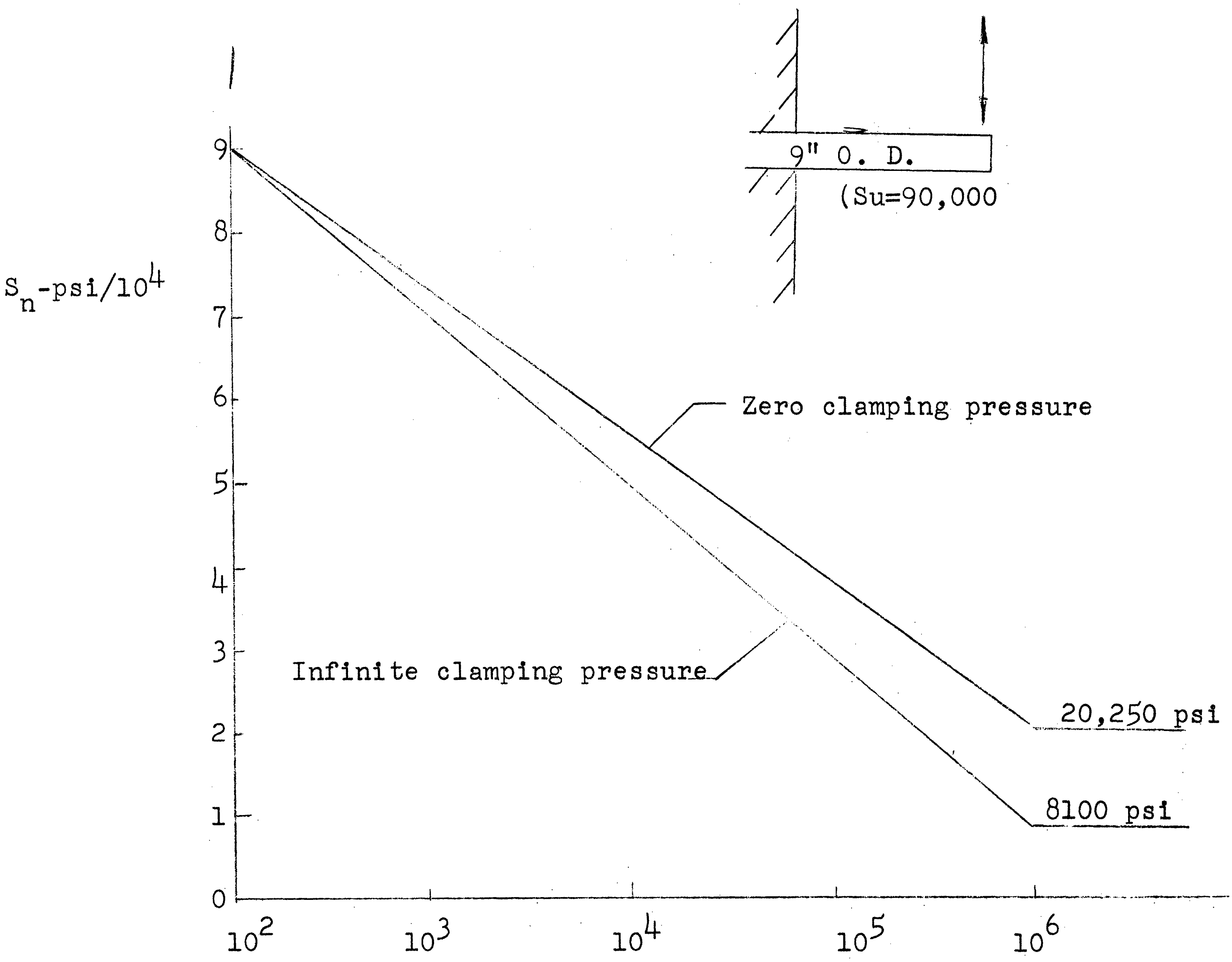
In conclusion fretting corrosion reduces the fatigue strength of a material however, the amount of reduction is a function of other factors, the most important of which is the load or clamping pressure.

5. PREVENTION OF FRETTING CORROSION

Before we discuss the actual preventive measures for fretting corrosion, it would be well to define the two situations in which fretting corrosion occurs. The first is that fretting which occurs in parts that were designed to undergo relative motion part or all the time. Examples of this situation are found in flexible couplings and bearings undergoing oscillatory motion. The second situation is where the parts are not designed to undergo any relative motion such as in shrink and press fits.

A device in which both conditions are found is the roller contact bearing. Fretting between the balls and the race is an example of the first situation while fretting between the face of the bearing and the holding surface is an example of the second situation.

There is some confusion in the literature today on the methods of preventing fretting corrosion. The reason for this



Number of cycles

Figure 13

S-N Curve Showing the Effect of Clamping Pressure on the Endurance Limit on Specimens Similar to Those Used in Reference 11.

Calculations for the above curves:

Zero clamping pressure:

$$S_n = S_u \times .5 \times (\text{size factor}) \times (\text{surface factor}) \times (\text{bending factor})$$

$$= 90{,}000 \times 0.5 \times 0.5 \times 0.9 \times 1 = 20{,}250 \text{ psi}$$

Infinite clamping pressure: $K_f = 2.5$ $S_n = 8{,}100$ psi confusion is that a remedy that is good for one of the above situations is usually not good for the other situation. There is of course some overlap in the remedies for the two situation that occur.

In the case of the flexible couplings and bearings the remedy most often used is to lubricate the parts either by liquid or solid lubricants. The operating conditions should be known in order to select the proper lubricant. A liquid lubricant, if it is going to be of any value at all, must have good "feedability" characteristics. The main function of the lubricant here is to prevent oxygen from reaching the surface and therefore the lubricant should be one in which oxygen has a low solubility and diffusivity.

Reflection on this idea leads to the conclusion that it would be best to have a lubricant that would have a solubility that would vary directly with temperature. If such a lubricant exists, it would take the detrimental oxygen away from the fretting area because of the higher temperature in this are due to the dissipation of energy. When the lubricant was not longer in the fretting region it would give up its oxygen due to its lower solubility at lower temperatures.

It was found that a liquid plus an additive such as Molybdenum disulfide (MoS_2) or other extreme pressure agents have little effect on preventing fretting. It should be remembered that liquid lubricants do not prevent fretting they only reduce the effect of fretting.

As a solid lubricant, Molybdenum disulfide, if it is applied correctly, will to a large extent postpone the beginning of fetting corrosion. The method used to apply MoS_2 to steel specimens is to bond it with a mixture of powder and corn syrup into intimate contact with the cleaned hot metal.

In the situation of press fitted assemblies, the best method for retarding the propagation of fatigue cracks initiated by fretting is to obtain "favorable residual thermal compressive" stresses on the surface. These residual stresses are obtained by subcritical quenching, that is, quenching from below the "critical" or phase transformation temperature, and were effective in Horger's experiment[9] to the extent of increasing the endurance limit 64%.

The effects of the type of steel, the tensile properties, and whether the shafts were quenched and tempered or normalized and tempered had much less effect on the fatigue resistance.

The shape of the assembly is also important but is a less important factor than the "favorable" residual stresses. The optium surface finish for fatigue resistance depends on the different situations that arise. Smooth finishes are best where relative motion occurs while in press fitted assemblies, if the roughness does not effect the truth of the fit, lubricants can collect in pits in the surface and thereby improve fretting resistance by acting as small oil reservoirs. The rough surfaces can also allow the oxide debris to escape easier than a smooth surface.

Another method that can be used to increase the fatigue resistance is to harden the surface either by heat treating or by cold working (See Figures 14 & 15). This will result in more of the microwelds breaking at the original junction giving smaller wear particles. Case hardening and nitriding will improve the fatigue strength in a test where fretting is present but how much of the increase in endurance strength is due to the reduction of fretting and how much is due to the improvement of fatigue strength is uncertain.

Using a rubber gasket or rubber cement will reduce fretting because the slip is taken up in the elastic material. Damping out vibrations that cause oscillatory motion is also benefical in reducing fretting damage.

A final remedy is to alter the existing conditons. If the relative motion or slip can be prevented by increasing the load, fretting will not occur. However if the load is increased but does not stop the motion the fretting damage will increase.

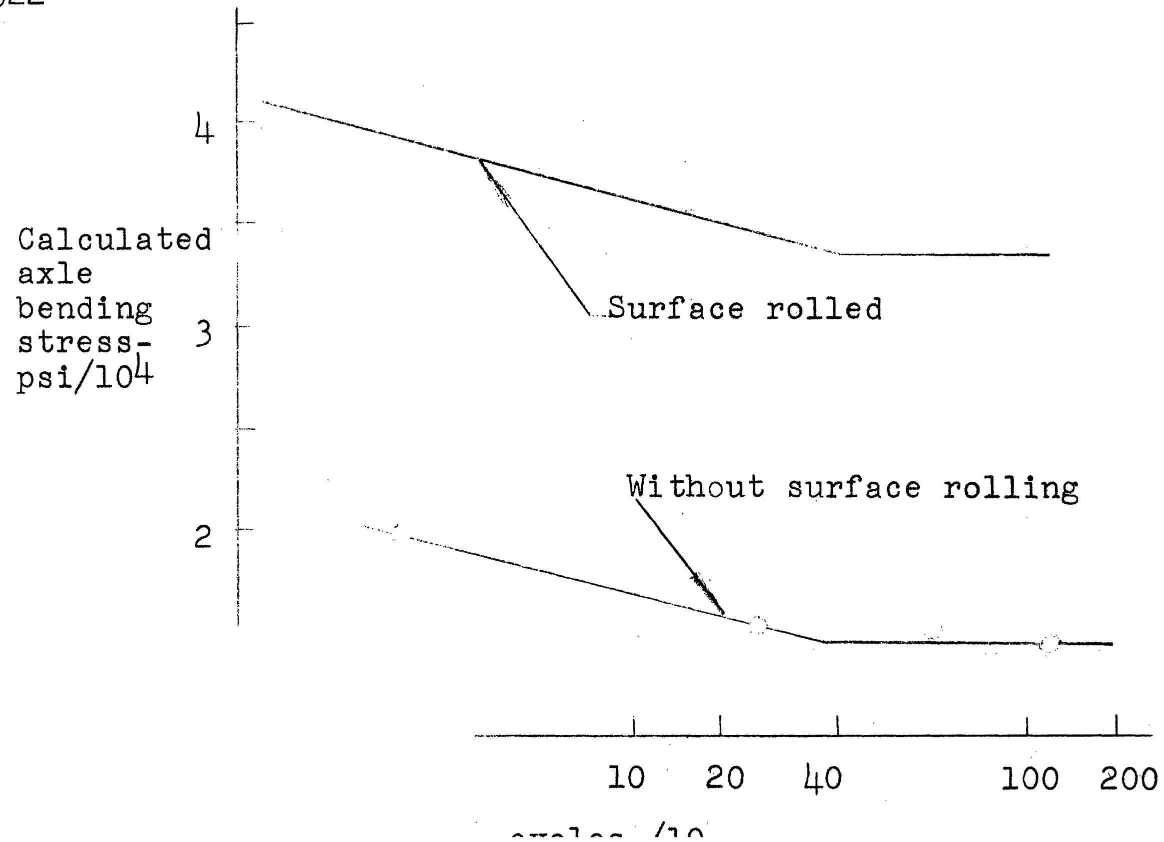


Figure 14

Effect of surface rolling on the fretting-fatigue strength of SAE 1045 steel axles.

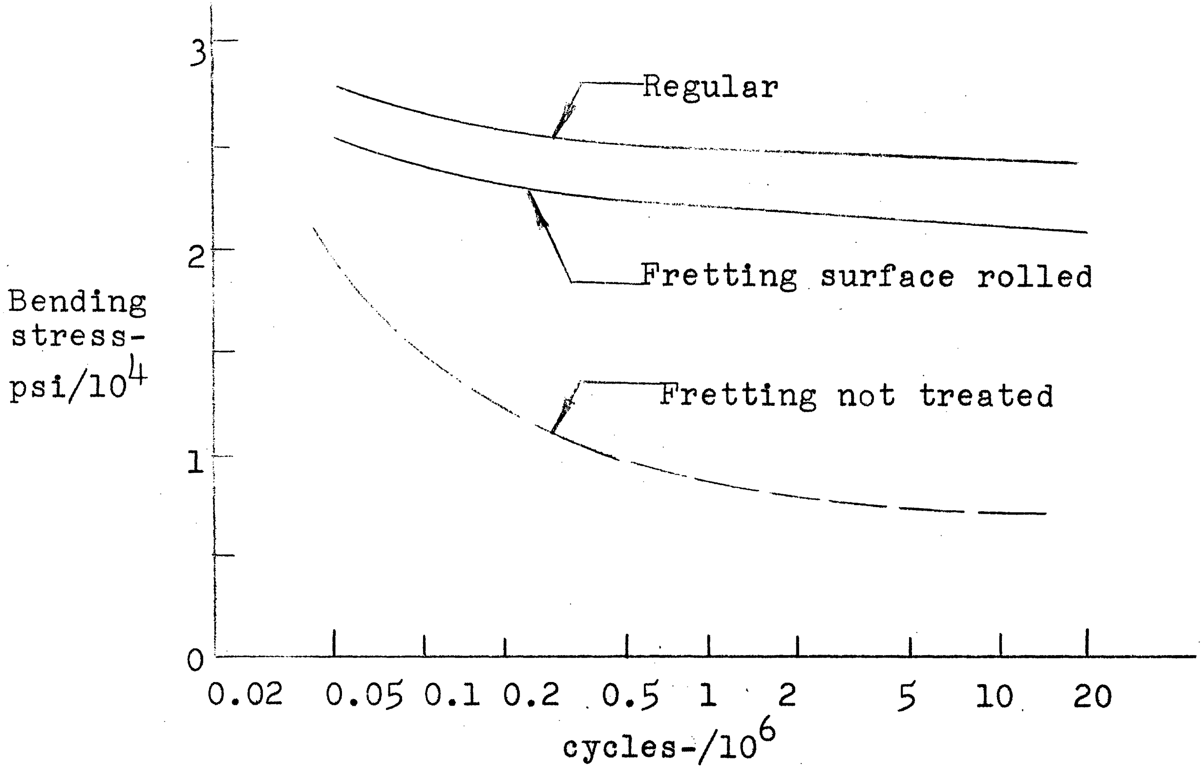


Figure 15

S N curves for regular fatigue and for fretting fatigue of surface-rolled and untreated specimens of magnesium alloys.

BIBLIOGRAPHY

1. Tomilinson, G. A., Thorpe, P. L., and Gough. H. J., "An Investigation of Fretting Corrosion of Closely Fitting Surfaces," Procedings, Institute of Mechanical Engineers, Vol. 141, 1939, p. 223.
2. Wright, K. H. R., "An Investigation of Fretting Corrosion," Proceedings, Institute of Mechanical Engineers, Vol. 18, 1952-53, p.556.
3. Feng, I. M., and Rightmire, B. G., "The Mechanism of Fretting," Lubrication Engineering, Vol. 9, p. 134.
4. Waterhouse, R. B., "Fretting Corrosion," Institute of Mechanical Engineers, Vol. 169, No. 59, 1955, p. 1157.
5. Fenge, I. M., Rightmire, B. G., and Ming, "An Experimental Study of Fretting," Proceedings, Institute of Mechanical Engineers, Vol. 170, p. 1055, 1955.
6. Uhlig, H. H. Feng, I. M., Tierney, W. D., and McCellan, A, "Fundamental Investigation of Fretting Corrosion, " NACA Technical Note 3029, December, 1953.
7. Uhlig, H. H., Tierney, W. D., and McCellan, A., "Test Equipment for Evaluating Fretting Corrosion," ASTM Special Publication, No. 144, June, 1952.
8. Uhlig, H. H., "Mechanism of Fretting Corrosion," Journal of Applied Mechanics, Vol. 21, No. 4, December, 1954, p. 401.
9. Godfrey, D., "Investigation of Fretting Corrosion by Microscopic Observation," NACA Technical Note No. 2039, Feb., 1950.
10. Horder, O. J., "The Influence of Fretting Corrosion on the Fatigue Strength of Fitted Members," ASTM Special Technical Publication No. 144, June, 1952.
11. Horder, O. J., "Fatigue of Large Shafts by Fretting Corrosion," Proceedings of the International Conference on the Fatigue of Metals, Institute of Mechanical Engineers, 1956, p. 352.
12. Warlow-Davies, E. J., "Fretting Corrosion and Fatigue Strength: Brief Results of Preliminary Experiments," Institute of Mechanical Engineers, Vol. 146, 1941, p. 32.
13. Fenner, A. J. Wright, K. H. R., and Mann, J. Y., "Fretting Corrosion and Its Influence on Fatigue Failures," Proceedings, Institute of Mechanical Engineers, 1956, p. 386.
14. Eden, E. M., Rose, W. N., and Cunningham, F. L., Proceedings, Institute of Mechanical Engineers, 1911, p. 839.
15. Bailey, J.M., and Godfrey, D., "Coefficient of Friction and Damage to Contact Area During the Early Stages of Fretting. II-Steel, Iron, Iron Oxide, and Glass Combinations," NACA Technical Note 3144, April, 1954.

CHAPTER 14

HIGH TEMPERATURE CORROSION

1. THEORETICAL ASPECTS OF CORROSION

Perhaps the first question to arise in the discussion of high temperature oxidation corrosion would be: How does this form of corrosion differ from oxidation reactions at lower temperatures? At normal temperatures iron and steel will not corrode appreciably in the absence of moisture. Therefore, when moisture is present the reaction that does occur is usually electrochemical and the result in as insoluble ferric hydroxide, more commonly known and referred to as ordinary rust. However high temperature oxidation appears to be due mainly to the direct chemical combination of oxygen with the metal. In iron and steel this reaction leads to a complex oxide layer.

The Iron Oxide Layer

The experiments of Doctor L. B. Pfeil have shown the oxide to be made up of three easily discernable layers;[1] FeO on the iron surface, an intermediate layer of Fe_3O_4, and a thin outside layer of Fe_2O_3 on the outside. The outer surface of the outer layer of the scale has a smooth velvet-like appearance while the under surface has a bright lightly etched appearance. The thickness of the layer varies with time, temperature and air supply. The structure is crystalline in appearance.

The outer surface of the middle layer has a bright lightly etched appearance and gives a distinct indication of being crystalline in structure also. The inner surface of the middle

Layer	Temp. of Oxidation-°C	Period of Oxidation-Days	Limiting Dia. of Layers-in.	Iron %	Approx. A %	
					FeO	Fe_2O
Specimen A						
Layer 1			0.581 to 0.551	71.6	21.3	78.
Layer 2	850	10	0.551 to 0.497	75.5	71.4	28.
Layer 3			0.497 to 0.444	76.5	84.1	15.
Specimen B						
Layer 1			0.635 to 0.589	71.55	20.6	79.
Layer 2	1050	9	0.589 to 0.498	75.45	70.7	29.
Layer 3			0.498 to 0.380	76.0	77.8	22.

NOTE: Pieces Originally 1/2" in Dia.

Table 1

Oxidation of Pure Iron

layer shows perfectly in reverse all the details of the original surface of the iron before oxidation, such as identification marks or numbers and similiar details. The position of the inner surface corresponds with that of the external dimensions of the iron before oxidation. This feature has special significance which will be brought up later in the paper.

The outer surface of the innermost layer has a marked crystalline appearance. The layer consists of minute crystals of almost perfect cubes lightly fitted together. They become more consolidated as the metal core is approached. The outer surface of this layer is so very porous that contact with the inner surface of the middle layer is only made at scattered points.

The color of all three layers is black. This oxide formation, described above as consisting of three easily discernable layers, is produced regardless of the chemical composition of the steel. About 10% of the total scale thickness consists of the outside layer, and about 40% of the innermost layer.

The three layers vary considerably in iron content. The outside layer is the lowest and the innermost layer the highest in iron. A typical analysis of the three layers formed on pure iron oxidized at temperatures of 850°C is given in Table 1.

The Mechanism of Oxidation

The formation of coherent scale deposits on iron is difficult to explain, owing to the absence of the disruptive action of the surface which might be anticipated as a consequence of the expansion accompanying oxidation. As the scale expands the outer surface retains the velvet-like smoothness previously mentioned and no cracks appear. This is particularly puzzling where the oxidation occurs on severely curved surfaces. If it were supposed that as oxidation proceeded, the outer scale was gradually expanded by the pressure exerted by a new scale forming beneath at the metal surface, the outer scale would certainly have to possess a pronounced degree of plasticity to refrain from cracking. While it is possible that scale is plastic at very high temperatures, Dr. Pfeil found that at temperatures of the order of 1000°C. and lower the scale behaves as a brittle and non-plastic substance. Even supposing for a moment that the scale were sufficiently plastic to undergo appreciable distortion without fracture; there remains the difficulty of accounting for the loose scale which forms in the inner layer and is in contact with the corroded metal. This loose scale, by reason of its very porous nature, could not possibly exert the pressure required to keep pushing the outside layer outward during scale formation. Therefore the suggestion that the progressive oxidation of iron from the metallic core by the pressure exerted by the new scale forming beneath is illogical and a new explanation must be sought.

The Pfeil Theory

Dr. Pfeil offers the following hypothesis: At the iron scale interface, iron is continually being dissolved in the scale (i.e. the iron is being converted into oxide which then forms part of the scale deposit) forming an iron-rich solid solution and the diffusion of the iron occurs outwards through the scale to the scale-air interface.[2]

To give weight to this hypothesis Dr. Pfeil showed, by a long series of experiments, that the following phenomena are true. He found that the outer rich layer of the scale occuring in the high temperature oxidation of pure iron contains approximately 72% iron and that the inner oxide layer contained approximately 77% iron. He then used a sample of each scale and heated them together in the presence of air. He found that the scale whose oxide had the highest iron content lost weight, while the scale of lowest iron content gained weight until and equilibrium condition existed, and the percentage weight of iron in both specimens became the same. Dr. Pfeil also found that this same phenomena occurred between the iron and the scale. In this case the scale, of course, was of lowest iron content. These two experiments strongly indicate that iron actually does diffuse through the scale from the iron to the scale-air surface.

The Oxidation Phenomena

Now to apply Dr. Pfeil's theory to the oxidation phenomena. Dealing first with the outside layer: the molecules which compose this layer are ever changing, for iron is continually being brought to the surface by diffusion. This tends to thicken the layer and at the same time the thickening tends to reduce the rate at which oxygen reaches the inner part of the layer. The inner portion, becoming rich in iron, then joins layer two.

Actually, as was previously indicated in the description of the inner surface on the intermediate layer, the scale grows both outwards and inwards from a position determined by the original dimensions of the iron. The scale found outside the original dimensions of the iron reaches its position by diffusion. That scale found within the original dimensions of the iron grows deeper into the metal as corrosion progresses and becomes increasingly porous as the metal is erroded away by diffusion.

It is now clear why the inner surface of the middle layer reproduces the surface characteristics of the original iron, for it was formed at the very start of oxidation, and the first formed scale would of necessity follow the surface irregularity of the iron. The surface characteristics of the iron are not visible on the outside surface of the middle layer because of the depositing of oxide on the outside surface which gradually covers the markings. Similiarly, no markings are visible on either surface of the outside layer. The outer surface of the innermost layer is too porous to show clearly defined details, although it was in contact with the inner surface of the middle layer on which the markings were most clearly defined.

The Rate of Corrosion

The rate of corrosion or scaling of iron and steel at elevated temperatures is influenced by a number of factors, those best established being cleanliness or roughness of the metal surface, environment and temperature of the metal, time at a given temperature, and the composition of the metal.[3]

Influence of Cleanliness

Cleanliness of the surface is a factor whose significance is too often neglected. The presence of grease has been found to cause a more porous scale with a consequent increase in the rate of corrosion particularly in the initial stage of the process. As the scale thickness increases, the rate of corrosion will become more dependent on the diffusion of iron molecules (not only on the rate of oxygen diffusion) thereby decreasing the detrimental effects of the porous oxide. Nevertheless, the initial rate of oxidation is decidely increased due to surface grease. Also, surfaces are seldom "unclean" to the same degree over the entire surface, the rate of corrosion frequently varies on different parts of the surface. The varied rate of corrosion causes cracks and blisters to occur in the scale and this, in turn, causes a further increase in the corrosion rate since oxygen diffusion is no longer restricted by scale in these regions.

Influence of Surface Roughness

The influence of the surface roughness was determined by experiments run by M. J. Day and G. V. Smith[4]. They compared the rate of oxidation at 1100°F. of several steel specimens having surfaces ranged from milled to polished. The polished surfaces were least attacked and, as may be expected, the milled surfaces were attacked to the greatest degree. This is attributed to the fact that the milled specimens had a slightly greater surface area (due to the relatively deep grooves caused by the mill) than the polished pieces. Also it was found, quite logically, that the influence of roughness is greatest at the beginning of corrosion and diminishes as the original surface corrodes away.

Rate of Oxidation in Gases other than Air

The use of iron and steel at high temperatures in other gases than air also has a marked effect on the rate of corrosion. Both iron and steel commonly scale more rapidly in oxygen than in air. The magnitude of this difference is indicated by the data presented in Figure 1, which shows bands representing the range of oxidation in air or oxygen at constant temperaure.

Although the rate of corrosion of iron and steel in gases other than air or oxygen has not been widely investigated, it has been established beyond question that carbon dioxide, carbon monoxide, steam, and sulfur dioxide are active scaling agents. This is thought to be due to the formation of a particular pervious scale by each of these gases in high temperature oxidation. Quantitatively, data collected is not entirely in agreement, but qualitatively there is general agreement that, at a given temperature, the rate of scaling in carbon dioxide is less than in air because of a less pervious scale, whereas in steam,sulfur dioxide, and carbon monoxide it is significantly greater than in air due to a more pervious scale.

Linear Oxidation Rates

Linear oxidation rates occur when the oxide film which forms is volatile in the environment, or one which fails mechanically,

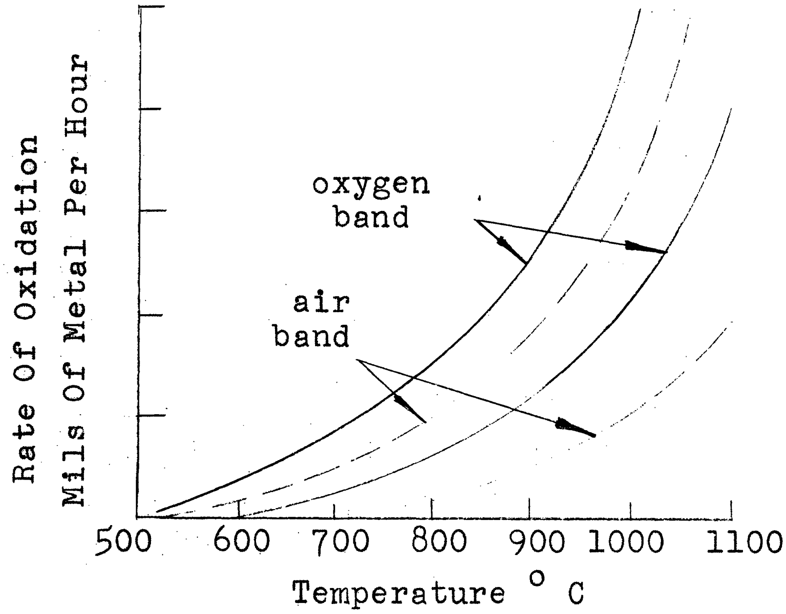


Figure 1
Rate of Oxidation of Iron or Steel in Air or Oxygen

BeO	10^{-12}	NiO	10^{-2}
Al_2O_3	10^{-9}	Cr_2O_3	10^{-1}
CaO	10^{-7}	CaO	10^{1}
SiO_2	10^{-6}	Cu_2O	10^{1}
MgO	10^{-5}	FeO	10^{2}

Figure 2
Oxide Conductivity at 1000°-(OHM-CM)$^{-1}$

	C	MN	S1	CR	N1
3140	.4	.8	.3	.7	1.3
XB	.8	.4	2.0	21.	1.5
Sil-1	.5	.4	3.3	8.5	---
21-12N	.2	1.3	1.0	21	11.5
21-4N	.6	9.	.2	21	4.
TPM	1	2.	1.	15	BAL

Figure 3
Typical Stainless Steels

that is cracks and flakes away from the surface. The molybdenum oxide, MoO_3, for instance, is volatile at temperatures above 850°. This leaves the bare metal open to further attack. Other metals, such as sodium and calcium, form porous or non-adherent films which, again, cannot protect the metal surface. Generally, these oxides are such that tho volume of oxide produced is less than the lost metal volume, causing the oxide to crack and allowing further oxidation. For all metals in this classificiation, the linear corrosion rate prohibits their use under high temperature service conditions.

Parabolic Corrosion Rates

Parabolic corrosion rates occur when oxide films which form have a specific volume greater than the parent metal. These oxides will often remain continuous, and have the property of adhering to the surface, thus preventing continuous corrosion.

Since the metal ion must diffuse through the oxide film to the surface to react and oxidize, corrosion will decrease as the thickness of the protective film increases. From the model of diffusion and film buildup, the 'parabolic rate process can be derived. The reader is referred to Reference 5 in the bibliography for a rigorous derivation of the relation. The result is

$$W^2 = t + C$$

where $\underline{W}$ is the weight loss by corrosion, $\underline{t}$ the time elapsed, and $\underline{C}$ a constant for the metal oxide, and primarily dependant on its conductivity.

Since the oxide coating determines the rate of corrosion, alloying for high temperature service must therefore involve adding metals which form parabolic-rate oxides. Price and Thomas reasoned that since the constant above depended on the conductivity, oxides of low conductivity would be good alloying elements. Experimental evidence agrees well with this idea. Figure 2 shows the conductivity of various oxides, and many of those with low conductivity are in use in today's high temperature metals.

2. PHYSICAL ASPECTS OF CORROSION

Actual corrosion cases, as evidenced by exhaust valves, are usually complex reactions invovling three basic types of high temperature wear; metal oxide corrosion, galvanic action, and gross burning.

Metal Oxide Corrosion

This type of attack is often seen in engine valves. Metal oxides in the exhaust gasses react with the protective film, thus exposing the bare metal to further oxidation, and resulting in a somewhat linear corrosion rate. The problem has become most acute since the addition of tetraethyl lead to automotive gasolines. The lead oxide which forms reacts with the protective chrome oxides according to the equation:

$$2Cr_2O_3 + 6PbO + 3O_2 \longrightarrow 2Pb_3Cr_2O_9$$

Many valve materials, such as Silchrone #1 and Silchrone XB, were widely used in exhaust valves before the use of tetraethyl lead. These two alloys of chromium and silicon now find use only in valves where moderate service temperatures are involved.

Under certain conditions, an extreme phenomenon known as catastrophic corrosion may occur by a method similar to that described above. In such cases, the rate of corrosion is initally parabolic, but proceeds to a rapid linear rate at higher temperatures. It has been found that this is due to the presence of a liquid oxide, usually MoO_3, which forms just before the rapid reaction starts. The primary difference is that the protective oxide reacts with the liquid oxide,rather than with the environment. In the case of molybdenum oxide and steel in such a reaction, the liquid oxide reacts with the pure metal at the metal surface as follows:

$$MoO_3 + FeCrMo\ (alloy) \longrightarrow Cr_2O_3\downarrow + Fe_2O_3\downarrow + MoO_2$$

Oxides of the parent metal and alloying elements precipitate out and the MoO_2 picks up another oxygen atom at the oxide surface and reacts again:

$$2MoO_2 + O_2 \longrightarrow 2MoO_3$$

In the case of steels, sufficient additions of nickel halts this reaction by reacting preferentially with the liquid oxide. Metals of the Mo-Fe-Ni-Cr series, with about 40% Ni and 20% Mo seem to be well out of the range of catastrophic corrosion.

Galvanic Action

The second process of high temperature corrosion, that of galvanic action, is essentially the same in nature at elevated temperatures as it is at normal temperatures. It is somewhat accelerated at higher temperatures due to an increase of conductivity and rate of reaction.

Valve faces are often made of materials different from the base, thus introducing a galvanic couple near the center of the head. Since valve failures usually occur at the face, (Figure 4) this seems to have no great effect on valve life. Considerable material can be lost near the center of the valve head before it becomes structurally unsound. Deposits of metal oxides from the exhuast gasses will also introduce galvanic couples, although to a lesser extent.

The effects of stress variation and discontinuties in galvanic action have been discussed in another chapter, and its conclusions agree almost completely here.

Gross Burning

This area of valve corrosion is usually the final stage of one of the other two areas of wear presented above. It can occur alone, however, and is usually closely tied to the valve area geometry in the engine. The problem will be considered later under that heading.

Metallurgy

Alloys used in exhaust valve design must exhibit optimum properties in many areas aside from corrosion resistance. Extreme pressure, high numbers of stress cycles, and high impact loads require the material to exhibit high hot hardness, high fatigue

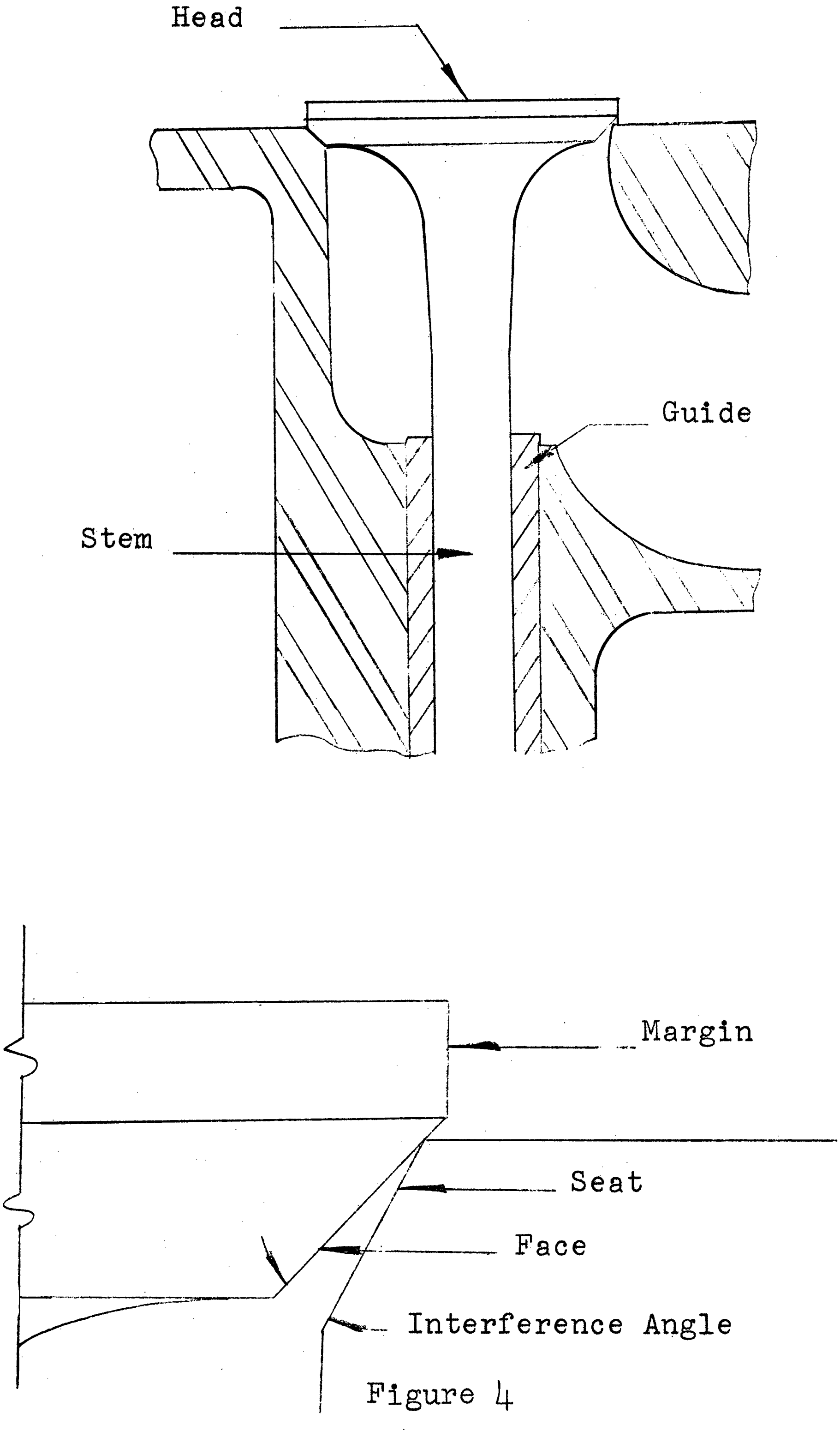


Figure 4

alve Nomenclature

strength, and low hot creep, for example. The metallurgist must compromise the requirements in these varied areas in choosing metals for high temperature service.

Stainless Steels

Valve materials are often chosen from the broad classification of steels listed in Figure 3. They usually contain high amounts of chromium, nickel, and molybdenum, with lesser amounts of other elements. Austenitic steels are usually used, although ferritic steels may be used in areas of moderate service.

Austenitic steels are usually typified by the high nickel content which stabilizes the austenitic structure at room temperature. Before the use of lead in automotive gasolines, materials such as AISI 3140, Silchrome #1, Silchrome XB, and other chrome-silicon alloys were used extensively. The lead oxide reaction now restricts their use to areas of moderate service conditions, however, More recently, materials such as Silchrome XCR, 2112N, 214N, and 10N have been found to stand up well under extreme conditions of stress and corrosion. They are generally high in both nickel and manganese, with less than 0.5% silicon. (See Figure 5).

The effects of silicon has only come into complete understanding within the past few years. It was felt for some time that additions retarded the corrosion rate, when actually the complete removal of the element results in a much smaller corrosion rate. (See Figure 6) Although the addition of silicon is helpful in pure oxidation of the metal, it will react with metal

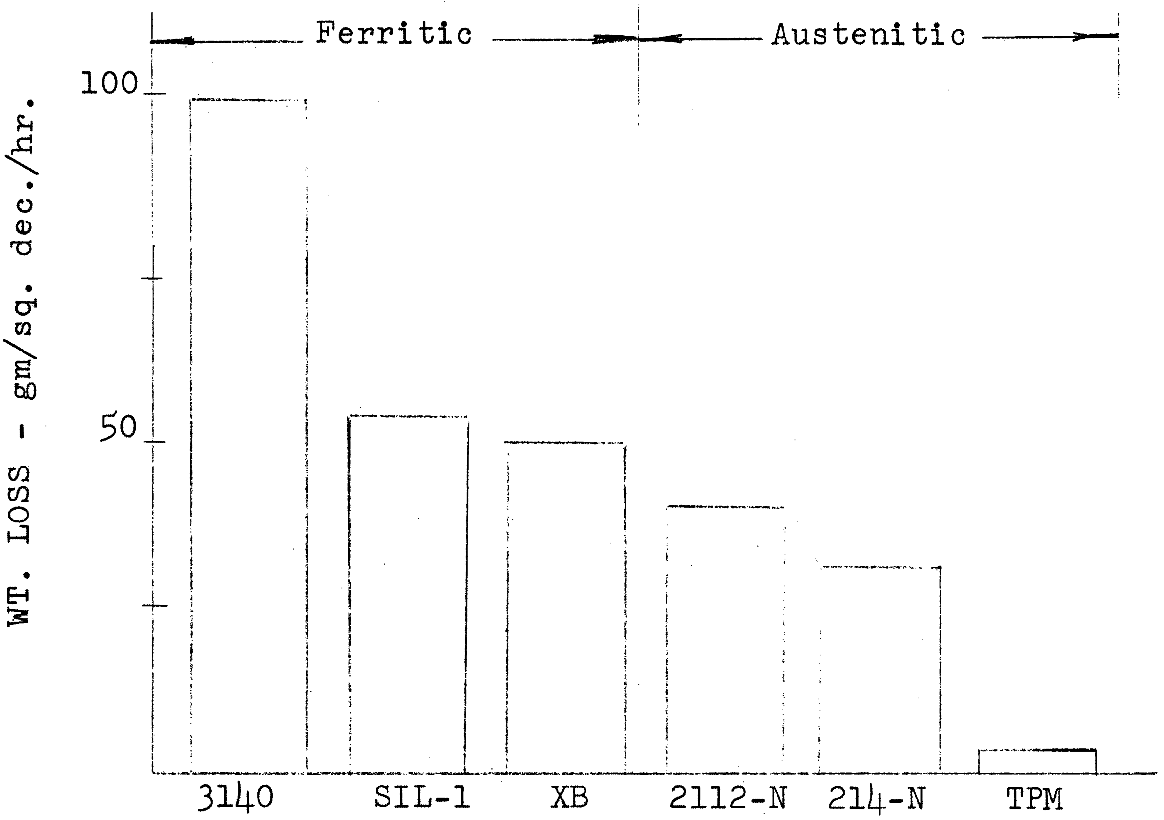


Figure 5

Corrosion Rates for Stainless Steels

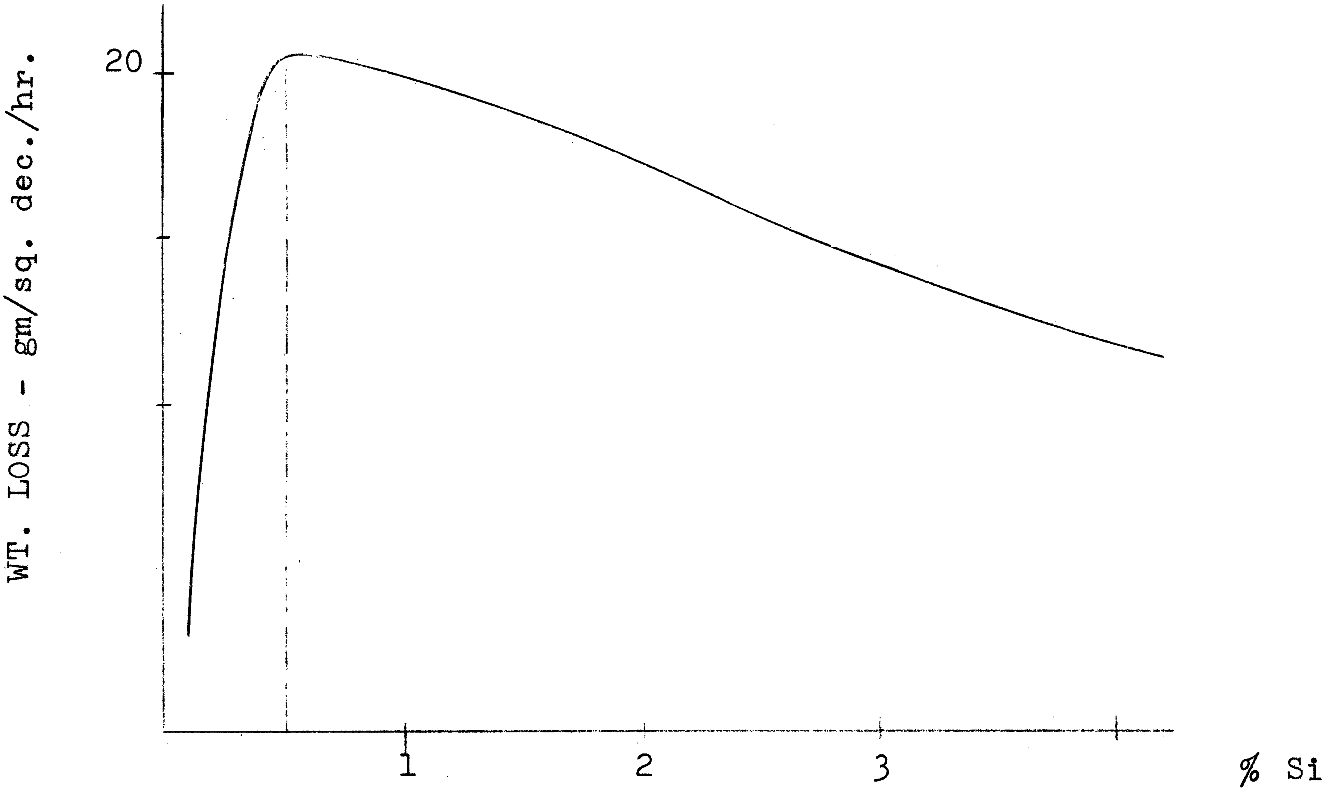


Figure 6

Effect of Silicon on Corrosion

oxides in the combustion gasses, with results as mentioned earlier. For this reason, newer alloys show a marked decrease in silicon content, and have a much longer useful life.

In most austenitic steels, manganese and chromium are interchangeable, with additive effects up to about 10% Mn.

Nickel Base Alloys

Nickel base alloys are replacing the stainless steels in extreme situations, such as in aircraft engines. The latest alloys include chrome-nickel-molybdenum and chrome-nickel-cobalt. The molybdenum alloy has been considered primarily for its high strength, but it tends to oxidize at high percentages of molybdenum. The cobalt alloy shows a pronounced corrosion drop, but it economically unfavorable in most areas.

It was felt that nickel addition above 60% would actually decrease corrosion resistance, for some time. More recent studies have traced the increased corrosion at high nickel content to a corresponding increase in silicon content, and higher nickel alloys are now available. It should be mentioned that aluminum and silicon are both detrimental to nickel alloys and should be avoided, even in small amounts.

Generally, nickel alloys will contain magnesium and titanium to about 3% to lower corrosion rate, with tungsten to increase the hardness. Aluminum and silicon will always be under 0.3% in good grade alloys.

Several new processes are becoming available for use in high temperature metallurgy. The use of vacuum melting apparatus has had good results so far. This process prevents the oxidation of the liquid metal during melting. Valve castings for materials too hard to forge have been quite successful in the use of so-called super-alloys for valve materials. Finally, protective coatings for valve materials are being investigated. The "Al-Dip" process, designed by General Motors, coats the entire valve with pure aluminum by a dipping operation. The steel and aluminum react to form complex compounds at the interface with resultant high hardness and corrosion resistance.

Valve Geometry

Generally, design changes in the valve shape, or design changes of the valve area in the engine are attempts to lower the temperature of the valve, (Figure 7) either by increasing heat transfer from the valve, or by protecting it from hot exhaust gas. As in the earlier metallurgical considerations, optimization of geometry from the valve life standpoint may be prohibitive from other considerations, and compromises must always be made.

The valve head is usually made heavy enough for strength and heat dissapation, but cannot be made so thick as to extend too far into the cylinder. Since many failures occur at the valve face, certain nickel valve-facing alloys are often welded around the face. The margin of the valve often becomes too thin after the valve grinding operation. This thin edge will heat up quickly and often chips, thus impairing the sealing ability of

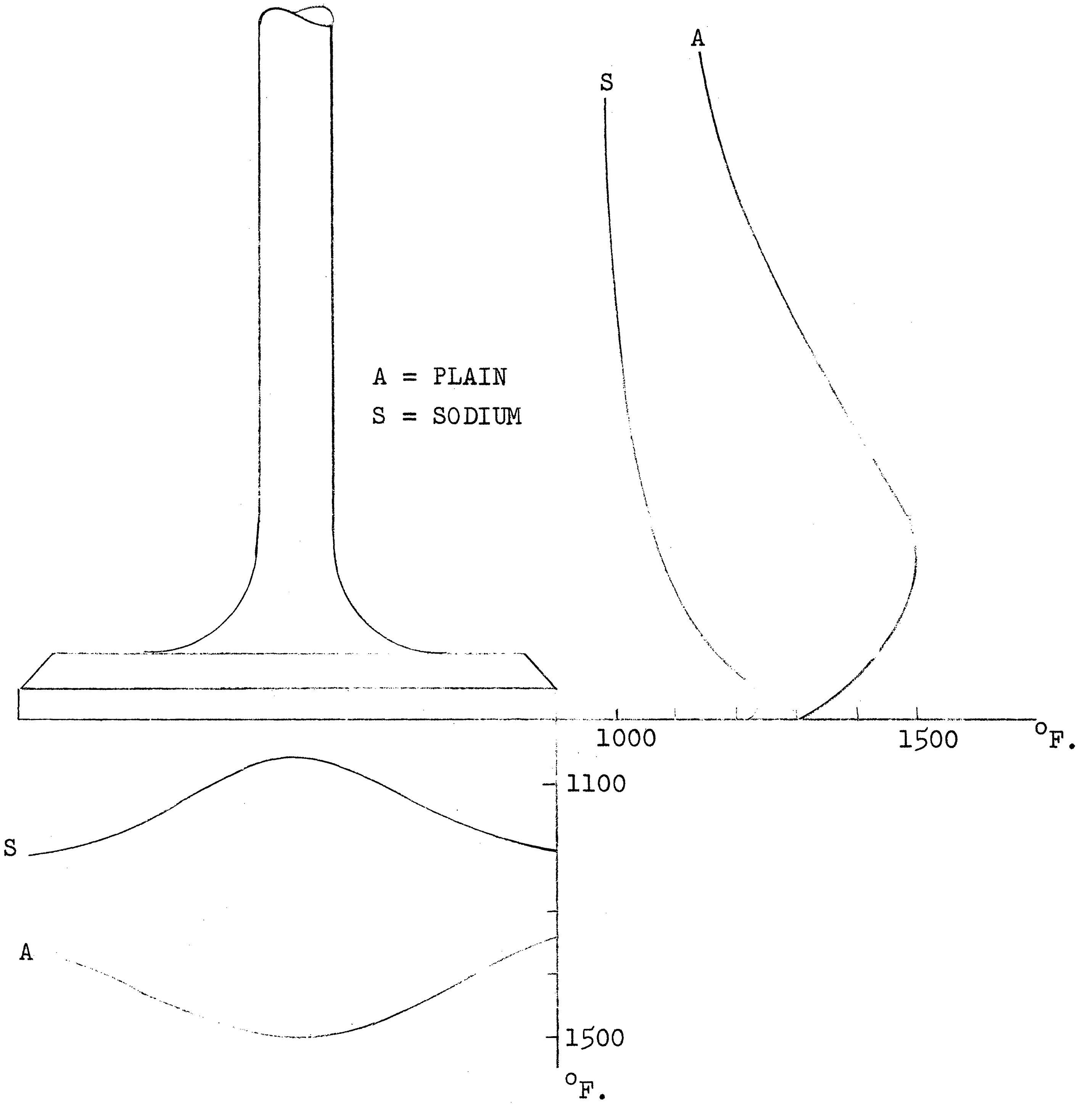


Figure 7
Valve Temperature Profiles

the valve. Margins which are too wide tend to collect engine deposits and thereby hamper the sealing action also. This commonly occurs also in faces and seats which are too wide.

The valve face is usually ground so that there is a small angle between the face and the valve seat, with contact only near the top of the valve face. This use of the "interference angle" provides a more positive seal, and offers less change for foreign particles to become caught between the seat and face as the valve closes.

Seat distortion also hampers the sealing action. Such things as uneven tension in cylinder head bolts or overheating of the engine can cause such distortion, with resultant burning of the valves.

The use of sodium-cooled valves and hollow-head valves can decrease temperatures considerably. (See Figure 7) The heat can flow down into the valve steam, where it can be transferred thru the guides to the water jacket. Sodium valves can reduce the temperatures by as much as 400°, while hollow-head valves are somewhat less efficient.

The greatest benefit so far, however, has been obtained from valve rotators, a device developed by Thompson Products in 1938, and improved considerably since that time. This device is attached to the head or tip-end of the valve, and rotates the valve in its seat during operation. (See Figure 8) This provides a constant wiping action which prevents the buildup of deposits on the valve seats. Similarly, if leakage is occurring at some

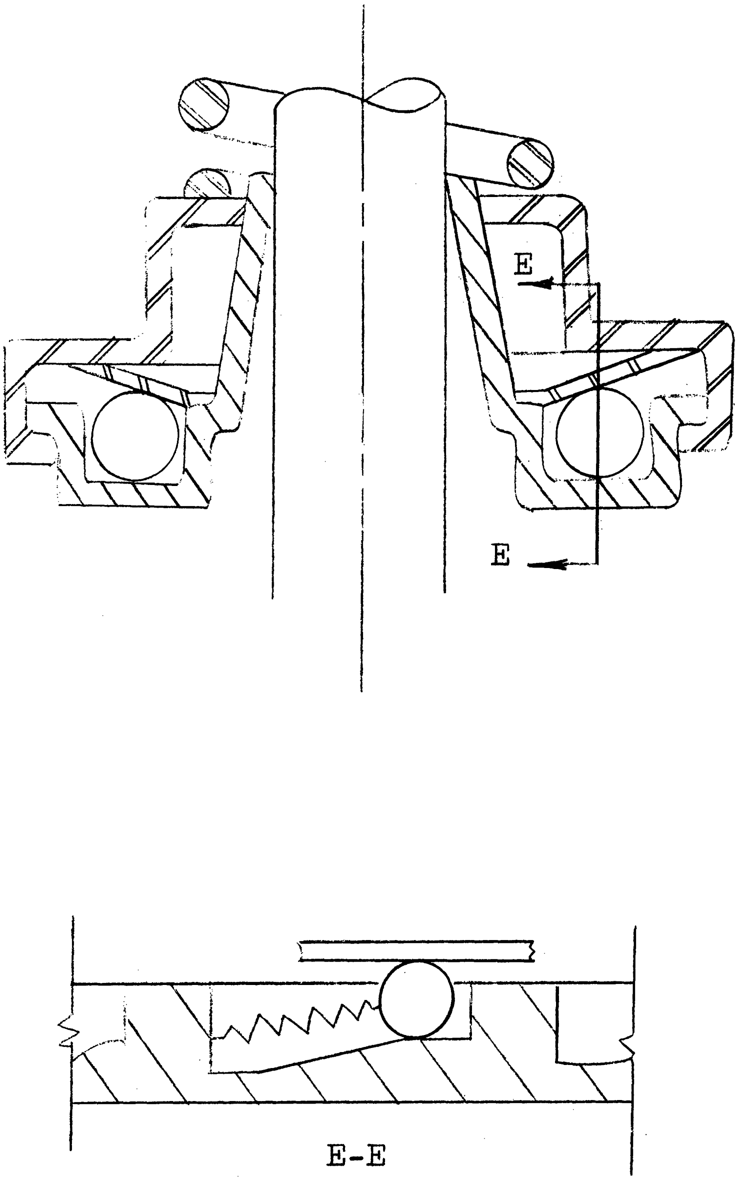


Figure 8
Valve Rotator

point on the seat, the rotation will prevent the continued burning of any single area on the valve face. Valve life has been increased by as much as ten times using such rotators, and they are usually operative for about 100,000 miles, so that replacements can be made during other maintainance work which will become necessary during that time. The positive type rotator uses a conical spring and several steel balls which allow the valve to rotate, giving a positive rotating action. The release type depends on engine vibration for movement of the valve, and such motion is random. The Ford Motor Company has developed a rotator similar to the release type.

3. LABORATORY AND STATISTICAL TECHNIQUES

In present day laboratory testing of engine valve life, no accurate and reliable test exists which can predict life under actual service conditions. At best, the available bench tests can serve to generally indicate promising alloys which can then be subjected to actual service testing on the road. In actual service conditions, there still exists a large spread in data. It can only be said that all the factors mentioned in this discussion have varying effects in each individual engine, and the extent of variance tends to give very unpredictable results.

The Lead Oxide Test

The lead oxide test has gained wide popularity and gives fairly good results. A cylindrical specimen is placed in a magnesia crucible with 40 grams of PbO and heated to 1675° for one hour. The specimen is then cleaned and the weight loss taken as an indication of the corrosion resistance. The shortcomings of the test are obvious, since only the metal oxide reaction is considered in the corrosion of the metal. However, this type of attack is certainly one of those most common in actual service, since its effect was noted in several instances earlier in this discussion. Also, as was mentioned, fairly good correlation has been obtained from the test so far.

Variability in Valve Data

Since variability is inherent in valve studies, with mere chance playing so important a part, statistical methods must inevitably be used in extracting reasonable conclusions from the data. A few examples of the method of approach should be mentioned here.

In Figure 9, valve life is defined in the following way by the Ethyl Corporation. The valves are run until two failures are obtained. After the first failure the time is noted and the valve replaced. The test is continued until a second failure occurs. Valve life is then defined as the average time of the two failures. Some increase in the reliability of data has been obtained in this way.

Figure 9 is very useful in setting up tests for valve life. From it, given the expected per cent increase in valve life (or other variable), the number of tests necessary to eliminate effects of spread in the data can be found. Conversely, given

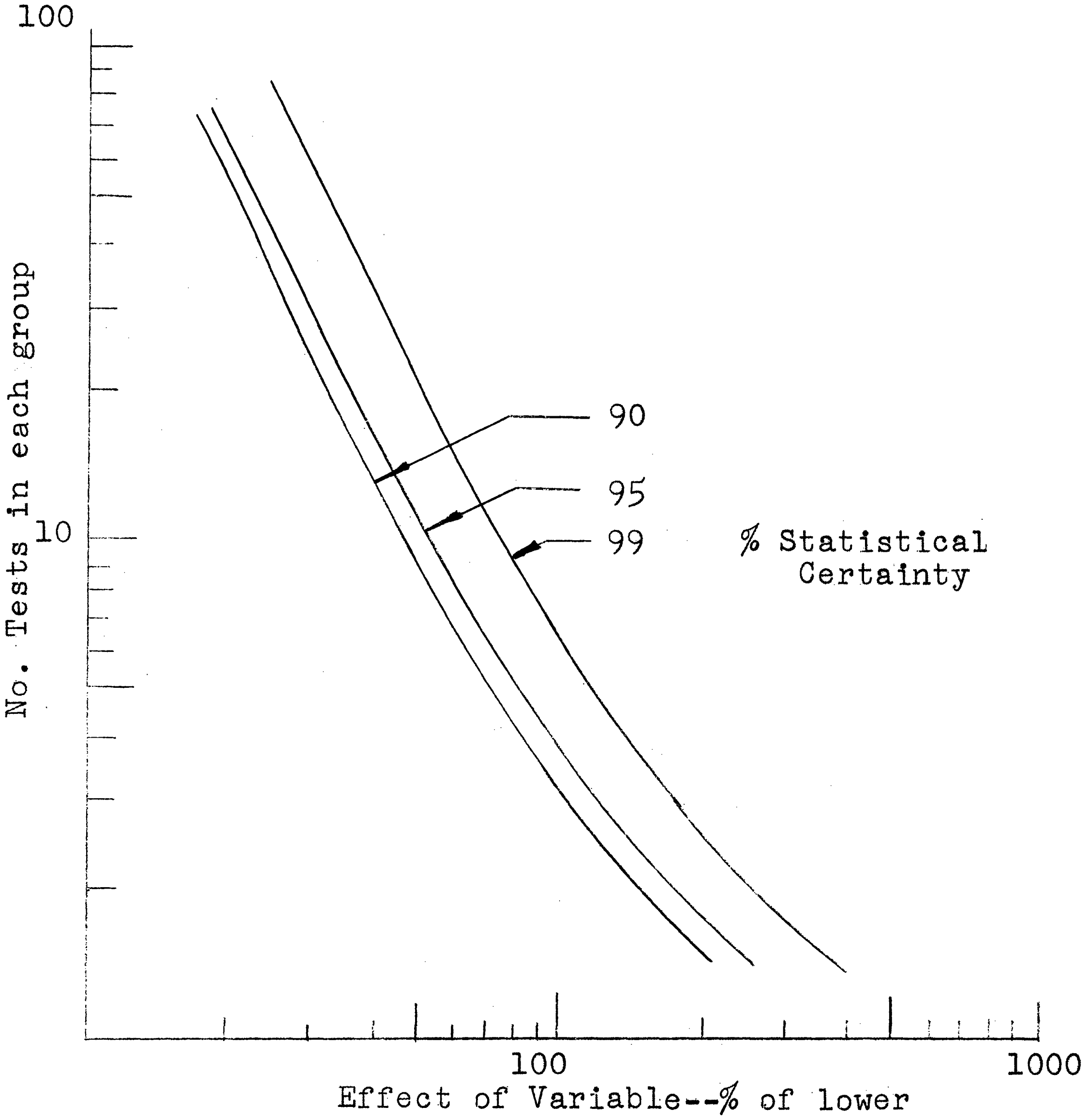


Figure 9

Variability of Exhaust Valve Life

a certain number of tests, the smallest per cent increase in valve life which can be measured by this number of tests can be found. Note that three different probability levels are given. For example, if a change is expected to give 30% difference in valve life, no less than 15 tests are necessary, and the probability of correctness of this examination will be 95%. For a 90% probability of correctness, fewer tests would have been necessary, while for 99% probability, more tests would have been required. Fortunately, in valve test work the expected effects are usually large, perecentage-wise. It probably would not be worthwhile to make a change in the valve to obtain improvements much under 30-40%. Rather, improvements of the order of 50-100% and more are usually anticipated. The existance of this variance in valve testing data does not imply that the data is not reliable, but only that the tests must be properly designed and the evaluation of the data be properly done.

BIBLIOGRAPHY

1. Speller, F. N., Corrosion-Causes and Prevention, McGraw-Hill Book Co., Inc., 1951.
2. Uhlig, H. H., The Corrosion Handbook, John Wiley & Sons, Inc., Chicago, 1948.
3. Pfeil, L. B., "The Oxidation of Iron and Steel at High Temperatures", Journal of The Iron and Steel Institute, XIX, 1929.
4. Day, M. J., and Smith, G. V., Corrosion Rate", Industrial Engineering Chemistry, 1943.
5. "High Temperature-A Tool for the Future", Symposium at Stanford Research Institute, University of California, pp. 87-126.
6. "Engineering Approach to Surface Damage", Symposium at the University of Michigan, C. Lipson & L. V. Colwell.
7. "Marks' Mechanical Engineers Handbook", Theodore Baumeister, Ed., Section 6, pp. 40-46.
8. "High Temperature Properties of Metals", Presented at 32nd National Metal Congress, Chicago, October 23-27, by the American Society for Metals.
9. Pfudstein, K., and Baily, J., "Factors Affecting Tractor Valve Performance", S.A.E. National Tractor Meeting, Milwaukee, September 11-13, 1951.
10. "Engine Valve Failures", Quaker State Refining Corporation, Bulletin, p. 809, 7-47.
11. "Lubrication", Texas Co., Vol. 34, No. 7, July 1948.

CHAPTER 15

STRESS CORROSION CRACKING

1. GENERAL INFORMATION

Definition

The current edition of Metals Handbook[1] includes the following definition:

> Stress corrosion cracking - spontaneous failure of metals under combined action of corrosion and stress, residual or applied.

Although various other definitions of stress corrosion cracking are offered in different sources, the point generally in common to all is that the phenomenon is directly linked to the effects of two lesser phenomenon, stress and corrosion. With this common ground as a basis, a discussion of the nature of stress corrosion cracking follows.

The Nature of Stress Corrosion Cracking

Failure of metals by stress corrosion cracking appears in many different circumstances and evidences itself in as many different ways. Quite frequently the "attack" is sudden and rapid,occurring without any warning. Railroad truck wheels "exploding" while stored in a warehouse are good examples of failure with no advance indication. Generally, localized corrosion of a part takes place forming pits or other surface defects that affect a local stress concentration. The presence of stresses affords the possibility of exceedingly high stresses and cracking once the stress concentration points have developed. The combined action of corrosion and stress then results in crack propagation. The earliest record of stress corrosion cracking was reported in 1886 by Robert Austen[2] when he observed cracking of a gold-copper-silver alloy in a solution of ferric chloride. At that time the cracking was a "mystery". Today some manifestations, season cracking of brass being an example, tend to the "mysterious." As indicated in the introductory material, many varied ideas about the mechanism of stress corrosion cracking presently exists. There is, however, a generally accepted common understanding of the factors governing this most insidious phenomenon.

The Factors Governing Stress Corrosion Cracking

Generally speaking, the factors, that govern stress corrosion cracking are stress, environment, time, composition, and internal structure of the metal. It can immediately be seen that some of the above factors are those governing galvanic corrosion. This is to be expected since galvanic corrosion is a major part of the stress corrosion cracking mechanism.

Composition

So far as composition is concerned, it is possible to flatly state that there exist very few alloys that are not susceptible to stress corrosion cracking. It is definitely true that some metals and alloys are more or less susceptible than others; much the same situation exists for galvanic corrosion. Nonetheless, extensive work[3] has shown that stress corrosion cracking can be produced in nearly every engineering metal. This fact itself points out the importance of knowing about this phenomenon.

Internal Structure

In considering internal structure dependency there is considerable "mysticism." A study of galvanic corrosion reveals the dependency of this specific mechanism on internal structure. A comparable study of stress corrosion cracking shows a definite structural dependency. In most cases, the cracking is intergranular, but a change in internal structure, a phase change for example, may change the cracking pattern to transgranular. It is known that strength gradients exist within a given structure. The cracking will preferentially follow the path of least resistance. This means that specific patterns of cracking will vary considerably as various strength gradients are set up by different internal structures.

Time

Time as an entity is not a governing factor of the stress corrosion mechanism, but it does aid in the description of what constitutes a stress corrosion cracking failure. The point is that some metals and alloys enhance extremely slow stress corrosion cracking damage under certain conditions. The time element is used to indicate exactly what a "failure" is; a unit length crack per unit surface area for a given unit of time defines a stress corrosion cracking failure. With such a definition of failure, it is possible to classify " failure resistent" alloys for specific applications.

Environment

The environmental consideration is directly related to galvanic corrosion. For a specific environment "pure galvanic corrosion" may or may not proceed depending upon the environments acting as an electrolyte.

An additional noteworthy, but unexplained point is that a change of environment frequently causes cracking to be transgranular instead of the original intergranular variety, or vice versa. As stated, this change is unexplained as yet, but has been observed many times as reported in Morris's experiments.[5]

Stress

Of all the factors, the most important, aside from the basic galvanic considerations, is that of stress. The specific role of stress in the mechanism of stress corrosion cracking is not understood. A number of facts are known, however. It is generally accepted that the stresses involved in stress corrosion cracking failure are tensile in nature, or at least have a tensile component at the metal surface exposed to the environment.[4] As stated in the original definition, the stresses may be residual or operating, but in all cases they must be tensile in nature. Much experimental work has been done[3] to induce stress corrosion cracking utilizing compressive stresses. To date not one case of stress corrosion cracking has been obtained under these conditions.

Another fact supported by the experimental work of Morris[5] is that stress corrosion cracks generally proceed in a direction perpendicular to the maximum tensile stress component. Even in a test specimen in "pure torsion" the crack developed at an angle of 45° with respect to the axis of torsion, or 90° with respect to the plane of maximum shear stress, In many cases, however, grain elongation and orientation influence the direction of crack propagation as was suggested in the previous material.

Dix[4] of Alcoa indicates that the tensile stresses required for stress corrosion cracking are fifty to seventy-five percent of the yield strength of the material. Independent research has shown that considerably lower stresses can also result in cracking, other conditions being similar. The variation of stresses for cracking then leads to the search for a "stress corrosion cracking limit." This limit would be the maximum stress that a material could sustain under specific conditions, all being based on statistical life information such as that used for bearings. Testing has been done to determine values pertaining to certain sets of conditions, but the limit values obtained have been so erratic that it appears "impossible" to find such limits, even resorting to various mathematical reasoning. It is felt that although the same test conditions excepting the variations in stress were "held," different sizes and shapes of pits are initially formed.[6] The varied pits enhance varied stress concentration factors, thus the erratic results are explained: The "failure" stresses are actually the same for the cases considered, but with different stress concentrations, the number of stress concentration-times-actual stress products to yield the same "failure" stress is infinite.

As implied above, no general understanding of the role of stress in the stress corrosion mechanism is available to date. A number of proposals have been made however. The following sections briefly review some of these propositions to obtain an insight into what the stress corrosion mechanism entails.

Before ending the discussion of the factors governing stress corrosion cracking, it is noteworthy that some investigators claim that there are still other factors aside from those mentioned above. Factors other than those discussed are not generally accepted by a representative group from the metallurgical field, however.

In particular, temperature is indicated as a factor with some regularity. More likely, temperature should be considered along lines of a secondary factor. Its connection with corrosion revolves around the temperature dependence of general rate process reactions, and would be compared to the significance that time as a factor has been given in regards to the mechanisms.

Having the above material as a background, brief observations of some thoughts about specific stress corrosion cracking mechanisms are in order.

2. THEORIES OF THE STRESS CORROSION CRACKING MECHANISM

Introduction

After investigators agreed on factors that govern stress corrosion cracking, various theories of the mechanism involving these factors followed. Being engineers, as opposed to scientists, it may at first seem to be of little interest to study various specific theories of the cracking mechanism. However, a brief survey of some current thoughts about the mechanism will add to the fundamental understanding of this unusual phenomenon.

In the recent past, a considerable amount of confusion arose when trying to explain the stress corrosion cracking mechanism. Even at the present time there exists a large number of radically different theories. Of the theories presented many are based on limited research and the resulting explanations tend to be applicable to "special cases." However, about fifteen years ago general explanations began to be produced. It is with later general theories that this section is concerned.

Four theories are presented in the following material. These explanations are very similar in some respects yet are radically different in others. The reader should view each theory with one main question in mind, "What is the specific role of stress and galvanic corrosion for each of the theories presented?" Being able to grasp the role of the lesser mechanisms of stress and corrosion forms the basis for further stress corrosion cracking study.

The Electrochemical Theory of Stress Corrosion Cracking

Based on experimental acceleration of corrosion in aluminum alloys by application of high constant stresses, E. H. Dix and associates formulated the electrochemical theory of stress corrosion cracking.[7,8] According to this theory two conditions must exist for corrosion acceleration upon application of a constant stress: The metal must be an alloy susceptible to selective galvanic corrosion along more or less continuous paths; secondly, the stress must be acting in a direction tending to separate the metal along the paths. The susceptible areas may be grain boundaries anodic to the bulk of the material. Also, susceptible areas may be created by precipitations at grain boundaries, the precipitate being anodic to the matrix or the matrix depleted of alloy by the precipitation being anodic to the richer solid solution within the grains.

A preferential galvanic attack is begun at the surface providing "galvanic requirements" are fulfilled. The presence of the high stress combined with the stress concentration affected by the corrosion pit initially formed yields stress conditions great enough to cause the metal to separate, fresh film free (free from corrosion product film) metal being exposed. The exposed metal is anodic to the film covered material. With great rapidity, the corrosive environment migrates along the "crack" already formed and allows further galvanic action to take place at an accelerated rate. The further corrosion results in further tearing. The combined effects of corrosion and tearing result in accelerated propagation of the crack until failure. It should be noted that this explanation requires tensile stresses and an internal structure susceptible to galvanic corrosion, all taking place in a corrosive environment. The main role of stress is to expose the fresh material for galvanic corrosion to proceed.

Various researchers have shown that definite electropotential differences necessary for galvanic action exist between areas of certain material, especially in precipitation hardenable alloys.

There is opposition to the acceptance of the situation following precipitation in studies by Giesler[9] and Totten.[10] In general, however, there is much information to support the electropotential-electrochemical theory. One of the chief proponents is the fact that stress corrosion cracking, having started, has been arrested in various material by the application of cathodic protection principles using impressed currents.[8,11]

The Film Theory of Stress Corrosion Cracking

A second theory postulated is that localized paths for the initiation of stress corrosion cracking are caused by the local breakdown of corrosive films under the influence of stress or changes in environment, as well as being the result of composition differences within the material. This idea is quite similar to that of the elctrochemical theory presented above. The main point of concern here is the notion that the corrosive films formed at the grain boundaries are less protective than the films formed on the crystal faces. The reason for this is the fact that the irregular atomic distribution at grain boundaries yields films more easily broken compared to those formed along the crystal faces where atomic distribution is more regular. An independent researcher, Champion[12], feels that the real effect of stress in the whole mechanism is just prior to failure, that is after the part has been weakened by "cracking." The stresses do play a part in the crack propagation in that they are high enough, stress concentration being accounted for, to cause local plastic deformation of the metal, but do not "tear" the metal as previously hypothesized. The deformation ruptures the weak films and a film free metal to filmed metal cell is set up. Galvanic action takes place and new films are formed. Champion feels that the new films which might be stronger than the already present ones might prevent further cracking by stopping further galvanic corrosion, this arresting the pit formation at the crack apex where the pits easily affect a stress concentration. However, it is generally felt that the required stress concentration for additional film rupture is reached well in advance of the cessation of galvanic action, the result being the recurring of the plastic flow-film rupture-corrosion cycle until the part becomes so weakened that mechanical failure occurs rapidly. The overall outlook presented by this theory is that stress corrosion cracking is dependent upon the film forming properties and deformability of the metal. All of this implies that stress plays no direct part in crack propagation, not inclusive of the mechanical failure which results after crack propagation has progressed to a critical point.

Some British research has been done to demonstrate the importance of oxide formation in stress corrosion cracking. However, the extension of the film based theory is highly questionable, even to the point that it is doubtful that film formation occurs in some of the well-known examples of stress corrosion cracking. Another point to notice is that this theory does not apply to alloys exhibiting transgranular cracking although such cracking is quite common. In general, the film theory is not so well accepted as some others, especially in view of work done recently.

The Mechanical Theory of Stress Corrosion Cracking

Keating[6] has recently hypothesized that the primary function of stress in the stress corrosion cracking mechanism is confined to the development of mechanical damage such as tearing the metal apart. While supporting the electromotive theory for the initiation of stress corrosion cracking, he does not feel that the galvanic corrosion plays such an important part in the crack propagation as the localized stress conditions. Keating views the initiation of cracking as a build up of stress concentration by local galvanic action coupled with existing stresses. A critical stress concentration is reached for the local stresses to cause fine cracks to develop. The cracks will propagate thru the material until an "obstacle" is encountered, the "obstacle" reducing the stress concentration at the crack apex to the point that the stress concentration - stress couple becomes less than enough to plastically crack the metal. Such "obstacles" might be non-metallic inclusions, lattic discontinuities, or unfavorably oriented grain boundaries. At the "obstacles" crack propagation is arrested, awaiting the diffusion of the corrosive environment to the apex. During the arrest local corrosion may take place along the crack already formed, the corrosion following laterally branching paths. This lateral corrosion might explain the "branching effect" frequently observed in stress corrosion cracking. When the corroding medium reaches the apex, galvanic attack occurs until a stress concentration yields the critical stress level. Again the crack propagates until failure of the part or until another "obstacle" is encountered.

The electrochemical theory explains why the corrosion at the arrest points would be intergranular. It also seems probable that some aspect of the corrosion process accounts for the intergranular weakness although study to date does not "shed any light" on this notion. For some alloys the ambient temperature may be near the equicohesive point. This being the case, very slight changes in temperature would influence the mode of cracking, that is, intergranular or trangranular. The net result of Keating's work does not yield information that is drastically different from the electrochemical theory. The work does, however, indicate a new role of stress in the cracking mechanism as well as a new role of galvanic corrosion, these being the significant points.

The Strain Accelerated Decomposition Theory of Stress Corrosion Cracking

The mechanisms postulated above are based on the pre-existence of metallurgical conditions that render a material susceptible to stress accelerated corrosion, and, in general, the theories apply to polyphase material. On the basis of the aforesaid theories, it becomes difficult to explain stress corrosion cracking in "homogeneous" alloys such as alloy brass or some particular stainless steels which exhibit good resistance to galvanic corrosion in the unstressed condition but demonstrate serious susceptibility in the stressed state. A reknowned example of stress corrosion cracking unexplained by the above theories is the transgranular cracking of 18-8 stainless steels in acidic choride solutions, even when the alloy is in the sensitized condition which normally would yield intergranular cracking, if any at all. The depleted zone theory as implied in Dix's work was expanded by Waber[13] to a novel line of thought, that of strain accelerated decomposition of metastable phases to provide the dissimilar materials for the galvanic reaction. In his "Generalized Theory of Stress Corrosion" Waber asserts that the major function of stress is to generate "precipitation reactions" and order-disorder reactions setting up local anodic-cathodic conditions. In the cases of the pre-existence of the local anodic-cathodic conditions, Waber's theory reverts to that of Dix. A noted difference in Waber's work is that he feels the initial cracks are formed as a result of surface inhomgenuities (asperities, etc.) as opposed to galvanic action. Once formed, the cracks yield the now familiar stress concentration at the apex. In the region of high stress, metastable material decomposes at a rate greater than the bulk of the substance. On decomposition, one or more of the products is anodic to the parent material and galvanic action proceeds. The crack now propagates, "generating" new anodic material ahead of it. In other words, the high local stresses which set high strains accelerate the formation of galvanic microcells.

This theory is based on the assumption that the strains will accelerate the decomposition reaction of a single, metastable, apparently homogeneous phase. In general, the decomposition reactions are of two types: one in which a single reaction product is formed as an austenite to martensite reaction, or one in which two or more products result as in precipitation reactions.

Thermodynamics demands that any reaction which is spontaneous results in a net decrease of free energy for the system (entropy increases). This is the underlying principle of current thoughts regarding the precipitation hardening process.[9,10] Again referring to thermodynamics, the "First Law" apparently dissolves Waber's theory. The theory suggests that the products of the decomposition are anodic while thermodynamic principles in fact require that the products be cathodic with respect to the parent material. Waber does explain this apparent fallacy by suggesting that the corrosion occurring in the cracking mechanism actually takes place during the coherency stages of the decomposition reaction. At this point in the reaction there are definite areas anodic to the parent material by virtue of the increased strain energy produced by the stressing at the crack apex. For single phase materials, the martensite reaction being an example, the "metal in transition" and the undisturbed parent metal set up the galvanic microcells.

Some people may question the thermodynamic reasoning used in Waber's work, but great numbers of studies have been made on the metastable reaction with the result being a substantial amount of support for this line of thought. In countless tests it has been shown that a simple cold working process will cause virtually corrosion resistant material to fail most rapidly by accelerated stress corrosion. In the end, this particular theory more closely approaches a "blanket theory" for the stress corrosion cracking mechanism. In further work to be done, it will most likely be shown that the strain accelerated decomposition theory is that closest to the actual mechanism.

Summary

Having considered varied theories of the stress corrosion cracking mechanism, a general summary of pertinent points will suffice the theoretical background of the topic. As can be drawn from the theories, localized corrosion must occur "somewhere" in the process, most likely from composition differences, rupture of protective corrosion films, or crystallographic orientation differences within the material. The localized galvanic action produces pits or trenches, especially at the free surface of the material exposed to the atmosphere. Under the influence of stress, stress concentration at the apex of the notch formed by the galvanic action will yield considerable strain energy which promotes crack propagation. The propagation results from a) rupture of corrosion films, setting up galvanic action between filmed and non-filmed metal, b) accelerating the decomposition of a metastable phase, causing "migration" of galvanic microcells, or c) mechanical propagation, i.e., actual tearing of the metal underlying the notch.

The above statements represent but a small part of the general theories presented. However, it is possible to see from just these points that the role of stresses, though not "pin-pointed," is definitely one of the most important considerations in the stress corrosion cracking mechanism.

3. STRESSES INVOLVED IN STRESS CORROSION CRACKING

Introduction

The extreme importance of stresses in the mechanism of stress corrosion cracking has previously been indicated. Thus a brief discussion of the various "stress conditions" concerned with the mechanism will lead directly to methods of control of the phenomenon.

To begin, it is best to organize the discussion of stresses by categorizing the "stress types" to be considered. As a broad classification, the stresses may be divided into two groups, residual and operating. The residual stresses include stresses resulting from heat treat operations, manufacturing processes, or, in general, stresses that would be inherent in a part before its assembly with other members, and certainly not inclusive of the operating stresses resulting from the application of the part in service. The operating stresses may be simply described as those not included as residual stresses, but actually are encountered in the use of the part.

The following is a discussion of the origins of the various types of stresses classified as residual or operating. From the study of the origins of stresses, prevention of stress corrosion cracking approaches being a trivial matter.

Residual Stresses

The topic of residual stresses is itself an extremely complex subject. Although the origins of stress of the residual nature are varied, it is possible to classify all residual stresses into two types, macroscopic or microscopic,[3] or into body or textural as denominated by other factions. The specific origins of the various residual stress manifestations approaches infinity in number. To attempt a broad discussion of the various origins is beyond the needs of background material for stress corrosion cracking considerations. However, an observation of some common examples of residual stresses will aid in the general understanding of the stress corrosion problem. It should be noted at this point that in every case of residual stress, the cause of the retention of the stress is the occurrence of inhomogeneous plastic deformation.[14]

Macrostresses - Cold working of metal always results in some form of residual stressing. However varied stress patterns, depending on the type of operation, occur. The following figures are representative of longitudinal stress patterns plotted across the diameter of a two inch steel rod.

In general, if the working process is severe enough to cause deformation to occur throughout the cross section, the pattern of stresses that would be expected is that of Figure 1-b below, that is, tensile stress components at the surface. The classic example of season cracking in brass cartridge cases is predicted by the tensile stresses set up in the free surface fibers, especially at the casing corners. All the stressing in the cartridge cases is residual in nature resulting from the deep drawing operations in manufacture. (see Figure 2 on page following.)

Although the stresses resulting from working of the material are considerable, thermal effects causing non-uniform volume changes within a part are the major source of residual stresses. The thermal effects may be considered as 1. Thermal expansion on heating or cooling, 2. A change from one solid state to another solid state accompanied by a change in volume, and 3. A change in volume with a change in state, liquid to solid for example.

Casting of railroad car truck wheels demonstrates the thermal expansion-contraction effect sets up residual stresses.

In the manufacture of the truck wheels, one foundry used a chill in the tread region of the mold cavity (See Figure 3 on the page following.) to produce "white iron" in this part of the wheels.

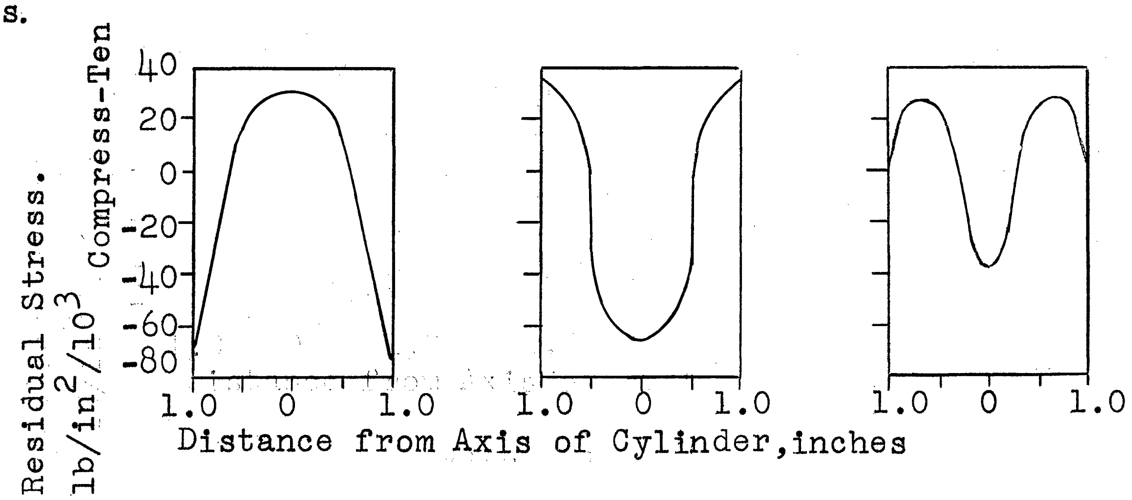


Figure 1

Representative patterns of longitudianl residual stress plotted across the diameter of a 2 inch steel rod. (a) Stress pattern produced by surface rolling; surface layers are in compression. (b) Stress pattern in cold-drawn steel; surface layers are in tension. (c) Stress in cold-drawn steel given a final surface rolling; surface layers in compression.[14]

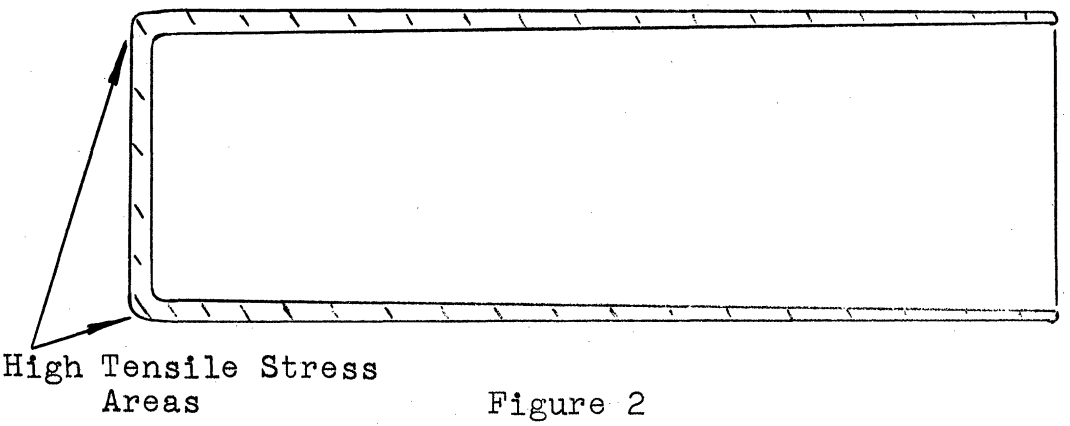


Figure 2

Cross sectional view of brass cartridge casing.

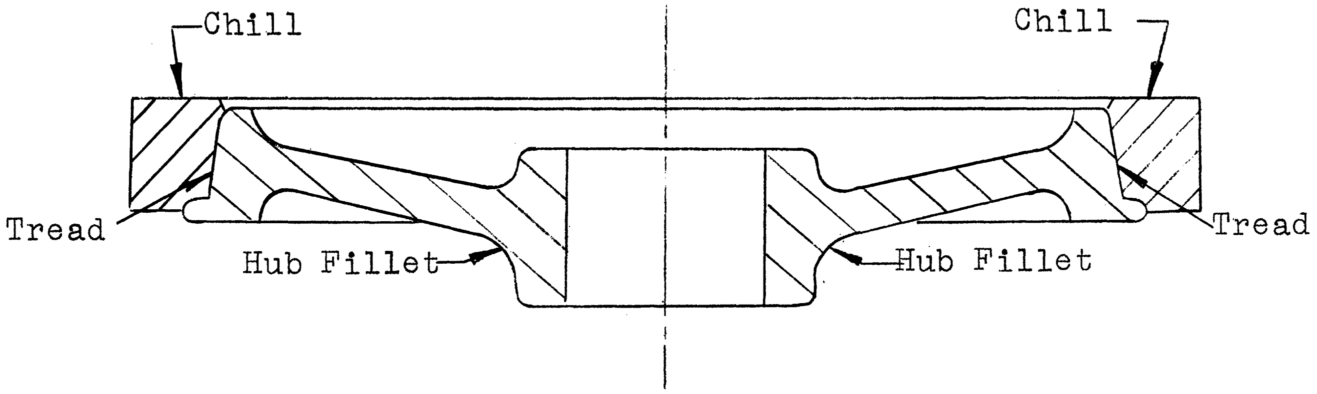


Figure 3. Cross sectional view of a railroad truck wheel.[15] In one manufacturer's process, residual stresses of the tensile nature, 32,200 psi in magnitude, were developed at the hub fillet. Accelerated cooling of the hub section by forced air flow resulted in residual stresses of the compressive nature, 27,000 psi in magnitude.

The chill sets up a rapid rate of heat transfer from the tread face after the liquid metal is introduced. The heat transfer is so rapid that a large thermal gradient is set up between the tread fibers and the inner fibers of the rim section. The outer fibers will cool to room temperature before the inner fibers, the inner fibers setting up compressive stresses in the outer fibers since contraction of the inner section is restricted. Since the rim section is chilled and cools at a relatively rapid rate, and since the web section is relatively small, the hub section, being large and unchilled, remains at a temperature greater than any other section of the wheel. Depending on the "hub cooling" used in a particular process, the hub can "lag" the other sections on cooling by many hundereds of degrees F. This means that the hub will attempt to contract from its higher temperature point to room temperature after the web and rim sections have already cooled to the ambient. The contraction will be restrained, tensile stresses being set up in the web section.

Depending on the "hub cooling" process used, the stresses in the hub fillet can range from 33,000 psi in tension to 1000 psi in compression.[15] The point to be noticed is that these stresses are considerable. In some cases the stresses are so great that the wheels actually will crack on cooling since stresses above the ultimate tensile strength are generated.

Being familiar with the thermal expansion-contraction effect yielding residual stresses, the origin of stresses resulting from solid state reactions becomes quite analogous. The most familiar situation which results in residual stressing in this manner occurs in steels through the austenite-to-martensite and austenite-to-ferrite plus carbide reactions. In the case of the austenite-to-martensite reaction there is a definite volume increase. Analogously, there is a larger volume increase in the austenite-to-ferrite plus carbide reaction. Thus, a large part of steel quenched from the austenite state will undergo two different volume changes dependent upon the metallurgical reactions that take place. Most generally, tensile stresses can be found in the martensite region while compressive stresses exist in the retained austenite and the ferrite plus carbide areas. Cracks denominated as heat treat or quench cracks are formed in this manner, resulting from residual stresses exceeding the strength of the material. Prevention of these stresses comprises much of the study done by process metallurgists.

The third type of thermal residual stressing results from the liquid to solid volume change on freezing of a metal. A simple example can best illustrate this phenomenon. Imagine a mold made to cast a one inch cube. Mild steel is introduced to the cavity (liquid metal), a skin or shell of the liquid freezes along the walls of the cavity. At this point the shell forms an impermeable container holding about one cubic inch of liquid metal. Eventually the liquid freezes with the volume change in the liquid to solid states being about a 3% loss, or decrease. This means that three hundredths of a cubic inch out of the total volume will be voids. These voids, found at the center of the cube, create partial vacuum in their formation. With the ambient atmospheric pressure and the partial vacuum in the center, the casting has compressive stresses at the surface fibers and tensile stresses in the "void fibers." It should be realized that this example is extremely simple, the whole volume change problem becomes quite involved with the increasing complexity of mold designs. Shrink voids, as these defects are called, present a major problem in mold design, often becoming the major cost in casting preparation.

Microstresses - The aforediscussed macrostresses have a characteristic nature depending upon part geometry, position in the part,etc., while microstresses are localized about certain microstructural features of the metal. Stresses across grain boundaries between adjacent grains resulting from differences in deformation behavior are a good example. Another example is the case of a solid phase precipitating within an existing solid phase, extremely high local stresses being developed. Many details of the microstress components remain to be studies. However, it is certain that these stresses are comparable in magnitude to the macrostresses and profoundly effect the behavior of the metals.

Closure - The information on residual stresses presented here is but a very brief discussion of some common cases. It should be noted that on a macroscale all types of stress patterns are possible; tension, compression, torsion, or combinations of components. On the microscale various local patterns of stress may also be developed. In magnitude the residual stresses may vary from nearly negligible to that of the ultimate strength of the material.

Operational Stresses

As indicated previously, operational stresses may be simply described as those stresses not considered as "residual" stresses. This is a gross simplification of the situation. However, this definition of operational stresses becomes more significant in discussing stress corrosion cracking if the stresses are limited to the constant or non-fluctuating variety.

Understanding operational stresses as being both non-fluctuating and non-"residual," it is possible to consider two types of operational stresses, fabrication and application. With these two subdivisions, cases of operational stresses observed in stress corrosions cracking are considered below.

Fabrication Stresses - As the name implies, fabrication stresses result from the assembly of units from one or more parts. It can immediately be seen that the source of constant non-fluctuating stress in assembly operations is infinite. A bolt having a constant tensile stress by virtue of its tightening torque is a simple obvious illustration. Also obvious is the fact that constant stress inherently results from press fitting assembly operations.

In Figure 4 (see following page) is a sketch of a bevel gear which has exhibited stress corrosion cracking. This gear was press fitted to a splined shaft. The root section of the spline affected a greater stress concentration that did other areas of the spline. With the press fit, inherent tensile components of stress existed in the root areas. The tensile components coupled with the stress concentration to provide the constant tensile stress necessary for stress corrosion cracking to take place.

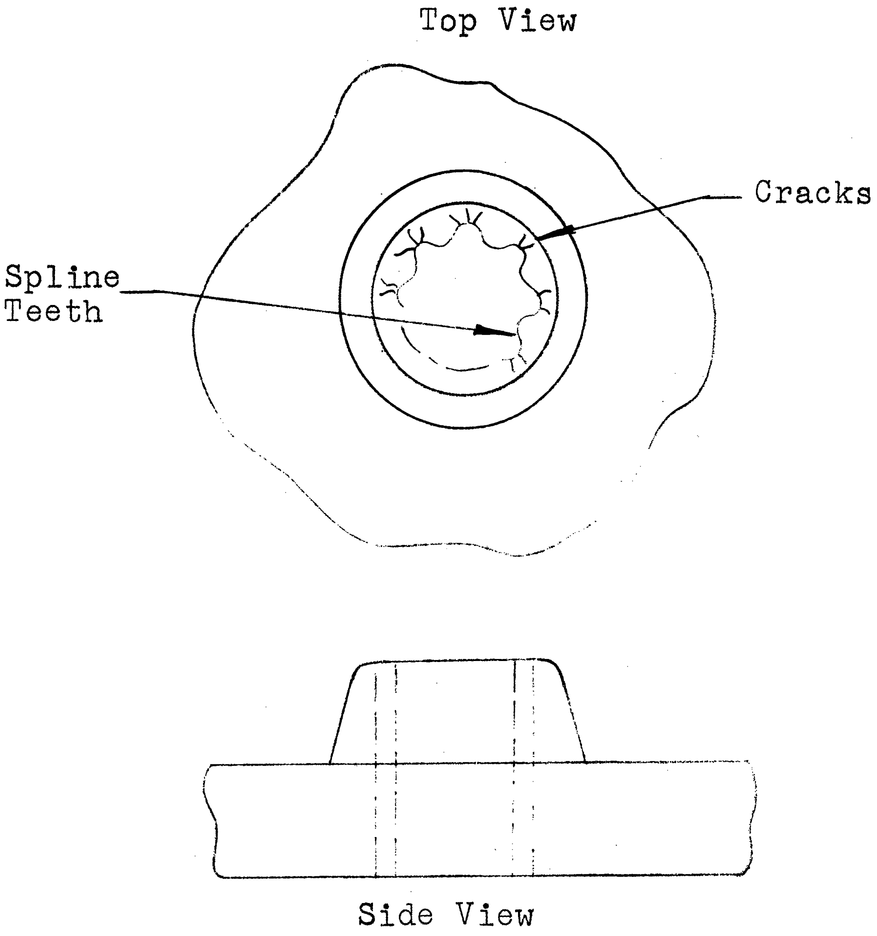


Figure 4

Spline-shaft section of a bevel gear. Note the cracks originating at the inner surface of the roots of the spline, propagating outward.

A second illustration of the origin of operational stresses may be observed in Figure 5. In this case, the large part is a casting which has a hole case in one wall to facillitate machining of internal fluid passages. After machining the interior regions, a tapered thread is cut in the cast hole so that a tapered plug can be inserted to seal off the fluid passage. Since the operation of the assembly includes high pressures in the fluid flow, the plug must be turned in until a high specified torque of tightening is reached. In this situation the plug is similar to a wedge being driven into a crack. The results now are obvious, the plug setting up tensile components of stress at the outer fibers of the casting shoulder. Under the proper ambient conditions, stress corrosion cracking set in most rapidly.

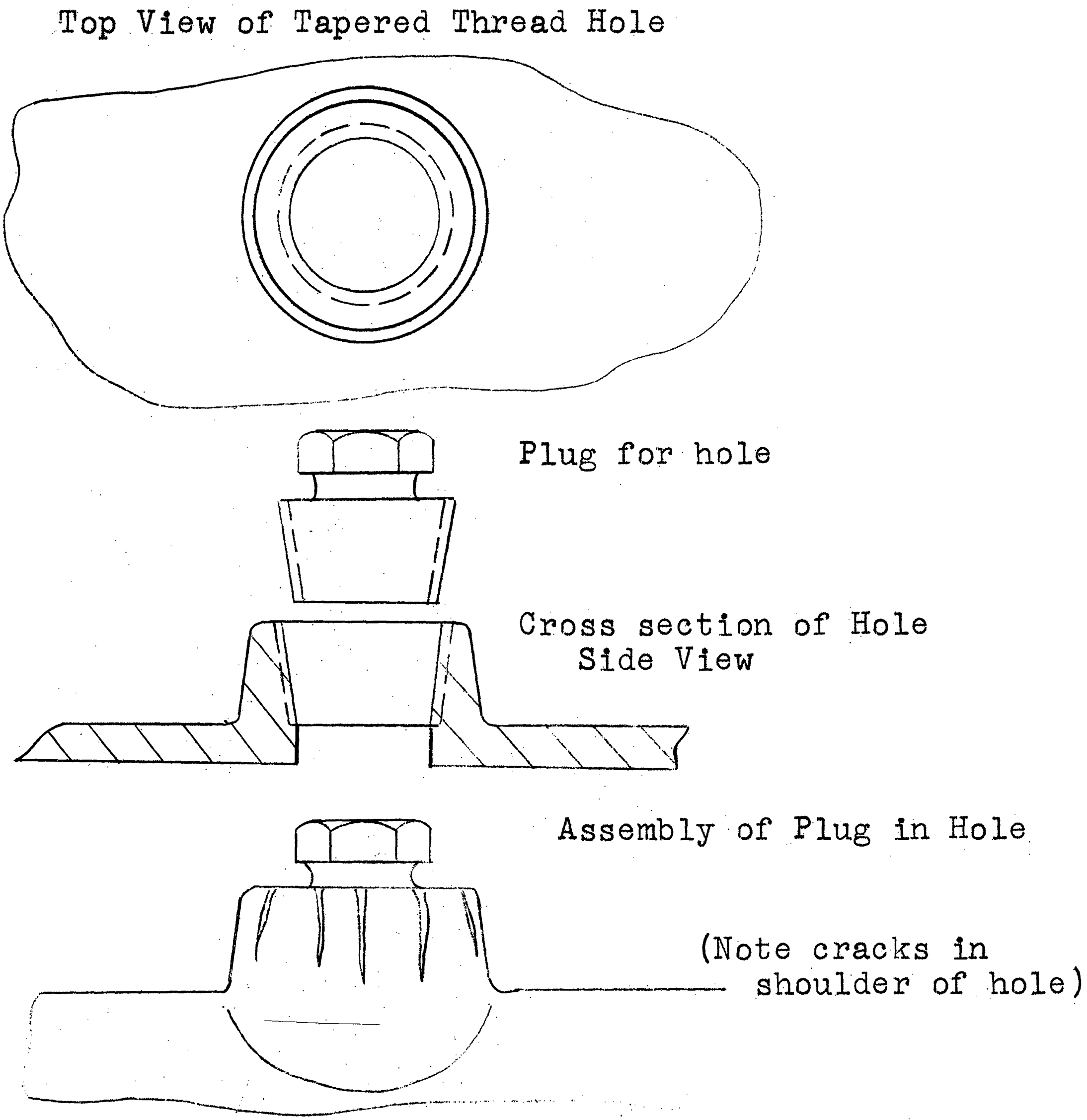


Figure 5
Three views of a plug-machining hole assembly in the wall along a fluid passage of a high pressure hydraulic system part.

Numerous other examples of fabrication stresses could be sighted. Field assembly of parts by welding offers literally thousands of situations for the required tensile stressing. In most cases of assembly, it is very obvious that constant tensile components of stress can be generated at free surfaces, all of which results directly from the particular assembly operations.

Application Stresses - Application stresses are those non-fluctuating stresses that result directly from the use of a part or unit. As was the case with fabrication stresses, infinite numbers of cases and situations afford the proper tensile components of application stresses to promote stress corrosion cracking.

One specific example for the sake of illustration can be observed in Figure 6 of the page following. This sketch is that of a fire extinguisher bracket. In operation, or use, the bracket is bolted to a wall and a heavy fire extinguisher tank suspended from it. Without any further comment, it can be said that tensile stress components occurred at the corner of the bracket, the components being great enough to promote stress corrosion cracking in that area.

Closure - In many cases of stress corrosion cracking, the cracking of a part is frequently attributed to some other phenomenon since the operational stresses appear to be so low. It must be realized that although the nominal operational stresses may be low, there exist many situation that afford stress concentration adequately enough to create a large acting stress from a nominally low operational stress. Also, it must be noted that any operational stress can be additive to the existing residual stresses. The addition of stress components frequently yields stresses of the magnitudes necessary for stress corrosion cracking where as neither the residual or operation stresses are "enough" alone.

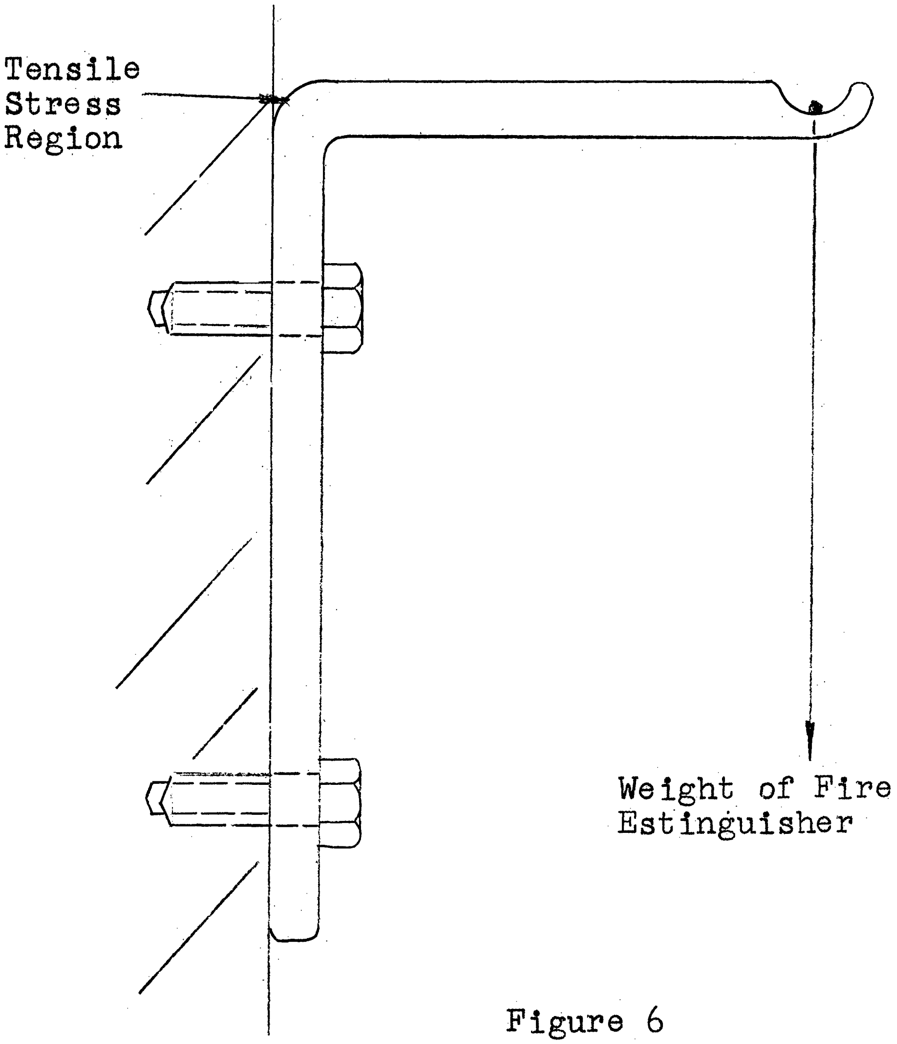


Figure 6

Assembly drawing of fire extinguisher bracket. Vector is representation of weight of fire extinguisher acting on the bracket.

4. THE CONTROL AND PREVENTION OF STRESS CORROSION CRACKING

Introduction

Understanding the information provided in the previous sections, control and prevention of stress corrosion cracking becomes the next logical subject for study. It is possible to attack this aspect of the overall study in a very complicated and sophisticated manner. However, a revertion to the original definition of stress corrosion cracking suggests that a less sophisticated study of stress control plus a similar study of basic galvanic corrosion will suffice.

Thus, the discussion of the prevention of stress corrosion cracking reduces directly to furthering the discussion of the general stress problems of previous sections and a subsequent study of corrosion control. By attacking the overall problem in this manner, it is possible to understand the prevention of stress corrosion cracking enough for the virtual prevention of the phenomenon in most situations.

Stress Considerations

In summary, the stresses of concern in stress corrosion cracking are either residual or operational. Again it is emphasized that various concurrent stress components may be additive or subtractive at any given point within a part. In many cases single stress components are not enough to promote stress corrosion cracking failure, but the addition of another stress component results in rapid failure. It then seems logical to think that removal of a single stress component would efficiently control the stress corrosion cracking rate, perhaps to the point that it could be said that stress corrosion cracking failure had been prevented. Keeping this in mind, brief studies of various stress condition control measures are of aid.

Residual Stresses - When faced with the problem of residual stresses, many engineers will proclaim "stress relief heat treatment" as the "cure all." That so called stress relief operations are in fact "cure alls" for many cases is well documented. However, in connection with stress corrosion cracking, the use of stress relief heat treatements demands some scrutiny.

That stress relief treatments generally relieve residual stress is not disputed. Since this is of primary interest for the case at point, various engineers overlook the extremely deterimental effects that may result from such treatments. In many cases the stress relief heat treatments produce an extremely corrosion sensitive material.[16] It is obvious that the purpose of the whole scheme of stress corrosion cracking prevention is lost by providing the "ideal" preferential paths for the cracking to follow although the residual stresses are in fact relieved. This is an example of the mysticism that hampered understanding of the prevention of the cracking mechanism some years ago. It was understood that elimination of residual stress should reduce stress corrosion cracking failure. When stress relief treatments promoted stress corrosion cracking instead of preventing it, much confusion resulted. Later theories of the mechanism removed much of the confusion.

From the above it should be obvious that some mechanical method of removing stress, specifically tensile components at the free surface of parts, would be highly advantageous. The mechanical method relied upon quite heavily with a high degree of success is that of shot and/or hammer peening. The peening operations deform the surface material, setting up compressive stress components or,at least, "cancelling" out the pre-existing tensile components. Without the tensile components at the surface of the metal, initiation of the stress corrosion cracking becomes difficult. Graphically, the results of a peening process can be qualitatively seen in Figure 1,c.

Variations of the above mentioned stress control methods form the used procedures for reducing the residual stresses connected with stress corrosion cracking. Other less popular processes are used with a high degree of success also. Nonetheless, the stress relief by heat treat operations or mechanical inducement of compressive surface stress remain the most generally successful methods.

Operational Stresses - Since the origin of operational stresses is so vast, it is not possible to specifically outline general prevention procedures to be employed. It is useful to consider fabrication and application stresses as separate entities. Again, not too much specific can be said for the two cases, but a few points of merit can be brought out.

As was pointed out previously, fabrication stresses result primarily from the assembly of parts into units. Information for the control of stress leading to stress corrosion cracking is usually based on experience from specific assembly operations and specific units. However, it is possible to apply a few generalities in some cases.

The familiar stress relief heat treatment operations discussed already find use, on a reduced scale, in the stress relief of assembled units. If the stresses are small scale, quite satisfactory results can be obtained using this sort of procedure. However, a blanket statement as to the success of heat treating is not possible, "trial and error" work being necessary to determine the applicability of heat treating to a specific situation.

If a particular assembly is critical in its application, implying that stress corrosion cracking cannot be tolerated, total or partial redesign of a part or unit may be necessary. In many cases something so simple as a "better" tolerance of a critical dimension will solve an apparently complex problem. A bolt, for example, through two mating members will stand less chance of necessarily being assembled by "force" if the holes are more accurately located.

Also, very simple changes aside from redesign may be sufficient. In the case of the plug and hole in a casting wall (Figure 5), a sealant applied to the plug might have satisfied pressure specifications of the unit, and the plug would not necessarily need to be tightened to the degree previously.

As stated above, generalities in fabrication stress control can not be made. One widely used approach is to check design information coupled with observation of product experience. This may lead to many varied solutions, extremely simple or extremely complex, but the problem most likely can be handled.

Considering application stresses, the most direct approach is to consider the actual service stresses on a given unit. Quite frequently design of a unit is altered when the stresses are considered to be "too high" in regards to stress corrosion cracking. In the case of the fire extinguisher bracket (figure 6) a change of section thickness at the corner reduced the service stress to a sufficiently low level.

Quite common in design work is the use of a "2 X" factor for the stress corrosion or galvanic corrosion considerations.[17] In the case of cast iron parts, it is quite common to double the thickness of a given section size, the original size being adequate for strength considerations. The doubling of the section "accounts" for the stress corrosion. It does not prevent stress corrosion as such, but it does "slow" it to a rate low enough that the part is considered as "safe" for a given life.

Again, the major generality that can be made regarding stress control, and its effect upon stress corrosion cracking reverts to experience with a specific part. In many situations the application stress appears to be "insignificant." Again the additive character of various stress components should be considered. An "insignificant" application stress may be the "straw that breaks the camel's back."

Galvanic Corrosion Considerations

It would appear that the galvanic corrosion considerations regarding stress corrosion cracking should not differ greatly from the approach taken for any ordinary galvanic corrosion problem. To a degree, this is the case.

For the combating of corrosion in any stress-less corrosion situation the following points are frequently indicated as the preventive methods available:[17]

1. Employment of alloyed materials, or selection of generally more corrosion resistant materials.
2. Use of metallic, inorganic, or organic coatings to exclude corrosion media.
3. Mechanical removal of the corrosion media.
4. Chemical steps utilizing inhibiters.
5. Use of substantial cathodic protection.
6. Design changes to incorporate the use of non-metallics, e.g. glass, rubber, etc.
7. Refinement or purification processes in the production of metals to provide more homogeneous structures.
8. Improvement of equipment design.

In general, the above steps do succeed in corrosion protection for many stress-less corrosion cases. However, one point should be noted in considering stress corrosion cracking aspects: The cracking mechanism, corrosion inherently playing a major role, quite commonly occurs in materials or situations normally classified as "absolutely" corrosion resistant. Stress corrosion cracking in stainless steels is a prime example, certain stainless steels being known as virtually corrosion resistant.

The obvious question then becomes, "What is to be done?" Of the above mentioned preventive measure, #8, improvement of equipment design, becomes a most important method of attack. In practice, components of units should be designed on standardized bases if stress corrosion cracking is expected to be a problem. Standardization, as referred to here, means simply to design in such a manner that a variety of materials or vendor components can be used for a given application. By having the standardization of design, it is possible to "experiment" for proper prevention of stress corrosion cracking rather than attempting to solve the problem theoretically. A part of stainless steel, for example, may be highly susceptible to stress corrosion cracking once put in service. An equally satisfactory (satisfactory from design criteria) cast iron part could be substituted and the result would be a virtually stress corrosion cracking resistant assembly.

Returning to the specific galvanic corrosion aspect of the stress corrosion cracking problem, it appears that removal of the corrosive media is an "easy" solution. The application of coatings and the like to critical parts immediately appears to be an answer. It must be pointed out that some of the stressless corrosion preventive methods, coatings being an example, do deserve consideration. It should also be noted that these methods can in fact result in extremely dangerous consequences. Copper coatings of parts, a frequently employed process, is an example of a "dangerous" method. Once a copper film becomes discontinuous, an extremely good possibility of such existing under the application stresses, pitting will start and incipient stress corrosion cracking begins.[17]

As with stress considerations, it also is true of galvanic corrosion considerations that general solutions regarding stress corrosion cracking do not exist. Again the most reliable guide post is the backlog of information from specific cases of stress corrosion cracking failure. Enough information exists that it becomes possible to observe "trends" rather than rules in certain situations. Following the trends with actual testing for specific cases allows one to design with a certain degree of confidence.

Summary

In summary, it is not necessary to repeat that which has already been stated above. Consistently through the information presented, the fact that stress corrosion cracking is a combined effect has been emphasized. That is the single most important point to be remembered when regarding prevention of the cracking mechanism.

Also, it is well documented that stress corrosion cracking does occur when the total stress condition is relatively low and galvanic promotion conditions are equally poor. Likewise, it is well known that stress corrosion cracking can be virtually eliminated in most service situations. The point to note is that stress corrosion cracking can pose a formidable problem, but that the problem can be controlled and contained to a suitable degree.

BIBLIOGRAPHY

1. "Definition of Metallurgical Terms," Metals Handbook, American Society of Metals, Cleveland, p 14, 1948.
2. W. C. Roberts-Austen, Proceedings of the Royal Institute, London, v 11, p 395, 1884-1886.
3. J.J. Harwood, "The Influence of Stress on Corrosion," Corrosion, National Association of Corrosion Engineers, v 7, September 1950, p 290.
4. E.H. Dix, "The Stress Corrosion of Aluminum Alloys,"Metals Handbook, American Society of Metals, Cleveland, p229, 1948.
5. A. Morris, Trans. of the American Institute of Metallurgical Engineers, v 89, p 249, 1930.
6. F.H. Keating, "Symposium of Internal Stresses in Metals and Alloys," Institute of Metals, London, p 311, 1948.
7. E.H. Dix, Trans. of the American Society of Metals, v 137, p 11, 1940.
8. R.B. Mears, R.H. Brown, and E.H. Dix, Symposium on Stress Corrosion Cracking of Metals, American Institute of Metallurgical Engineers, 1944.
9. A.H. Goesler,"Precipitation from Solid Solutions of Metals," Phase Transformations in Solids, John Wiley & Sons, New York, p 423-426, 1951.
10. H.C. Totten, "The Phenomenon of Precipitation Hardening," University of Michigan, Ann Arbor, p 23, 1959.
11. H.J. McDonald, "Cathodic Protection, a Symposium," National Association of Corrosion Engineers, p 192, 1949.
12. F.A. Champion, "Symposium of Internal Stresses in Metals and Alloys," Institute of Metals, London, p 468, 1948.
13. J.T. Waber, "A Generalized Theory of Stress Corrosion," American Society of Metals, 1949.
14. A.G. Guy, Elements of Physical Metallurgy, Addison-Wesley, p 299-305, 1959.
15. R.A. Flinn, Modern Cast Metals Practice, University of Michigan, Ann Arbor, p III 19-III 23, 1955.
16. U.R. Evans, "Stress Corrosion-Its Relation to other Types of Corrosion," Corrosion, National Association of Corrosion Engineers, v 7, July 1955, p 238-244.
17. Fontanna, "Stress Corrosion and Corrective Measures," Metal Progress, v 53, June 1948, p 838-840.

CHAPTER 16

DAMAGE BY ABRASION

1. DEFINITION OF ABRASIVE WEAR

Abrasive wear could be defined as wear that occurs at normal temperatures as a result of dynamic contacts between metallic wear surfaces and abrasive particles or fragments. The highest wear rates occur, of course, when the hardness of the abrasive is considerably greater than that of the abraded. Examples of common abrasive materials would be coke, gravel, dirt, scales, ore, and sand.

Because toughness will be discussed many times in this chapter, let it be defined as being an indication of the amount of plastic deformation a material can withstand before fracture occurs. Thus, in general, the larger the difference between the yield and tensile strength of a material the tougher it is. It should be realized that toughness is an indefinite term and hard to define. Certainly, this definition is open to dispute, but for the purpose of the chapter is adequate.

2. TYPES OF WEAR

Wear by abrasion can be divided into three major categories, namely gouging abrasion, high stress or grinding abrasion, and low stress scratching or erosion abrasion. All three may occur simultaneously on a wearing part, but usually the type which predominates can be recognized. It is important for the engineer designing for an abrasive wear situation, to be able to identify the type of wear so that he can make the proper selection of material or surface for maximum wear resistance. Although, the general method for abrasive wear resistance is through control of the hardness of the material, it will be shown that other facts such as microconstituents and their distribution are more effective methods in certain types of wear.

Gouging Abrasion

This type of wear is characterized by high stresses (usually caused by impact), and results in considerable micro-deformation of the surface. Another important characteristic is that the abrasive usually has good support so that its gouging action can be compared with that of a machine cutting tool. Sometimes these gouging forces are applied at relatively low velocity, as is the case of a shovel dipper digging into rock pile, while in other cases, they may be applied at high velocity, as in the case of the hammers or breaker bars in an impact type pulverizer.

Abrasive wear rates are usually discussed in terms of mils per hour which is the rate of metal removal normal to the surface of the material. This type of wear usually has wear rates that tend to be higher than for grinding or erosive wear (except for erosion at high velocities). This fact can be observed in Table 1 which gives the wear rates for all three types of wear and also further examples. These rates depend a lot upon the character of the ore being handled. The highest values given are for extreme rates. It should be realized that when the wear rates are excessively high, such as in the hammer pulverizer, the machine is usually not operated at this extreme.

In general, the hardness of a material is the chief control of wear resistance for abrasive wear, or, for that matter, of any type of wear. However, hard materials are brittle and fracture easily under impact. Since gouging abrasion is usually accompanied by impact, the practial choice frequently involves a compromise selection in which a material is chosen which has less than optimum abrasion resistance combined with adequate impact strength.

Table 1[1,2] gives a list of commonly used materials for gouging wear resistance. It should be noticed that the most wear resistant materials are at the top of the list and increase in toughness towards the bottom. The hardest materials are usually used as inserts on tough steel backings. Their application would be suitable, for example, in rock pulverizers where wear rates and impact forces are excessively high. The abrasion resistant alloys are normally used in those applications where white irons are unsuitable because of insufficient toughness. It should be remembered that the tougher alloys may also have better wear resistance than nickel-chromium, martensitic white irons for certain applications and degrees of impact.

One of the more commonly used materials for this type of abrasion is austenitic manganese steel. Its important properties[1,3] are its high toughness, its ability to work harden under impact, freedom from high residual stresses in large casting, and ready availability. It lends itself well to weld repair or the building up of hard facings. However, since it work hardens it is also hard to machine and so is usually only

	Wear Rate* Mils per Hour
1. Austenitic Manganese Steel in Gouging Abrasion	
a. Hammers in impact pulverizers	5 to 1000
b. Shovel dipper teeth	5 to 500
c. Wearing blades on coarse ore scrapers	5 to 100
d. Ball mill scoop lips	3 to 15
e. Crusher liners for crushing siliceous ores	2 to 20
f. Chute liners handling coarse siliceous ores	0.1 to 10
2. Low-Alloy High Carbon Steel in High Stress (Grinding) Abrasion	
a. Rod and ball mill liners in siliceous ores	0.5 to 5.0
b. Grinding balls in wet grinding siliceous ores	0.15 to 0.45
c. Grinding balls in wet grinding raw cement slurries	0.05 to 0.15
d. Grinding balls in dry grinding cement clinker	0.005 to 0.015
3. Pearlitic White Iron in Low-Stress Abrasion (Erosion)	
a. Sandblast nozzles	100 to 1000
b. Sandslinger liners	50 to 250
c. Pump runner vanes pumping abrasive mineral slurries	0.1 to 5.0
d. Agitator and flotation impellers in abrasive mineral slurries	0.05 to 1.00
e. Screw type classifer wear shoes in sand slurries	0.05 to 0.20

* Wear rate is based on rate of metal removal normal to the wearing surface of the part.

Table 1

Typical Rates of Wear in Various Types of Abrasive Wear[1] applicable to casting. Under certain conditions that would give an excessive amount of plastic flow, 2% molybdenium austenitic manganese steel, which has a higher yield strength, provides even better wear resistance.[2]

High Stress or Grinding Abrasion

This type of abrasion occurs when two wearing surfaces rub together in a gritty environment with sufficient force to produce a crushing action in the mineral particles or other abrasive entrapped between these two surfaces. The normal loads on the surfaces are usually low. However, since the actual supporting area (the grains) is considerably less than the area of load application, the stresses on the individual grains are exceedingly high. One experimenter[1] estimates the value may be as high as 300,000 psi. These high stresses are capable of causing microspalling or fracturing of brittle constituents such as coarse carbides which may exist in the structure of some wear resistant alloys. Since the compressive crushing strength of even the hardest minerials is relatively low, (for example, for quartz, CS = 30,000 psi) the grains are easily fractured. The broken abrasive grains are sharp and can scratch the hardest type of steel. Deterioration then occurs from scratching, local plastic flow and microcracking. Thus even though the abrasive originally occuring between the surfaces may be rather smooth (and seemingly relatively harmless), it is soon broken into a cutting, destructive abrasive.

Referring back to Table 1, it can be seen that the wear rates for grinding abrasion in mile per hour tend to be relatively low. However, since large surface areas are involved, the wear rate in pounds per ton of material ground is the important criterion and is usually quite high.

Norman, on the basis of his experience and various tests lists in Table 3 the steel matrixs in the order of decreasing wear resistance. He feels that the structure of the matrix has a much greater influence on relative wear resistance than the distribution or quanitity of any primary or secondary carbides which may be present.[1] The wear resistance of a high carbon, austenitic matrix may vary from that of martensite to that of soft pearlite depending upon heat treatment, alloy, and carbon content.

Norman also gives relative wear rates for grinding mill liners (Table 4) and this chart can be used as an indication of the type of steels to be selected for grinding abrasion. It should also be noticed that the wear rates increase with decreasing hardness.

Low Stress Scratching (Erosion) Abrasion

The primary factors causing erosion wear are velocity and low impact with motion being parallel to the surface. It ordinarily depends on hard and usually sharp particles for effective wear; however, the chief source of stress is velocity. The abrasive particles may be suspended and carried in a fluid such as air or water, or flow by its own weight (such as in a sand chute).

				Structure
Increasing Abrasion Resistance	1.	Tungsten carbide composites	Properties Sensitive to Carborn Content	Maximum abrasion resistance Worn surfaces become rough
	2.	High-chromium irons		Excellent erosion resistance Oxidation resistance
	3.	Martensitic iron		Excellent abrasion resistance High compressive strength
	4.	Cobalt base alloys		Oxidation resistance Corrosion resistance Hot strength and creep resistance
	5.	Nickel base alloys		Corrosion resistance. May have oxidation and creep resistance
Increasing Toughness	6.	Martensitic steels		Good combination of abrasion and impact resistance
	7.	Pearlitic steels		Inexpensive. Fair abrasion and impact resistance
	8.	Austenitic steels		Work hardening
	9.	Stainless steels		Corrosion resistance
	10.	Manganese steel		Maximum toughness with fair abrasion resistance. Good metal to metal wear resistance under impact

Table 2

Alloys for Abrasion Resistance[2]

Increasing Wear Resistance	High carbon martensitic matrix
	High carbon pearlitic matrix
	Bainite
	Soft Pearlite
	Pearlite and ferrite
	Low carbon ferrite
	*High carbon, austenitic matrix may vary from martensite to soft pearlite

* See the paragraph discussing this table

Table 3

Matrixs for Grinding Abrasion Resistance[1]

	Material	Hardness* R_c	Relative Wear Rate
1.	High chromium white iron (15% Cr 3% Mo)	66	89
2.	High chromium white iron (27% Cr)	64	98
3.	Martensitic 1% carbon Cr-Mo steel**	55	100 (Std)
4.	"Martensitic" Ni-Cr-Mo white iron, chill cast	59	107
5.	Martensitic 0.7% carbon Cr-Mo steel	58	111
6.	"Martensitic" Ni-Cr white iron, chill cast	55	116
7.	Martensitic 0.4% carbon Cr-Mo steel	55	120
8.	Pearlitic 0.8% carbon Cr-Mo steel	39	127
9.	Austenitic manganese steel	48	195
10.	Pearlitic low chromium white iron	48	195
11.	Normalized 1021 steel	$65R_b$	225

Table 4

Relative Wear Rate of Some Materials for Grinding Mill Liners in High Stress Abrasion[1]

* Hardnesses from the average of Rockwell hardness radings on the worn surface after service in the ball mill. On materials with austenite in their structure, this surface hardness is usually substantially higher than the hardness of the material below the cold worked surface.

**This "standard" material contained 1.0% carbon, 0.8% manganese, 6.0% chromium, 1.0% molybdenum and was hardened by an air quench from 1900°F.

The stresses involved with erosive wear are only occasionally great enough to break the abrasive grains. This means that the original smoothness or angularity is important since there is little change with use.

If impact is very low, the best criterion for wear resistance is to use the hardest steel possible with a high carbon content; the latter factor gives a high proportion of hard carbides in the structure. Figure 2 is a general curve of hardness vs. carbon content and it can be seen that a high carbon content is necessary to obtain a high hardness. Figure 1 illustrates the importance of using high carbon contents to improve relative wear rates (lower abrasion factors). Carbon content preferably should be in the white iron region of composition for ferrous alloys (1.5-2.0%).

In the presence of higher impact, a tougher steel is recommended (Refer to Table 2) depending upon the magnitude of impact. Table 5 lists various erosion resisting steels. It should be noted that hardness is a good indication of wear resistance, but that high carbon content appears to be somewhat more important. It has been found that pearlitic matrixs with a high carbide content provides better wear resistance than a low carbon content martensitic matrix. Quanitity, distribution and hardness of the carbides control the wear for abrasive particles harder than martensite.[6]

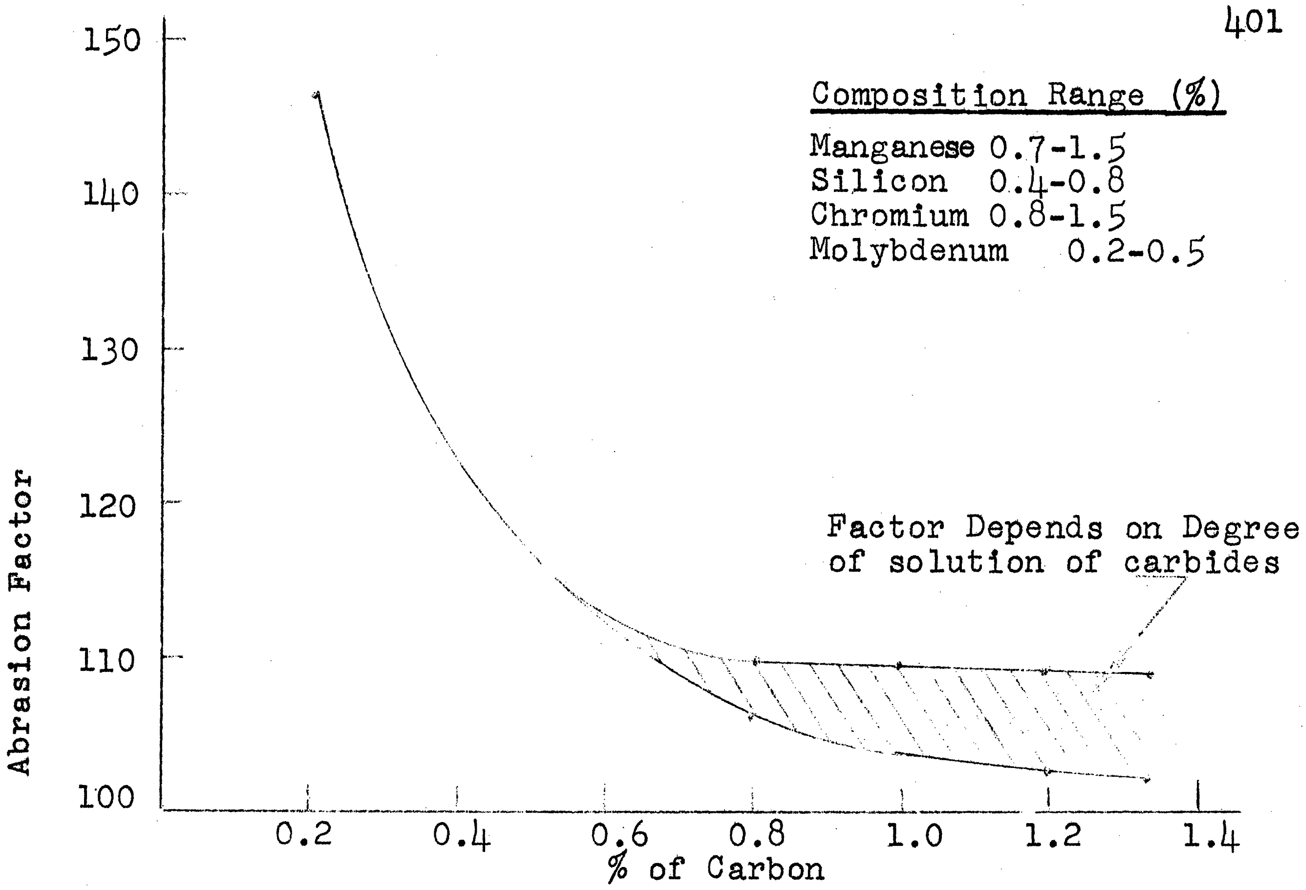


Figure 1

Influence of carbon content on the wear rates of martensitic cast steels when grinding a siliceous ore.[1]

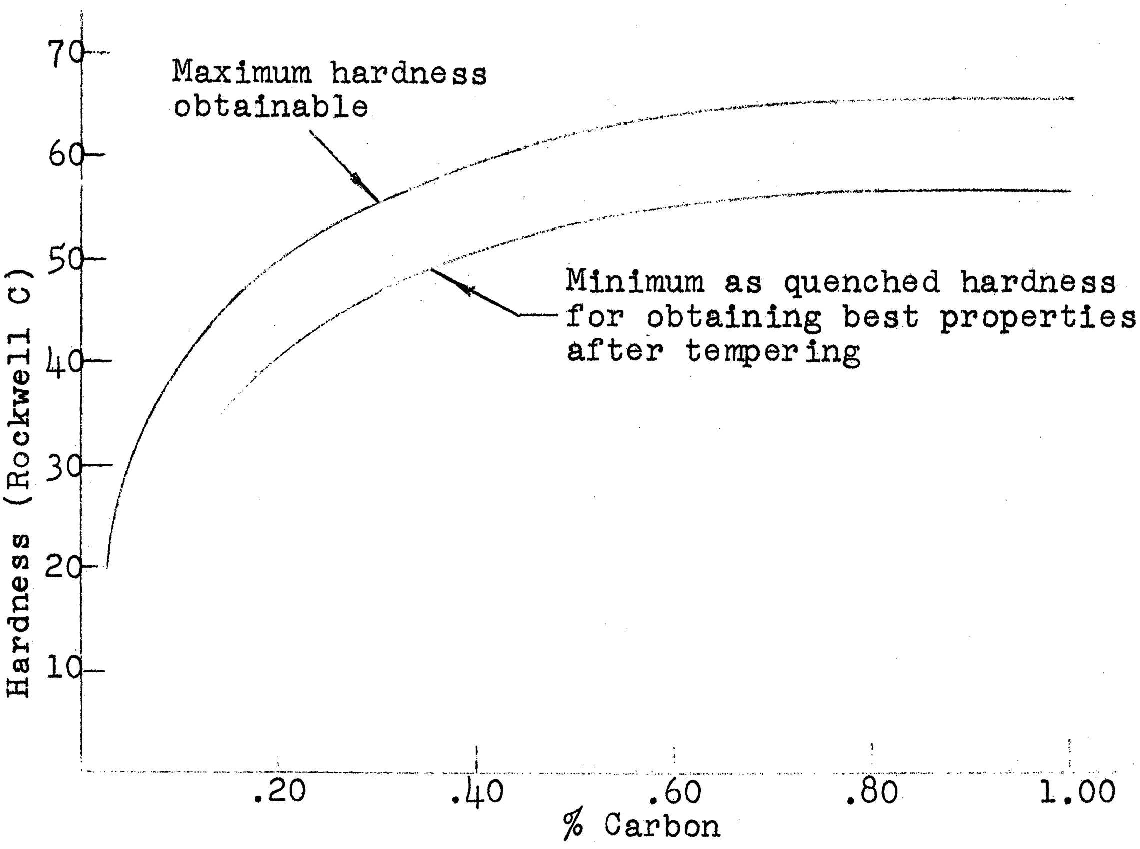


Figure 2

As-Quenched Hardness for best properties after tempering curve and maximum hardness curve[3]

Technique	Applicable To	Effect
Electroplating Chromium Nickel	Most ferrous and non-ferrous materials	Hard, smooth, usually thin surface layer Degree of porosity can be controlled.
Anodizing	Aluminum, magnesium, zinc	Hard, thin oxide layer
Diffusion (impregnation) Carburizing Carbonitriding Cyaniding	Steels only (Usually low carbon steel)	High carbon or high nitride (or both) case; surface hardness
Spraying Metallizing Ceramic spraying Flame plating	Metallic and non-metallic base materials	Layer of interlocked particles usually part oxidized, and porous
Hard facing Gas welding Arc welding	Ferrous and non-ferrous metals	Heavy alloy or carbide layer with properties depending on the materials deposited
Chill casting	Gray iron, some steels	White iron surface
Surface hardening by flame or induction	Iron and steel	Local surface hardness, depending on the area heated, time, temperature, quench and material.

Table 5

Surface Changing Techniques for Abrasion Resistance[7]

Rubber For Erosion Abrasion Resistance

One of the more frequently used materials used for the handling of highly abrasive particles in this type of wear is rubber. The abrasion of rubber is governed by four factors,[4] namely, the tear strength (The result of cross-linking and the presence of an active filler); elasticity (again a result of cross-linking and an active filler, and influenced by the molecular structure of the rubber); the heat conducting ability of the compound used; and the avoidance of degradation especially by oxidation.

It has been concluded by researchers, that under low impact the individual chains of rubber are capable of accepting and storing a considerable amount of energy, the energy of impact being spread over many carbon-carbon bonds along the chain.

If however, the velocity of impact is very great, there is insufficient time for the energy to be shared along the chain and the maximum strain might fall entirely on the single bond between two carbon atoms. In such circumstances the plastic properties of rubber disappears and it becomes a harder and more rigid material, which would be expected to be relatively brittle and to have a much lower abrasion resistance. Thus for application of low impact, rubber is a good material to consider for wear resistance.

3. INFLUENCE OF THE ABRASIVE MATERIAL

It is generally accepted that for a material to be scratched or appreciably worn, the abrasive must be harder than the material.

To a large extent this appears to be true, though a small amount of wear still occurs even when relatively soft minerals rub against very hard wear resistant materials. There is, however, usually a very marked drop in the wear rate when the hardness of the abrasive mineral is less than that of the material which it is abrading.

It can be seen from Figure 3[1] that in grinding quartz, which is approximately as hard as the martensitic steel and definitely harder than pearlitic steels, the wear rates of the three steels are relatively close to each other, with a spread of only 25% between the hardest and softest steel. In grinding feldspar, which is definitely softer than the martensitic steel, but harder than the pearlitic steels, a spread of over 100% exists between the two types. This is due to the marked drop in wear rate of the martensitic steel when it is grinding a mineral softer than itself. In grinding calcite, which is softer than all three steels, the wear rates of all three are very low, though there is still a spread of over 100% in relative wear rates of the martensitic and pearlitic types.

However, in many cases, the hardness of the abrasive is not its most important property for high abrasion. In Figure 4[2] the wear rate for two minerals is plotted. As shown, both minerals have the same hardness and composition but the tougher mineral, chert, has a much larger effect on the wear rate. The tougher minerals are able to withstand fracturing under load, (such as in grinding abrasion) while the harder particles are reduced in size. It has been found by several researchers[2,5]

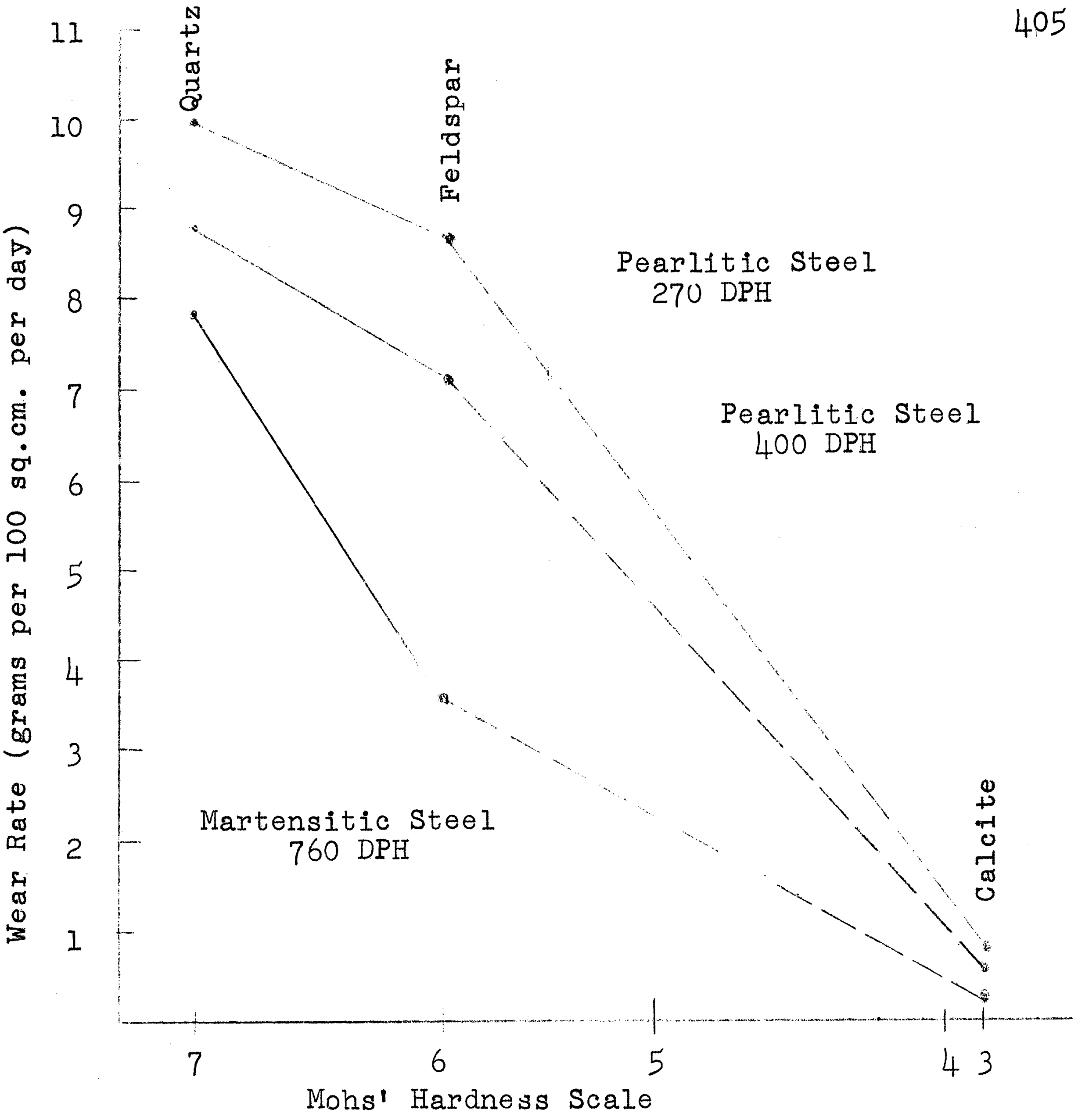


Figure 3

Influence of hardness of the abrasive mineral on the wear rate of 0.80% carbon steel grinding balls at three hardness values[1]

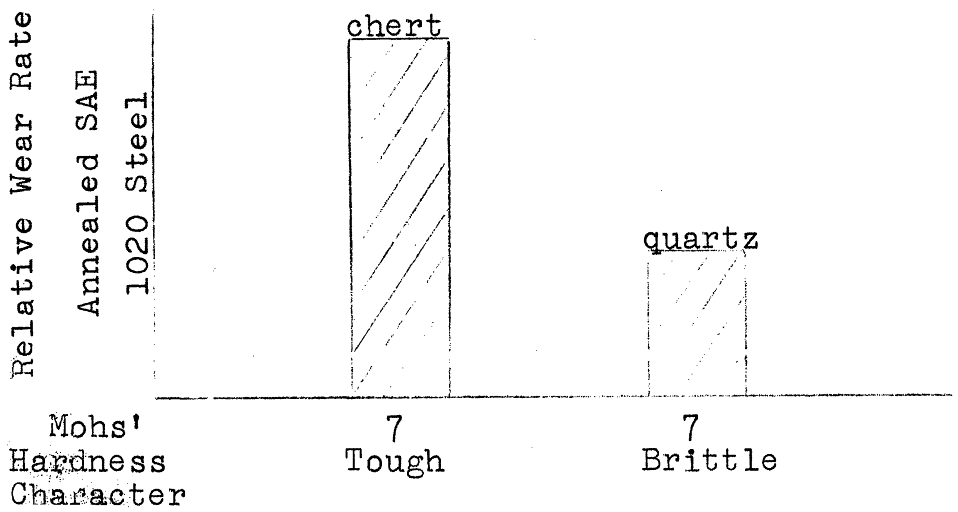


Figure 4
Relative abrasiveness of two mineral abrasives

that the wear rate decreases with a decrease in grain size. This was found to be particularly true in harder alloys that could also resist and fracture the grains. Another mineral characteristic that seems to have a greater influence than hardness on the wear rate is angularity. (That is, the number and sharpness of the cutting points on a grain.) Again harder alloys tend to increase fracture of the cutting points and reduce the wear rate.

One researcher[6] has pointed out that for dry abrasion resistance, (particularly for rapid abrasion), the hot hardness of the aggregate may be as much or more of a governing factor than the room temperature hardness. Therefore the latter may be of limited usefulness in predicting the behavior of materials in this type of service.

4. IMPORTANT FACTORS

So far, this discussion has dealt with the methods of avoiding abrasion. However, there are other considerations that require experience and intelligent judgement by the design engineer. Some design factors to be considered are strength, machinability, and corrosion resistance (important for abrasive carrying corrosive fluids). A high first cost may be justified by a long wearing life and low frequency of replacement with a consequent reduction in downtime costs, while a low first cost may be desired for a competitive product. Availability of the material to be used is important in quantity production.

Another consideration is consistent performance which, of course, is necessary for a quality product and reliable machine performance.

5. CHANGING THE SURFACE FOR ABRASION RESISTANCE

The previous discussion has been concerned with abrasive wear of machinery concerned mainly with the handling of minerals. Consequently, the major control of wear resistance has been through a change of material. However, in a broader field, particularly machine design, the use of hard, alloy materials may be costly and unjustified because of a low wear rate. However, it may be desired to reduce the wear rates even more. Although increasing the hardness of a low-cost steel through heat treatment may be the best method,changing the surface is another important control. Table 5 lists the more common techniques, their materials application and effect. However, it should be cautioned that these methods should only be used for mild wear conditions because in vigorous abrasive wear applications these coatings wear through too quickly to be effective. This point has been supported through research by Avery and Haworth.[3] They have also concluded that hardfacing deposits have to be at least 1/8 in. thick to achieve full abrasive resistance.

Of course, there are preventive measures to be taken in an attempt to exclude abrasive particles from wearing surfaces. Among these would be the use of air cleaners, oil filters, dust covers, seals, and lubricants to flush away grit, to mention a few.

6. WEAR TESTS

Wear tests are notoriously unreproducible. Repeated tests often give wear rates that differ by factors of 2 and sometimes by as much as 10.[7] This may be caused partly by undiscovered systematic differences among test conditions, but is also caused by the inherently random nature of the wear process. Several wear tests are presented here that agree roughly with the first part of this chapter.

Melting Point Theory

On the basis of sandblasting and sanding tests, H. A. Unckel of Sweden[8] proposes that abrasive wear is roughly proportional to the inverse of the absolute melting point of the material. A sampling of test results for the sanding tests are given in Table 6. This test is basically a type of grinding abrasion performed by holding a sample onto a Carborundum sanding belt for a length of time. In the discussion of this type of abrasion it was mentioned that although hardness has some effect on the wear resistance, the matrix was the major controlling factor. From the table, this can be seen to be roughly true, because materials with the same hardness give vastly different wear volumes.

Material	Melting Point $(^{\circ}F)$	Hardness BHN	Volume Loss (cu. in x 10^5) Time 5 Min.
Iron			
Annealed	2802	110	.3
Leaded Brass			
Rolled	2100	164	1.1
Annealed		76	.7
Pure Copper			
Rolled	1981	120	.7
Annealed	1981	48	.6
Aluminum			
Rolled	1220	36	1.6
Annealed		23	1.5
Aluminum Alloy (24S)			
Heat treated	935	104	1.1
Annealed		62	

Table 6

Partial Results of Sanding Tests of Materials on Carborundum Belts
Melting Point Theory - H. A. Unckel[9]

Material	Resistance Factor R
Plain carbon steel-annealed to Vickers 130-hardness diamond pyramid St. C. 16.61	1.0(Standard)
Standard Steel (above), carburized and quenched to surface hardness of R_c 68	4.0
Aluminum alloy Avional TS=54,000 psi	.5
Copper alloys	.42-.93
Cast irons	.70-.90
Cast steels	1.2-1.9 (hadfield)
Wrought steel	1.2-6.0
Alloyed cast iron Wear resistant alloy	1.3-5.5
White metals	11-18
Deep nitrided case (Vickers 950)	3.4
Hard chromium plate	11-12
Sprayed coatings, electrolytic deposits of nickel or siliconizing	1.0

Table 7

Sample Results of Stauffer Sand Erosion Tests[5]

The author disputes the concept of the Modell number presented by Oberle and discussed in another chapter.

$$\text{Modell} = \frac{\text{BHN} \cdot 10^6}{\text{Young's Modulus}}$$

This theory states (in part) that wear by sliding abrasion depends on the ratio of strength (represented by hardness) to elastic modulus. He (Oberle) attributes wear resistance to the amount of elastic deformation before plastic yield occurs. As one of his examples, Unckel mentions that the heat-treated aluminum alloys and commercial aluminum both have very nearly the same elastic modulus, very different hardness but about the same wear resistance.

Thus, although the Modell number may be applicable to many cases of abrasive wear, ~~it~~ should not be expected to be an absolute indication of the wear rate.

Sand Erosion Test

Another researcher, W. A. Stauffer[5] performed tests of sand erosion resistance. The test was to revolve five specimens (each of standard size) in a sand and water mixture and to measure the amount of volume loss after a length of time. The losses were compared with a standard specimen and a wear factor R determined. It is defined as the inverse ratio of abraded volume of reference standard to abraded volume of the test piece. The author states that deviation from the average wear volume of each material was 5%. Some values are given in Table 7.

These results agree essentially with Table 1 and the conclusion (particularly for erosive wear) that the main controlling factor is hardness. This table also indicates the value of a nitrided case, carburization, and especially a hard chromium plate. Sprayed coatings, electrolytic deposits of nickel, or siliconizing do not raise the erosion resistance of the standard appreciably above R = 1.0.

BIBLIOGRAPHY

1. Norman, T. E., "Abrasive Wear of Metals," Engineering Approach to Surface Damage, University of Michigan, Ann Arbor, 1958.
2. Avery, H. S., Surface Protection Against Wear and Corrosion, American Society for Metals, 1954.
3. Clark, D. S., and Varney, W. R., Physical Metallurgy, Van Nostrand, Princeton, New Jersey, 1952.
4. Avery, H. S., "Wear Resisatnce," Engineering Approach to Surface Damage, University of Michigan, Ann Arbor, 1958.
5. Stauffer, W. A., "Wear of Metals by Sand Erosion," Metal Progress," January 1956, p. 102-7.
6. Haworth Jr.,R. D., "The Abrasive Resistance of Metals", Trans.ASM, Vol. 41, 1949, p. 819-869.
7. Caldwell, Van, "War on Wear", Steel, Vol. 137, No.22, November 28, 1955, p. 98-101.
8. Unckel, H. A., "Wear Resistance: Is Melting Point the Key?", The Iron Age, Vol. 181, No. 12, March 20, 1958, p. 97-9.

CHAPTER 17

RESIDUAL STRESSES

1. CAUSES OF RESIDUAL STRESSES

Various presentations approach the study of this phenomenon in different ways,[1,2] some being concerned solely with mechanical causes while others have analyzed thermal influence. Perhaps an appropriate breakdown of causes should include the following:

1. Mechanical-loading is of sufficient magnitude to cause plastic deformation. Removal of load allows some elastic relaxation to take place, but due to incomplete relaxation, stresses remain.
2. Thermal-nonuniform temperature gradients cause differing degress of expansion and contraction. Although restraint must be applied, thereby introducing a type of mechanical influence, the primary cause of the final stress condition is thermal in nature.
3. Bulk Phase Transformation-particular phase transformations require a physical change in volume as one phase decomposes into its transformation products. Perhaps the most common example cited is the transformation of austenite in steels. Whether martensite or other decomposition products results, a volume increase occurs and residual stresses are created.
4. Phase Precipitation-the properties of certain metallic alloys are sometimes controlled by the process known as precipitation or age hardening. Due to a condition of supersaturation, one phase is precipitated and disperses itself throughout the grains of the matrix. Density differences between matrix and precipitate are considered to be the basic source of stress creation.
5. Combinations of the above-in most cases, the mechanism that causes residual stresses is not singular in nature. Quenching operations may induce thermal, phase transformation, and mechanical mechanisms to contribute to stress creation. Welding operations often introduce phase transformations as well as thermal gradients and plastic flow of the hotter metal.

2. MANUFACTURING OPERATIONS

The transformation of a design from a line drawing to a physical metal product requires the application of basic manufacturing processes. These include casting, welding, machining, plastic working, and heat treating. An understanding of the capabilities and comparative costs attributed to each of these processes is of primary importance to the design engineer. In addition the practicing designer should be cognizant of the potential influence that each process might exert upon the mechanical properties of the finished product. Since any commercial fabricating process will, in varying degrees, develop residual stresses in the metal being processed, the effects that these stresses might have upon factors such as fatigue life and functional specifiications are of immediate concern to those in design.

Although information both specific and general in nature is available in the literature, the means for predicting both the magnitude and sense of stresses that will result from a particular operation have not yet been developed to a useful and reliable state. From the vast amount of research devoted to this topic, however, applicable remedies have been devised and the following qualitative conclusions are formed:

1. The major causes of residual stresses are plastic deformation and non-uniform thermal effects.
2. Tensile residual stresses are generally detrimental, especially when they occur at or near the surface.
3. Compressive residual stresses usually provide beneficial tendencies.
4. Reduced fatigue life, warpage of the workpiece, immediate failure under service loads, and accelerated stress corrosion and pitting indicate some of the detrimental effects of residual stresses.
5. Core as well as case properties may be altered by the presence of these stresses. Consequently, consideration must be given to subsurface as well as surface stresses.

6. Most commercial methods of stress relief are either thermal or mechanical in nature. Thus, remedial rather than preventative measures are indicative of the present state of control.
7. Manufacturing operations cause great variations in both magnitude and sense of residual stresses. Typical examples are tabulated below and present the general tendencies to be expected in regard to surface residual stresses only.

Tensile	Compressive	Either
Welding	Nitriding	Carburizing
	Shot Peening	Rolling
	Flame and Induction Hardening	Casting
	Heating and Quenching Single Phase Materials	* Abrasive Metal Cutting
		Non-Abrasive Metal Cutting
		* Heat and Quenching Material that undergo phase transformation

* Majority of commerical situations would probably produce dominant tensile stresses

Charts, graphs, or other data included pertain to the specific examples presented and should not be interpreted otherwise.

Spot Welding

One of the characteristics of this operation is the presence of a heat source for extremely short periods of time. This creates a large temperature gradient which in turn leads to fast cooling rates as the heat source is removed. Any tendency of the hotter metal to expand is met by restraint from the colder surrounding metal. In addition, phase transformations occur where such a possibility exists. The predominating stresses are usually tensile and a typical biaxial stress distribution is shown in Figure 1.[2] This particular distribution involves a material that had not

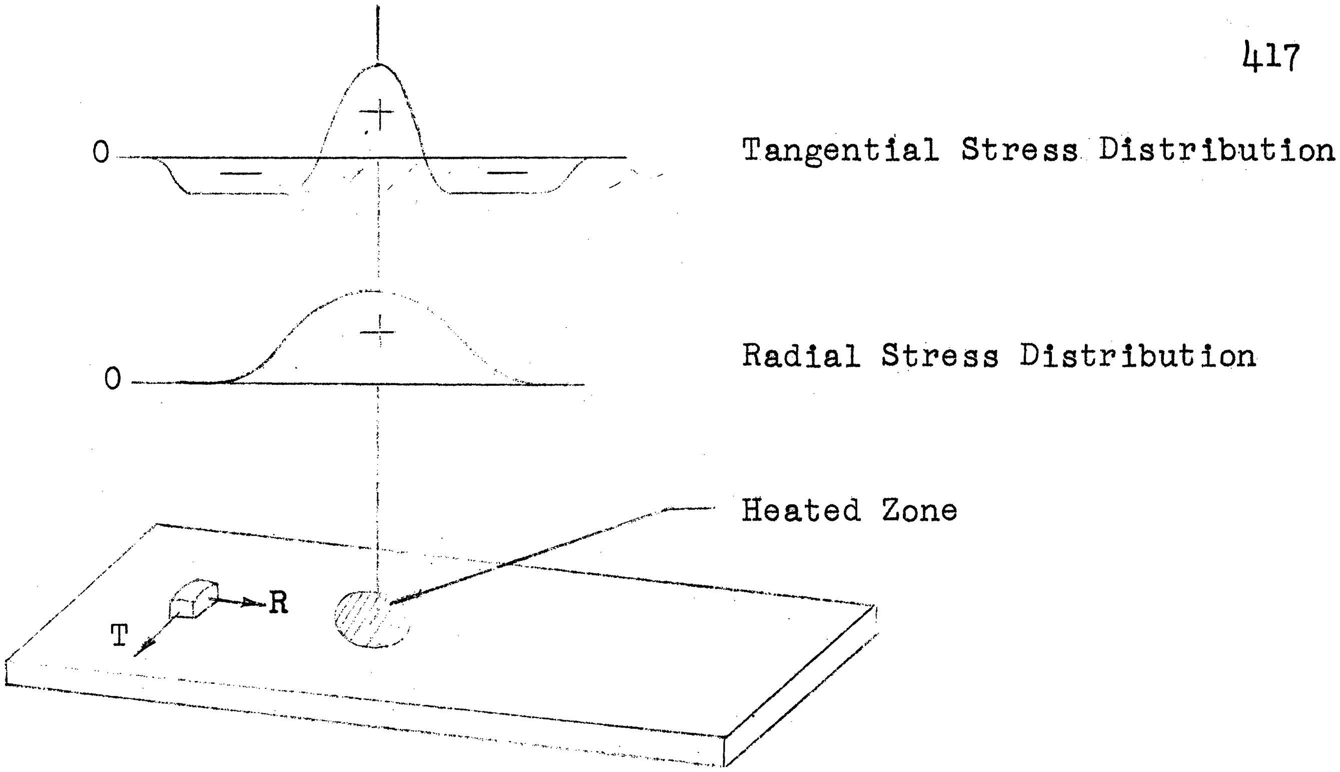


Figure 1

Residual Stress Distribution in a Thin Strip Subjected to a Spot Weld. (No Phase Transformation). Reference No. 2.

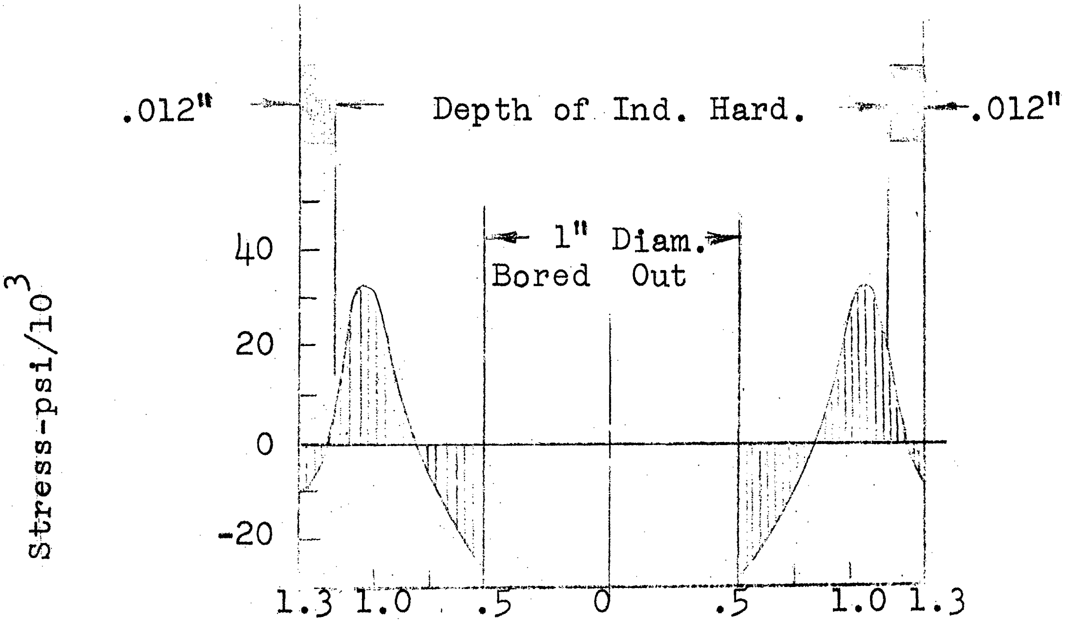


Figure 2

Residual Stress Distribution in an Induction Hardened Crankpin.[2]

undergone phase transformation. If this mechanism were also included, the peak tensile stresses would probably be increased. Most other welding operations tend to create tensile residual stresses at the surface.

Induction Hardening

This method of surface hardening consists of heating a relatively shallow layer of a piece of steel to the austenite region then quenching to produce martensite. Phase transformation, thermal effects,and mechanical restraint all influence the final state of stress which is compressive at the surface. Figure 2[2] shows a typical distribution that results from this operation. Nitriding, shot peening, and flame hardening are other processes in this category which tend to create surface compressive stresses. Carburizing, however, may lead to either type of stress at the surface. The major difference, from a stress viewpoint, is due to the process itself. Since the entire piece, rather than the surface alone, is heated to high temperatures, the transformation characteristics and cooling rates in the core section can markedly alter the over all stress pattern.

Heat Treating in General

The ability of many metals to be hardened by various heat treating operations is one of their most desirable properties. Since heating, then quenching in a cold medium are inherent in this process, it is not surprising to find a stressed end product. In many situations, excessive stresses occur and cause the phenomenon commonly called "quench cracking". Factors such as material composition, size and shape of part, and quenching media introduce so many possible complexities that wide variations in final stress patterns can result. Figure 3[2] and 4[3] indicate different patterns that have been found.

Casting

Fundamentally, the cooling of a casting as it solidifies is similar to quenching a piece of metal, but particular characteristics of the casting process cause added complexities. Mold restraint can have a decided influence upon stress patterns and may cause serious consequences, such as hot tears. Figure 5 illustrates how such results can occur. In addition, nonhomogeneity of the structure and uneven cooling rates may lead to centerline shrinkage, interdendritic shrinkage, and blowholes. The resultant stress distribution can be influeced by so many variables that no single comment will suffice.

Rolling

The mechanical working processes are characterized by plastic deformation of the workpiece and, for the purposes of this chapter, this single mechanism will be considered as the sole source of stress creation.

Apparently, the resultant stress distribution caused by rolling operations is influenced greatly by the depth of pene- or plastic deformation that is produced. Slight reductions in bar thickness are considered as "surface" or "nonpenetrating"

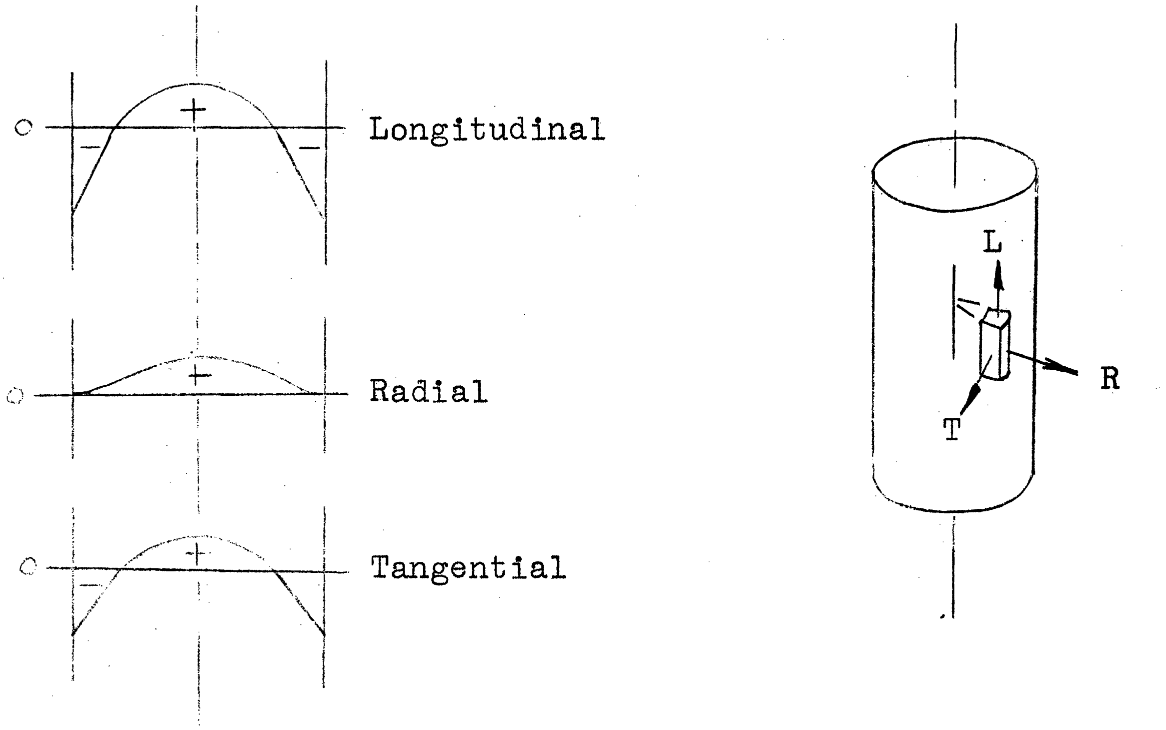


Figure 3

Residual Stress Patterns in a Quenched Cyclinderical Bar (No Phase Transformation). Reference No. 2

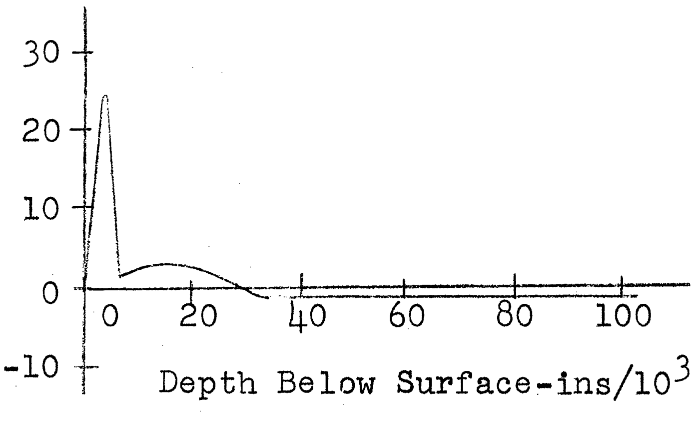


Figure 4

Residual Stress Distribution in a Falt Bar of AISI 52100 Stee after Heat Treating to Obtain a Hardness of R_c59. Reference No.3

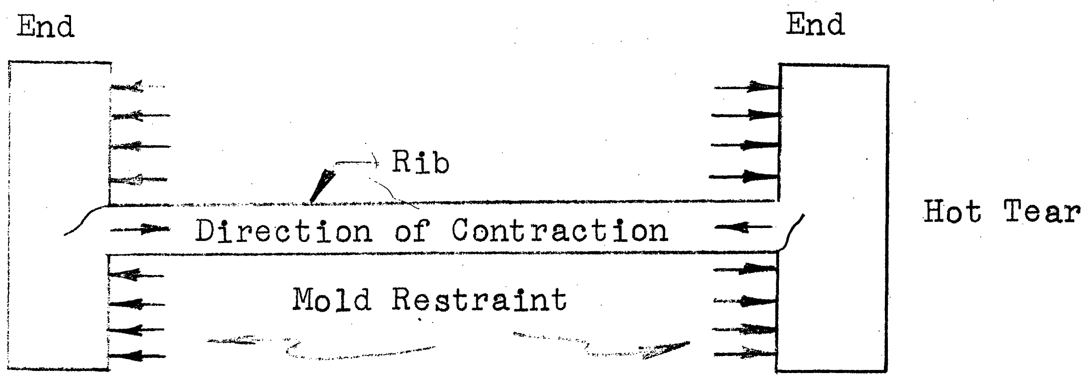


Figure 5

Development of hot tears in a casting as it solidifies from a molten state. The thin rib becomes fully solid before the ends do and as the rib cools and contracts, restrainst is afforded by the mold. If the induced stresses exceed the strength of the hotter ends, tears will result as shown.

rolling while large reductions are classed as "heavy" or "penetrating" rolling. Figure 6a[1] presents a schematic version of these differences as they affect stress patterns in the finished workpiece. A possible explanation of these differences is that surface rolling is similar to shot peening in that the tendency of surface expansion is held back by the interior thereby causing surface compressive stresses. During heavy rolling, plastic deformation results throughout the section and, under load, the full workpiece is subjected to longitudianl tensile stress which is maximum at the surface. This condition prevails because the outer regions of the section thickness experience a greater degree of plastic flow than do the interior regions. Upon passage through the rolls, the external load is released and elastic relaxation follows until an equilibrium of residual stresses is reached. Figure 6b is a schematic presentation of a simplified version showing the shift in stress pattern from a load to an unloaded condition.

Metal Cutting Operations

Non-abrasive

These operations employ a cutting edge of controlled geometry, and although temperature influences are possible, the consensus of opinion seems to be that mechanical affects are the predominant cause of residual stresses. Henriksen[4] conducted planing tests on a range of plan carbon steels and from his findings, concluded that the maximum surface stresses were predominantly tensile. Colwell[5,6] investigated the effect of milling and shaping operations

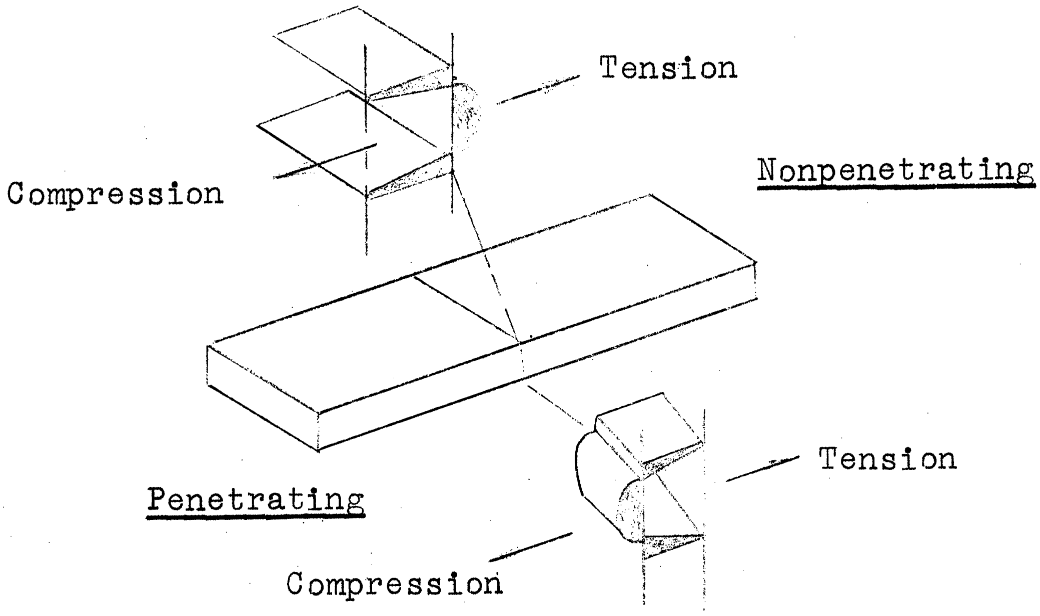


Figure 6a

Residual stress distributions caused by penetrating and non-penetrating rolling of flat strip stock. Refernce No. 1

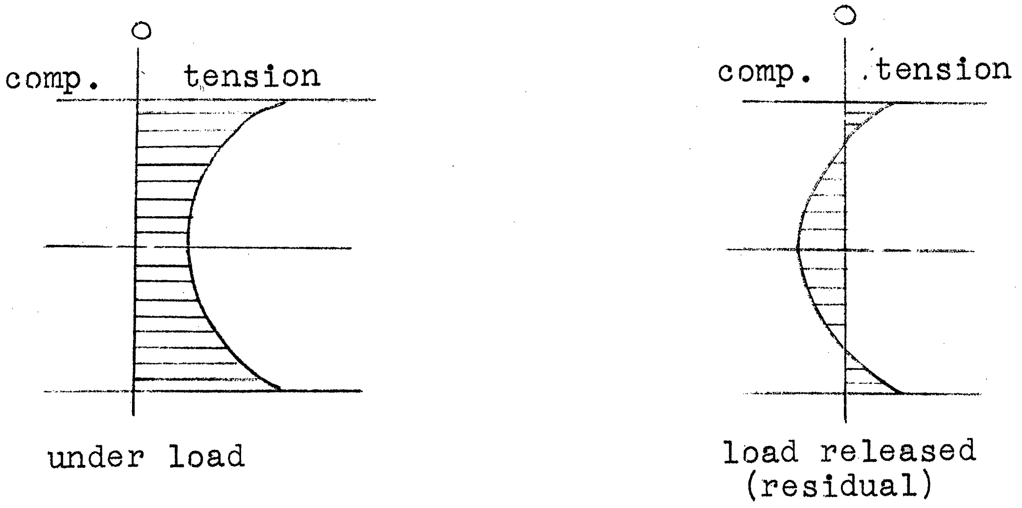


Figure 6b

Stress distributions caused by penetrating rolling. Residual is at the right.

and found the residual surface stresses to be compressive in nature. From this and other work he concluded that either type of stress can result. Apparently, the geometry and degree of sharpness of the cutting tool, as well as lubrication, size of cut, and overall rigidity of the machining setup all exert some influece upon the type of stress that results. Regardless of the end condition, the depth of stress penetration is relatively shallow with critical stress levels seldom exceeding 0.010 inches.

Abrasive

In recent years the major emphasis of residual stress studies in the area of metal cutting has been devoted to abrasive operations. A logical answer to the apparent overemphasis on this area is that a great many of todays machined products require the degree of precision and surface finish that can be attained only through abrasive techniques.

These operations involve the use of hard abrasive particles that serve as cutting edges having uncontrolled geometry. In addition, the workpiece is subjected to more severe thermal influences as compared to non-abrasive types of operations. Practically all of the investigations reported pertain to restricted test conditions and extrapolation of such specific findings might prove erroneous. Several general conclusions seem to prevail, however, and these are:

1. Stress penetration as the surface is quite shallow, and critical stresses seldom exceed several thousandths of an inch.
2. Predominance of the thermal influence usually produces tensile stresses. This would be characteristic of commercial conditions that are considered to be "heavy" cuts.

3. Compressive surface stresses can result when cutting conditions are "light" by commercial standards.
4. Peak stresses are usually greater than those that result from non-abrasive operations.

Tarasov, Hyler, and Letner[7] reported that stresses resulting from the grinding of a ball bearing grade of steel reduced the fatigue limit as much as 13% but could increase the limit by 38% depending upon the grinding conditions. Apparently, other investigators feel the detrimental aspects of these findings are conservative and may prove misleading for the general grinding situation. Letner[3,8], Colwell, Sinnott, and Tobin[9], Halverstadt[10], and Clorite and Reed[11] reported findings for different work materials and the results substantiate the general conclusions mentioned previously. Variations of measured stress distributions are pronounced and, as yet, unexplained. It is of interest to note that tensile residual stresses in excess of 200,000 psi and compressive stresses greater than 100,000 psi have been reported.

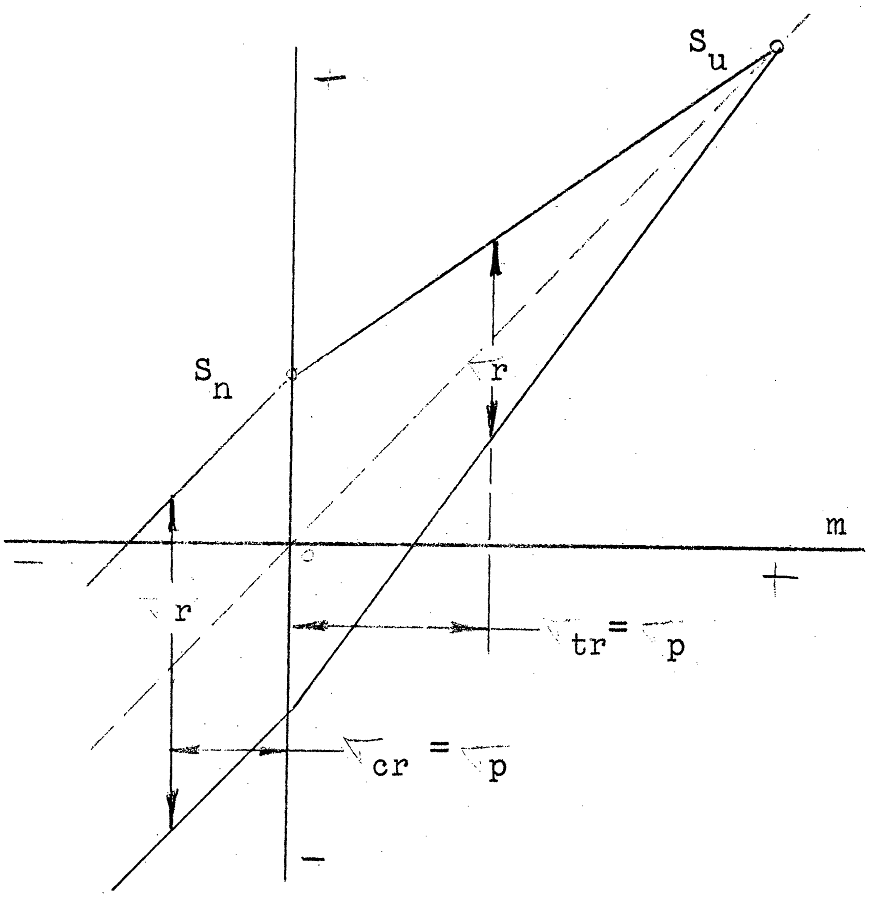
3. THE INFLUENCE OF RESIDUAL STRESSES ON THE PHYSICAL PROPERTIES OF METALS

Effect on Fatigue Strength

When the presence of residual stresses is accompanied by a change in fatigue resistance, the change is often attributed to their influence. However, the level of influence of residual stresses on the fatigue resistance of machine elements depends on many factors connected with the configuration of the part, the properties of its metal, the nature of the residual stress distribution, the magnitude of the residual stresses in the danger zone, and the nature of the external stresses acting during operation.[20]

Although considerable data has been complied dealing with the influence of residual stresses on the fatigue of fabricated parts, the exact relationship is still not well understood. There are those who oppose drawing a simple cause-and-effect relationship from this data on the grounds that the methods used to introduce residual stresses also introduce other effects that may be the immediate cause for increasing fatigue life. For example, while introducing compressive residual stresses on the surface of a metal, shot peening also cold works the surface, thereby raising the tensile strength and related fatigue strength. Evidence, however, favors a direct relationship and it is generally agreed that the fatigue life is affected beneficially or detrimentally by residual stress depending on whether they are compressive or tensile and whether they occur at or below the surface. Surface compressive stresses are commonly associated with increased life, whereas surface tensile stresses are associated with decreased life.

A basic understanding of the manner in which residual stresses affect the fatigue strength as discussed above can be gained by treating the surface residual stress as a preload stress (constant or varying as when fading occurs) and superimposing it on the alternating operating stress by use of a modified Goodman Diagram. This has been done in Figure 7 for an alternating compressive and tensile external load.



S_n = Fatigue strength under completely reversed stress

S_u = Ultimate strength

σ_{cr} = Compressive residual stress

σ_{ta} = Tensile residual stress

σ_p = Preload stress

σ_r = Allowable external operating stress

Figure 7

Goodman Diagram showing the Idealized Effect of Residual Stresses on Fatigue Strength

From this diagram, it is evident that in the presence of compressive residual stresses, higher externally applied stresses are tolerable than when surface tensile residual stresses are present. This principle also applies to subsurface zones which possess residual stresses.[21]

Although this principle of adding the residual stress to the externally applied stress is simple, in actual design practice it is seldom used. This is because the stresses should be combined not only at the surface of the part, but also throughout the crossection which implies that the magnitude and distribution of the stresses is known. In the case of externally applied stresses in an unnotched part, the distribution is uniform for axial loading or varies lineraly for bending and torsion. However, when notches are present, the stress distribution becomes more complex as in the case of residual stresses. In these cases the exact distributions are seldom known, rendering the addition of the stresses impossible. Also, the effect of residual stresses on the strength of a part is generally much smaller than the effect of externally applied stresses. For steel members, residual stresses of less than 25,000 psi can probably be neglected[22]. For these reasons, the absolute amount of surface residual stresses should not be taken, as it often is, as a measure of the increase in the fatigue limit of a constructional part.

Parts Lacking Stress Concentrations

For smooth parts (or for parts with no considerable concentrations of stress) the residual stresses do not give rise to very great changes in the resistance to fatigue. In this connection the greater the softness and ductility of the material, the less this influence will be.[23,24] On smooth samoles of structural steel, the endurance limit under flexure or axial stressing has been successfully raised by inducing compressive residual stresses, through surface cold-work and quenching, to a comparatively moderate magnitude (10-22%).[24,25] However, a considerable part of this effect must be attributed to strengthening of the surface layers (increasing hardness).

The maximum lowering of the endurance limit under flexure for smooth samples of structural steel owing to tensile residual stresses will generally not exceed 16%, regardless of the magnitude of the stresses.[24,25]

The apparent reason for the comparatively small influence of residual stresses on the fatigue endurance of smooth members is the removal (fading) of a considerable proportion of these stresses by cycling stresses close in magnitude to the endurance limit. This phenomenon will be discussed later.

Parts with Sharp Stress Concentrations

The part played by residual stresses on the fatigue endurance of metal parts changes abruptly if these parts possess stress concentrations (notches).[21,23, 24] These concentrations of stress can be brought about both by structural and by the operating or technological factors of grooves, drill holes, apertures, fixed mountings, breaks, corrosion damage, casting or welding effects, etc. In this case, the residual stresses can alter the magnitude of the endurance limit by as much as 50%,[24] both increasing it or reducing it depending on the type of stress.

This large effect of residual stresses on the fatigue resistance of parts with large concentrations of stress is explained in part by the following facts:

1. Residual stresses can, like stresses from external influences,be concentrated around various grooves, slots, etc.
2. The concentrations of compressive residual stresses at apertures and grooves offset, in part, the inherent weakness of surfaces that are subjected to repeated loads by creating local cold-working of the metal (in the danger areas of notches.) The softer and more plastic the non-cold-worked metal, the greater the capacity of the cold-worked metal for reacting to the residual stresses.
3. Removal of residual stresses when parts with notches are stressed is considerably more difficult than for notch free parts.

Fact three is particularly significant. In many cases, the high residual stresses in the area of a notch remain unchanged even after the part has been loaded to limits close to or exceeding the endurance limit.[24]

It should be noted that it is of particular pratical value that compressive residual stresses are more beneficial in the case of notched specimens than for smooth specimens since machine parts and structures usually have surface irregularities which act like notches and often are the points of initiation of fatigue cracks which ultimately lead to service failure.

Some Considerations in the Control of Residual Stress to Raise Fatigue Strength

In selecting methods to control residual stresses to raise the fatigue strength of a part the following points should be considered:

1. Residual stresses are of only slight significance in regard to the fatigue strength and endurance limit at low strength of a particular group of alloys.[19]
2. Residual stresses are also of little significance, in these respects, for polished, smooth (ie., cylindrical and prismatic) sections[24]
3. The effect of residual stresses, if they exist, become more pronounced as the number of cycles to failure increases[16]
4. The important effects of residual stresses on structures which are subjected to repeated loads relate to conditions where the fatigue strength and endurance limit are generally reduced in comparison to those of polished, smooth sections. This applies (a) to many high-strength materials and particulary to steels, heat-treated to high-hardness, (b) to the fatigue properties in the presence of high stress concentrations, simultaneous chemical attack or corrosion fatigue and of a simultaneous rubbing, fretting, or chafing action.[19]

Effect on Ultimate Strength and Ductility

Such properties as the ultimate strength and ductility of a metal are not altered significantly by residual stresses provided that a material of pronounced notch sensitivity such as hardened steel or cast iron is not used. These facts can be explained by considering a bar containing surface compressive residual stresses and tensile residual stresses in the core as in Figure 8. If an external tensile load is applied to the bar, a uniform tensile stress is introduced (b). In this case, the residual and externally applied stresses are directly additive to form a net stress as in (c). If, as the external load is applied, the net stress exceeds the yield strength local flow takes place until the stress peak subsides into the elastic range due to redistribution of the net stress as in (d). Therefore, since

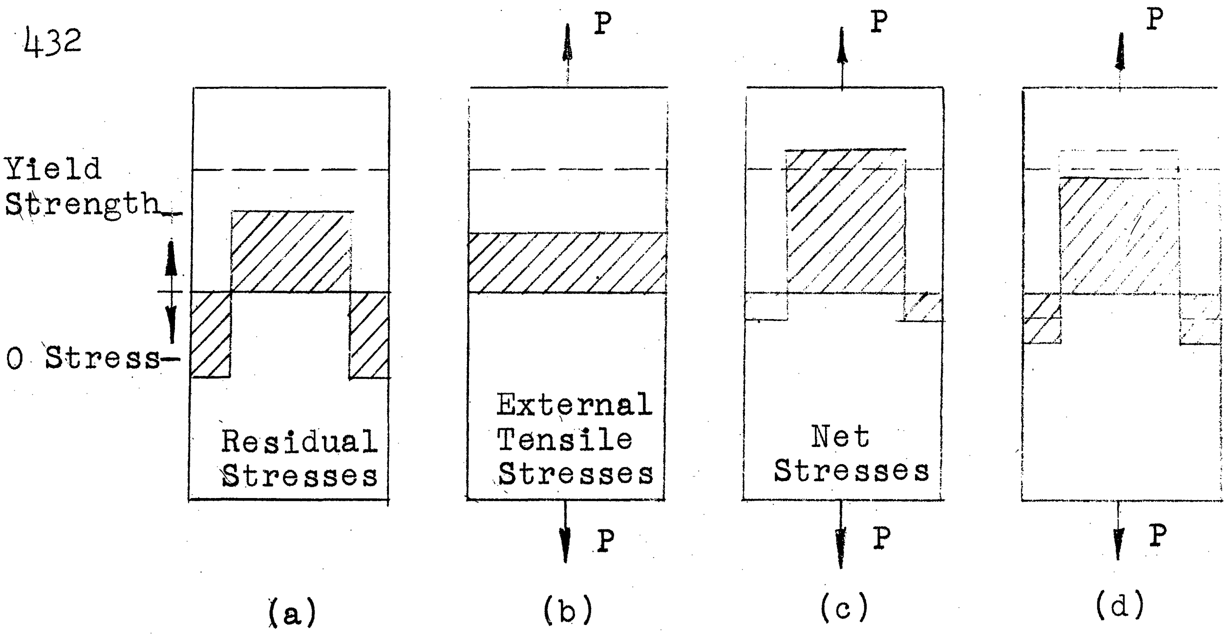


Figure 8

Effect of Residual Stresses on Ultimate Strength

the ultimate strength and ductility are properties of a metal above its yield point, they are for the most part not altered by residual stresses.

However, in the case of some parts with large crossections made of ductile materials, if the core is under a high all sided tensile stress, it is possible that the flow phenomonon will be hindered such that the sum of the applied and residual stresses will reach the tensile strength and a crack in the core will lead prematurely to total fracture. Such failures will occur especially if there is impact loading. Apart from these exceptional cases, the load carrying capacity of tough materials under stationary loading is not much affected by the presence of internal stresses regardless of the nature of loading (tension, bending or torsional).[25]

Effect on the Static and Impact Strength of Brittle Materials*

Materials of low ductility fail at low stress by brittle fracture under static and impact loads because they are not capable of yielding to relieve local high stresses. Brittle materials will, therefore, suffer from surface imperfections whether the applied load is static, impact, or fatigue.

Since completely brittle materials fail only by tensile stresses, their strength, under various kinds of loading, is increased by introducing residual compressive stresses in the fracture sensitive layers. Steel having the hardness of the case of carburized and nitrided specimens would be notch sensitive and therefore brittle except, as shown in Figure 9, that the hard layers are residually stressed in compression.

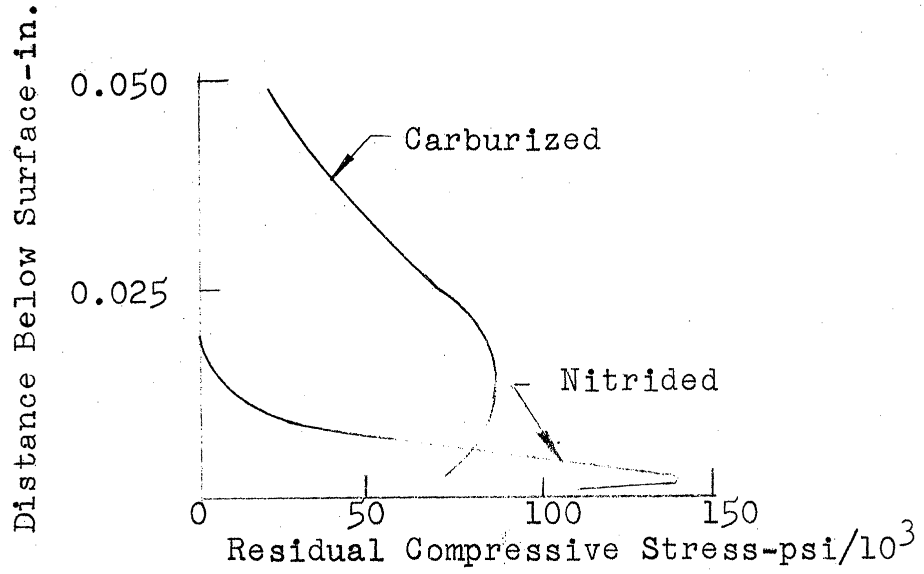


Figure 9

Residual Stresses developed in Nitrided and Carburized Specimens by the Expansion of the Hard Case during the Hardening Transformations

---

* This discussion is abstracted primarily from reference 27.

Prestressing by mechanical means can only be accomplished in materials have some ductility. However, Figure 10 shows that steels of hardness as high as Rockwell C64 (BHN 700) is is sufficiently ductile to respond to shot peening. The peening which was not of high intensity is seen to have developed a residual compressive stress that extended to a depth of 0.009 inches and a magnitude of 150,000 psi. Mechanical prestressing can, therefore, be expected to increase the static and impact strength as well as the fatigue strength of many hard metals.

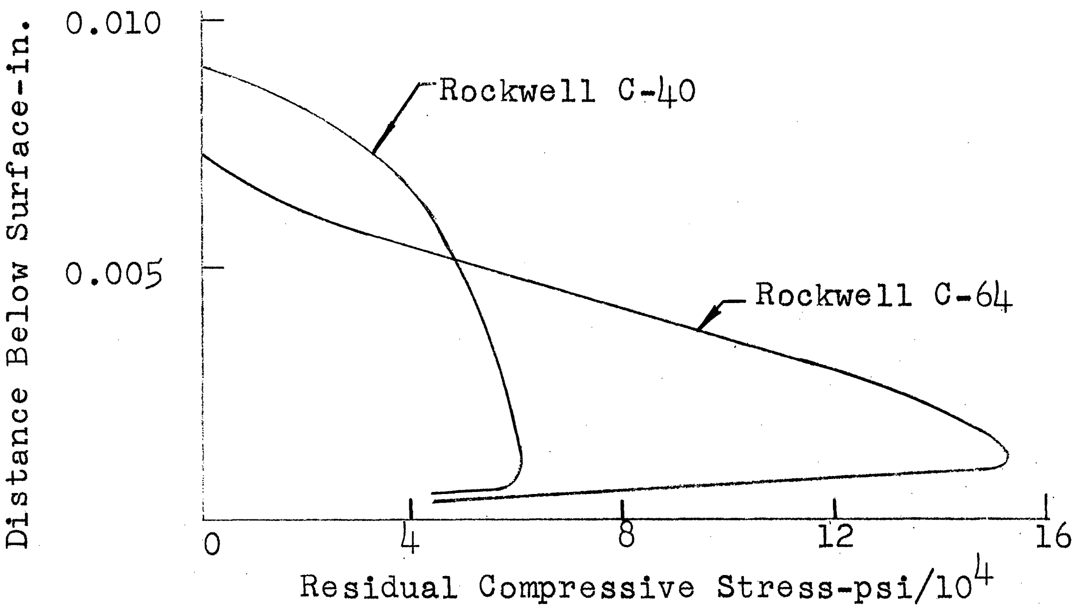


Figure 10

Magnitude and Depth of the Residual Stresses Induced by Shot Peening Varies with the Yield Stress of the Peened Material

Mechanical prestressing is also effective in reducing the "brittleness" acquired by hardened steel that is uniformly hard throughout, (tool steels) as well as reduce the hazard of fracture that often occurs because of reduced ductility at low temperatures. In general, it can be stated that the effectiveness of residual stresses increases as "brittleness" increases.

Effect on the Elastic Limit and Yield Strength

Overstraining a metal in tension or compression raises the elastic limit for stresses in the same direction and lowers it for stresses in the opposite direction. This is known as the Bauschinger effect[20] and it is directly attributed to the effect of residual stresses since it can be nullified by stress relieving.[26] The proportional limit is similarly affected. These effects are shown in Figure 11.

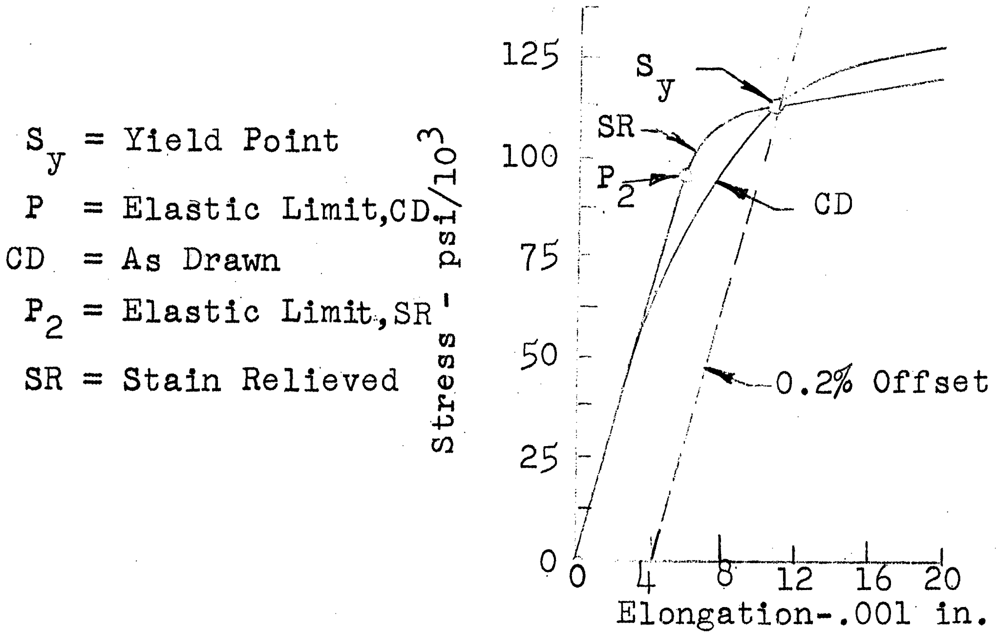


Figure 11

Effect of Stress Relief at 900°F on Yield Strength of a Cold-Drawn 0.45 per cent Carbon Steel

Effect on Resistance to Penetration

Tensile residual stresses decrease the resistance to penetration whereas compressive residual stresses increase the resistance to penetration, but to a lesser degree than the decrease due to tensile stresses.[27,28] The softer the metal, the greater this effect becomes, and it is most pronounced when the stress is at or near the yield point of the material. These relations are shown in Figure 12.

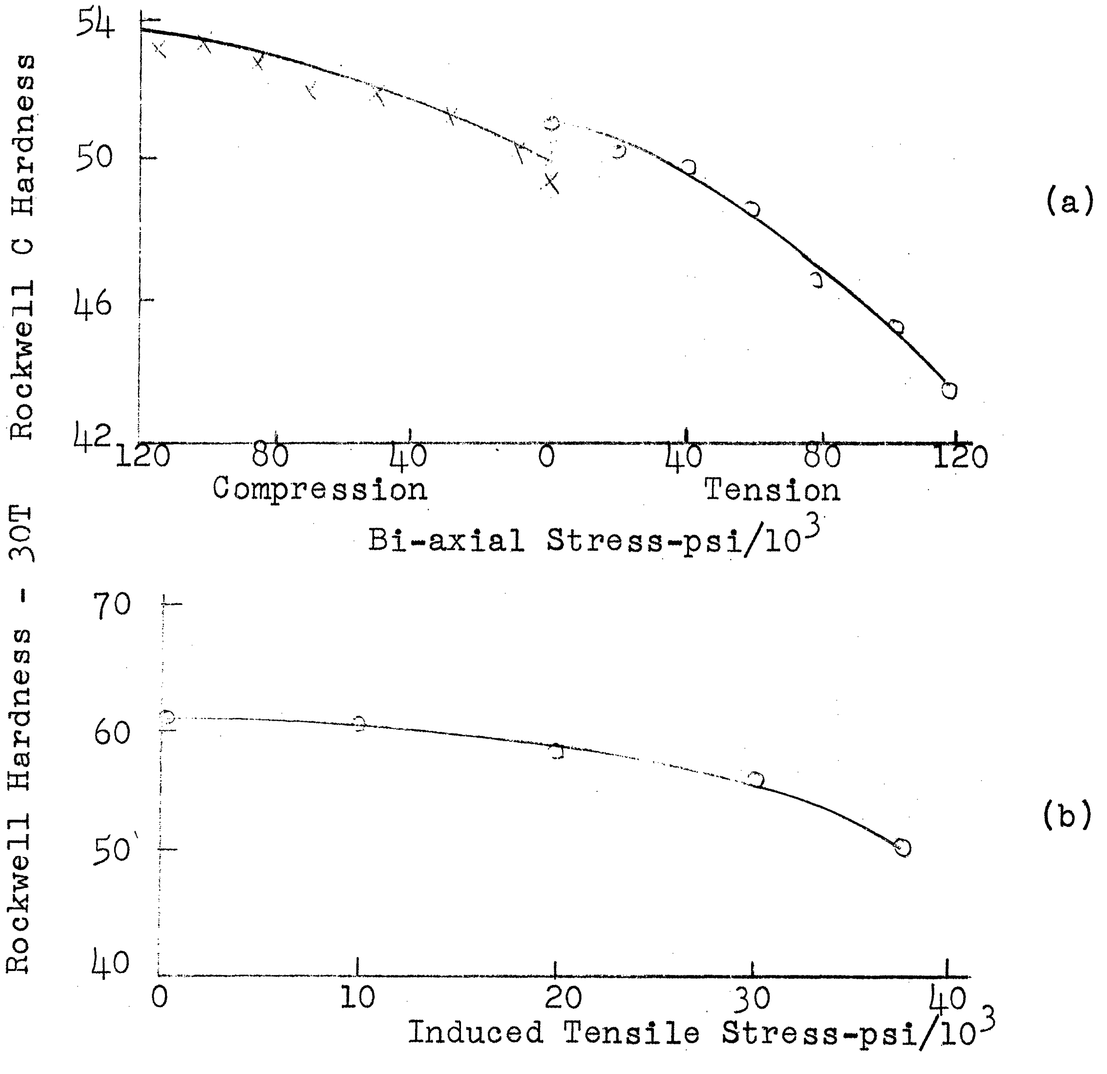


Figure 12

(a) Indentor hardness measurements and surface stress in steel disks that were deformed from a plane surface toward vertical shapes. (b) Hardness curves for annealed mild steel under uniform tensile stress. As the tensile stress increases, the hardness decreases approximately linearly in the elastic region. It departs from linearity as the yield strength of the material is approached.

This phenomenon has no apparent practical significance except for its detrimental affect on penetration hardness readings are a relative measure of resistance to penetration. Such hardness readings are increased in the presence of compressive stresses and decreased in the presence of tensile stresses.[28]

Although it has not been investigated, this phenomenon could affect the abrasive wear qualities of machine parts residually stressed on the surface since this type of wear depends directly on resistance to penetration. That is, abrasive or cutting type wear takes place whenever hard foreign particles are present between rubbing surfaces. These particles first penetrate the metal and then tear off the metallic particles.[28] Thus, the greater the resistance to penetration, the greater the wear resistance of the part. This effect is highly speculative and requires investigation to prove its validity.

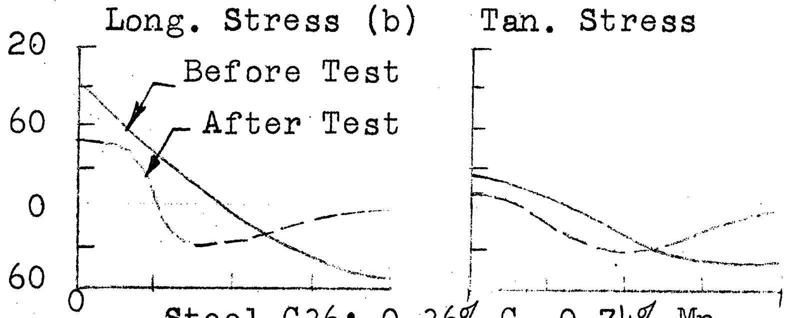
Effects on Other Properties

Residual stresses effect properties other than those discussed which may have an effect on wear resistance; for example, creep, electrode potential, acid solubility and corrosion resistance, the propagation of mechanical vibrations, the rate of transformation of phases, and the rate of aging.[30] These factors could affect frettage, cavitation erosion, galvanic corrosion or stress-corrosion cracking. However, very little is known about the exact relation of internal stresses to these properties; chiefly because the process of setting up stresses also causes a fragmentation or reorientation of grains, both being factors that have a powerful effect on these properties.[20]

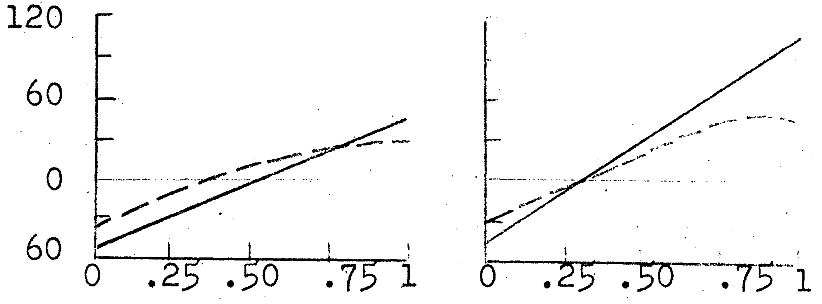
4. FADING OF RESIDUAL STRESSES

Residual stresses sometimes decrease in magnitude under cyclic loading. Earlier in this chapter this phenomenon was credited with the negligible effect of residual stresses on the fatigue strength of ductile, notch free machine parts. This decrease in residual stresses under cyclic loading is termed fading. Effects of this phenomenon are shown in Figure 13. Diagram (b) is of particular significance in that here the initial surface tensile stresses are seen to actually change into compression after various degrees of bending stress and a number of stress reversals were applied. Fatigue tests on the specimens from which these curves were obtained disclosed that the endurance limit was not decreased by the initial value of residual stresses. This fact is explained by the fact that the tensile stresses changed to compressive stresses.[31]

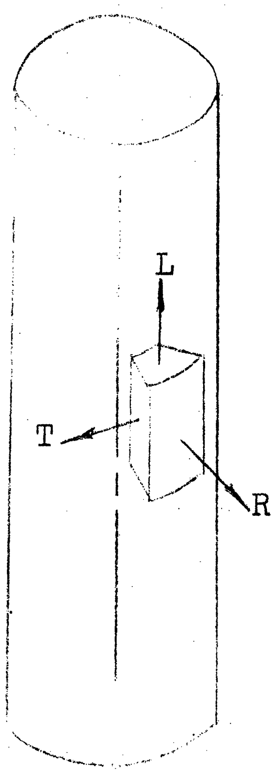
Residual stresses induced by various means fade at different rates and to different degrees depending on the magnitude of initial stress and the damping capacity or yield characteristics of the metal involved. A single cycle of applied stress above the yield point of shot peened metals can remove the greater part of the longitudinal residual compressive stress. Repeated cycles of stress below the yield strength act much more slowly. Serious reduction of residual compressive stresses usually does not occur for stresses lower than one-third of the yield strength for as many as one million stress reversals. In the case of specimens of steel (carburized,



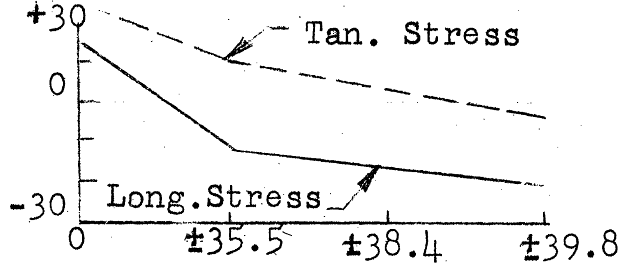
Steel C36: 0.36% C, 0.74% Mn. Cross section
Ice Water Quench From 1110°F.
0.97 10 Cycles at ± 48,000 psi
Endurance Limit 50,000 psi



Steel Ni 12; 0.04% C, 0.52% Mn.,
12.17% Ni. Annealed at 1650 F-
Furnace Cool to 680 F - Water
Quench from 680 F
236 10 Cycles at ± 51,200 psi
Endurance Limit 51,200 psi



(a)



First Load-1000 psi

0 5.04 8.31 7.63

Cycles at Load 10

0 0 0 4.41

Second Load-1000 psi

0 0 0 4.52

Cycles at Load 10

Specimen: O.D. - 1.075 in., Bore 0.315 in
Bore water quenched from 1110°F.
Steel C34; 0.34%C, 0.69% Mn.

(c)

Figure 13

Fading of residual stresses due to repeated stressing. (a) the direction in which the longitudinal stress L, the tangential stress T, and the radial stress R, act in a bar. (b) Steel bars as defined. (c) residual strssses on a hollow steel bar quenched from bore.

heat treated and then shot peened), the application of one million cycles of reversed bending with an applied stress of 40,000 psi did not appreciably reduce the surface residual compressive stress. Application of cycles of 50,000 psi stress (reversed) caused fatigue fracture, but a single cycle of applied reversed stress of 125,000 psi did not appreciably reduce the residual stress.*

Although very little is known concerning fading in cast iron, it has been observed that cast iron tractor wheels exhibited no appreciable fading of residual stresses in service.[33] The theoretical effect of residual stresses on the S-N curve

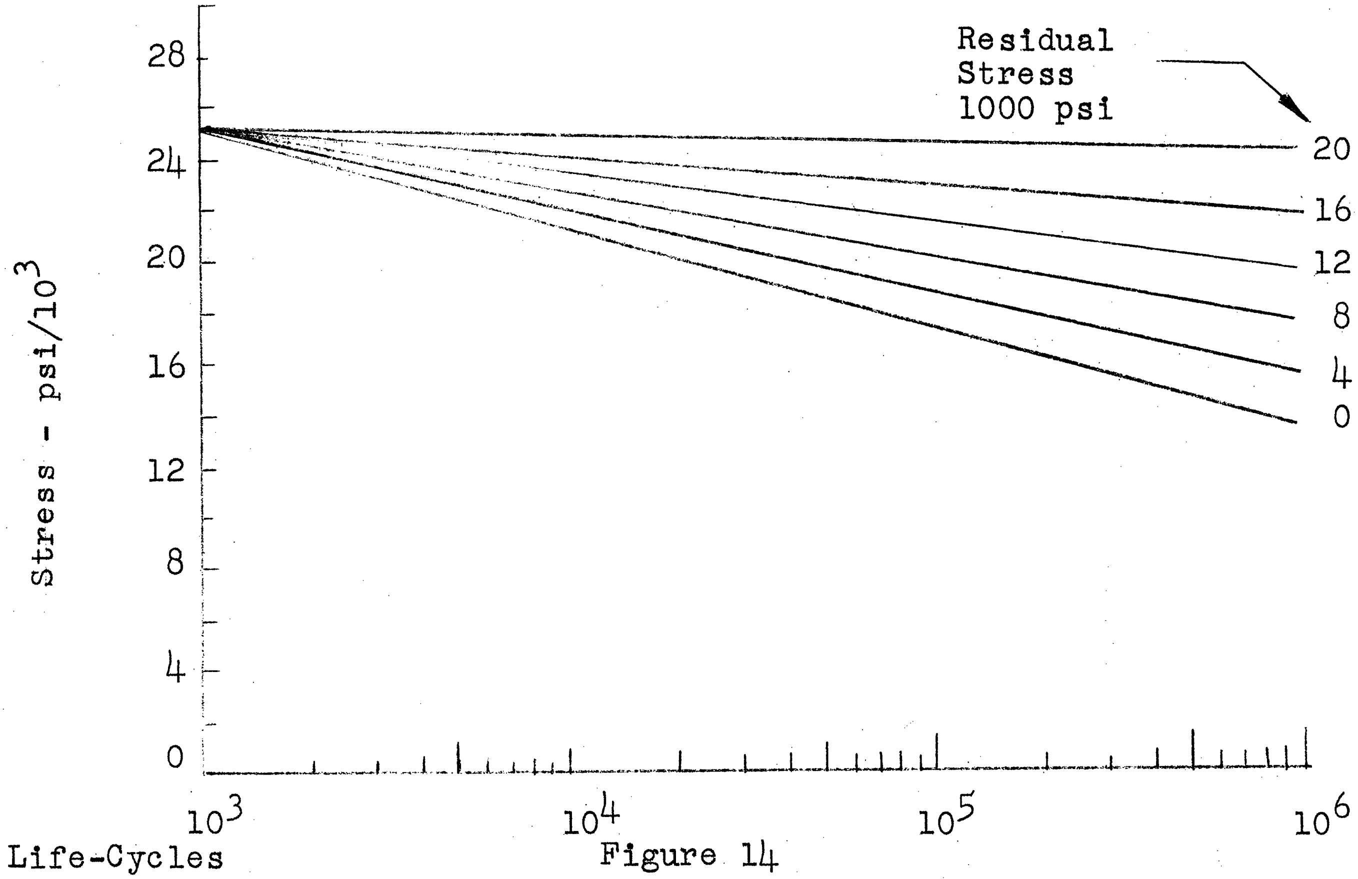


Figure 14

The Effect of Compressive Residual Stresses on the S-N Curve of ASTM No. 28 Cast Iron. Contructed by the use of a Goodman Diagram.[33]

---

* This paragraph was primarily abstracted from Reference 32.

(constructed from a Goodman Diagram) of ASTM No. 28 cast iron is shown in Figure 14.

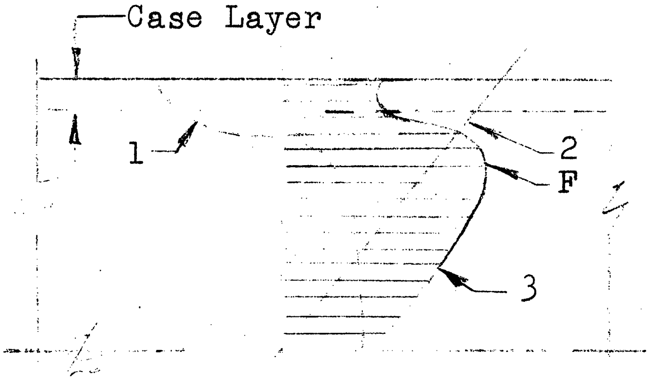
In general, it can be stated that fading of residual stresses under vibratory loads, in parts without large stress concentrations, occurs noticeably only when these loads are of considerable magnitude (close to or at the endurance limit). Where stress concentrations exist, even considerable overloads do not generally lead to removal of residual stress in danger zones.[16,24].

5. THE EFFECT OF RESIDUAL STRESSES ON THE WEAR AND SERVICE LIFE OF MACHINE ELEMENTS

Internal Notch Effect[16]

Parts which are surface-hardened by such means as case hardening, nitriding, and induction or flame hardening exhibit reduced fatigue resistance in the heat affected or transition zone between the hardened and unhardened regions. This weakness may occur in a 1. subsurface area or 2. surface zone between the hardened and unhardened area. Such failures are sometimes attributed to the fact that the physical and structural properties of the steel in the transition region are inferior to those in the hardened layer. However, the most popular explanation for this weakness is that the favorable residual compressive stresses in the hardened layer are balanced by opposing unfavorable tensile stresses in the transition zone. It is the residual tensile stresses which lead to an internal-stress peak; premature fatigue failure initiates through this region of high tensile stress.

The nucleus of fracture located below the surface may not necessarily occur at the depth of the hardened layers. Figure 15 shows that the initiation point of fracture F is located in the region of high tensile stress causing an internal notch. Maximum tensile stress occurs below the surface at F becuase of the combined effect of the residual stress from surface hardening superimposed on the stress due to external loading. Typical stress distribution due to various forms of surface hardening are shown in Figure 16. The steep stress gradient at the case core intersection in induction hardened parts makes them particularly susceptible to subsurface failure due to the internal notch effect.



F = Region of failure
1 = Residual stress due to case carburizing.
2 = Bending stress due to external loading
3 = Net stress from (1) and (2)

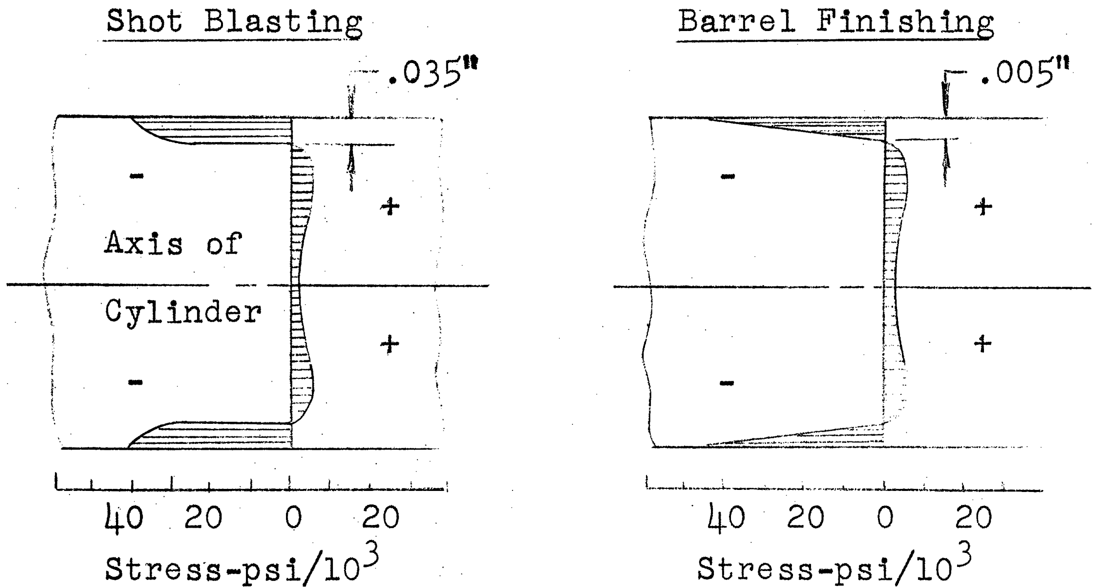
Figure 15

Stress Distribution Through Tension Side of Bending Specimen Case Carburized

The stress peak which arises in case hardened parts due to residual stresses can also cause premature failure by acting on subsurface inclusions, seams, cracks and other discontinuities. In parts under rolling loads and other localized contact pressures, a subsurface region of maximum shear combined with the residual stresses would be especially sensitive to these internal discontinuities. However, there have been virtually no studies conducted on the effect of residual stresses on rolling fatigue.[34] Under repetitive loading, subsurface cracks may propagate to the surface so that various forms of pitting eventually result.

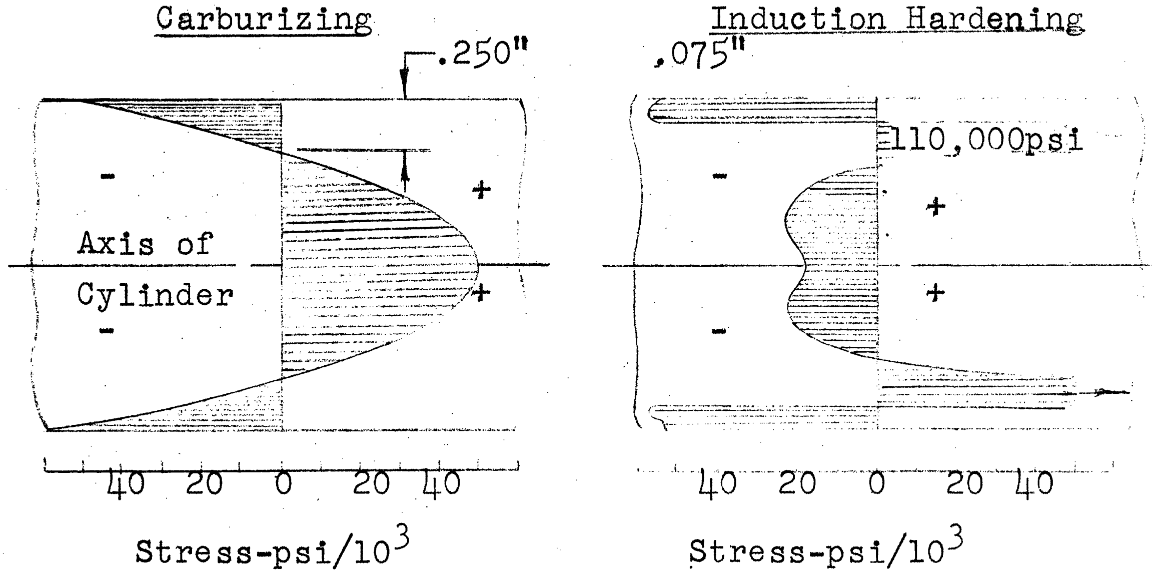
Spalling[35]

Spalling is considered a surface fatigue phenomenon characterized by the flaking of a surface layer as is sometimes encountered in gear contact surfaces. It is generally considered a special form of pitting. If spalling-type failure occurs, it is usually associated with a particular kind of residual stress distribution. The main characteristic which distinguishes spalling from pitting is that the fractured surface runs substantially parallel to the original surface. This is shown schematically in Figure 17 which shows a cross section of a ball bearing which failed due to spalling alongside the tangential residual stress distribution existing in the ball. The failure apparently started in the region stressed in tension below the surface and propagated radially until it encountered a zone which was residually stressed in compression. Then the fracture veered off, running parallel to the surface, and sub-



SAE - 1045
Normalized Bar Stock
Grit Blasted
9 minutes

SAE - 1045
Water Quenched
Drawn at 1000 F
Tumbled 10 hours



.30 C Amola
Normalized
Carburized .070"
Oil Quench - R_c 58

SAE - 1045
Drawn at 1125 F
.060" Case - Water Quench
Drawn at 400 F

+ Tension
- Compression

Figure 16

Typical stress distributions due to various forms of surface hardening

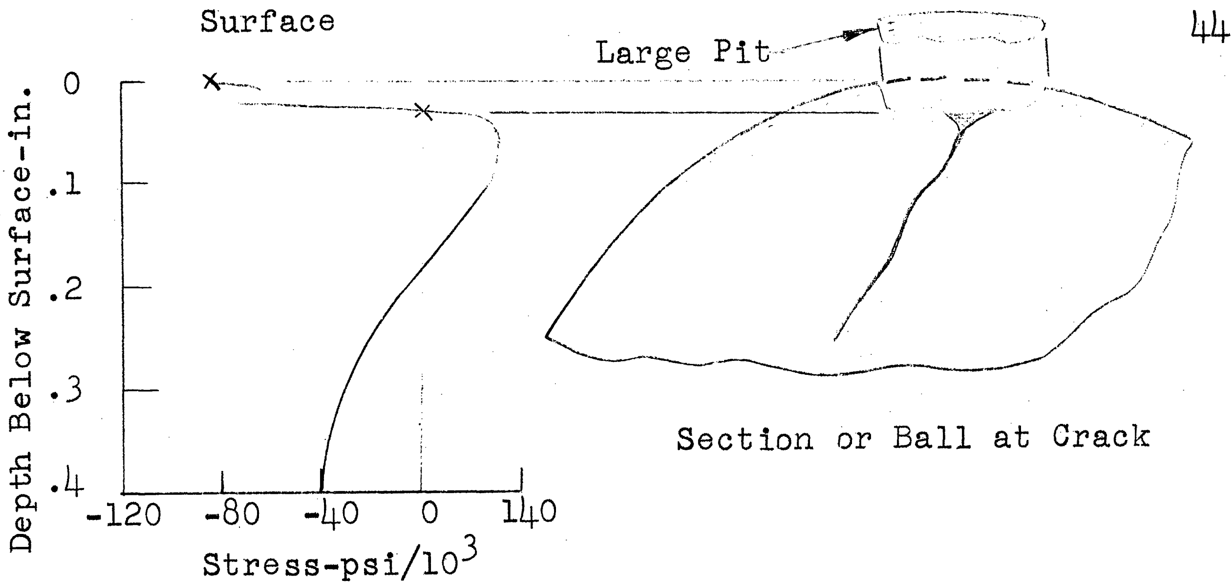
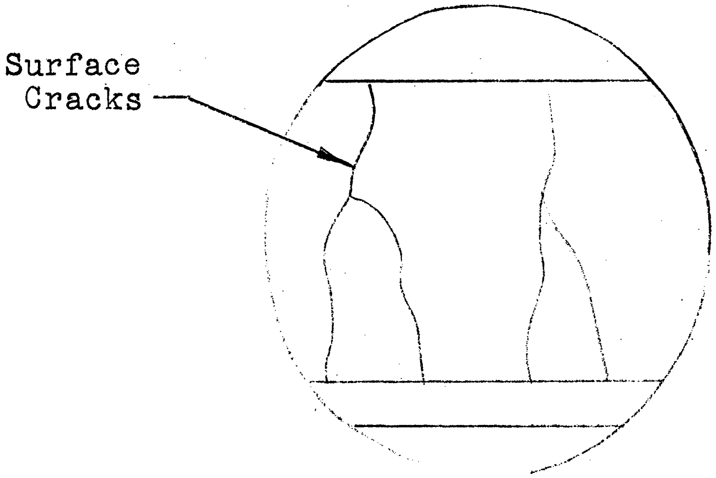


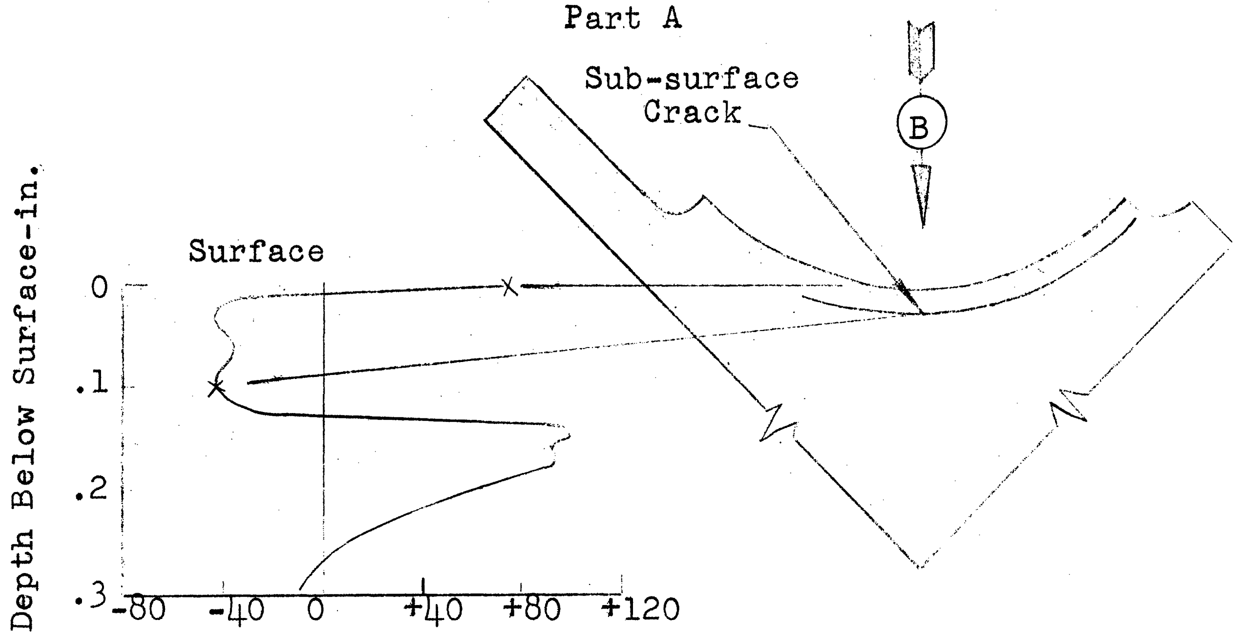
Figure 17
Spalling Type Failure in a 13/16 in. Diamter Ball Bearing

sequently broke through to the surface. It can be seen that the fracture path running parallel to the surface is located in the area of a high stress gradient. In this zone, the residual stress ranges from a fairly high compressive stress to a tensile stress in a very short distance. The fracture finds it difficult to penetrate the layer of metal residually stressed in compression, but apparently the steep stress gradient plus the stresses imposed by contact with the race were sufficient to encourage propagation in a direction parallel to the surface.

Another example of spalling is shown in Figure 18. Here the surface of an induction hardened bearing was found to be full of cracks, as shown in (a), and under tensile stress due to grinding. Results of a residual stress analysis are shown schematically in (b). Again the spalling is associated with a severe stress gradient. There is a tensile residual stress at the surface and a compressive stress beneath it. The fracture



View in Direction of Arrow B
Part A



Part B

Figure 18
Spalling Type Failure in a Large Bearing Race

penetrated vertically until it encountered a layer of material residually stressed in compression and then traveled parallel to the surface.

From these examples it can be generalized that if subsurface fracture planes parallel to the surface and vertical planes below are observed, there exists a surface layer stressed in compression with tensile stresses below (Figure 19a). If

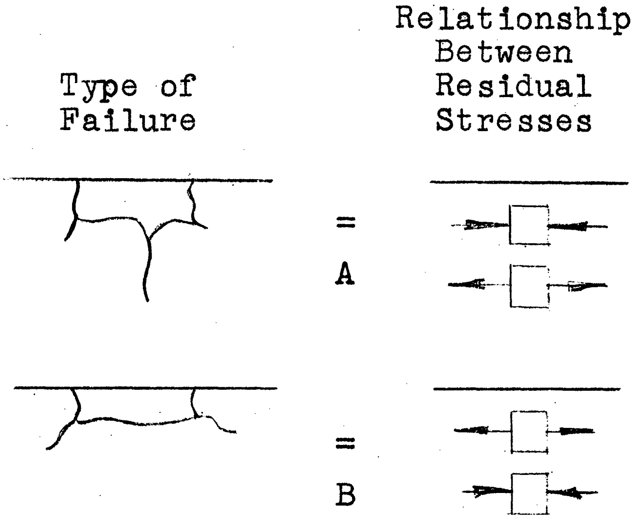


Figure 19
Spalling Versus Residual Stresses

vertical fracture planes in a surface layer in combination with a fracture plane parallel to the surface are observed, a steep stress gradient with surface layers in tension and the material below in compression can be expected. (Figure 19b)

Flaking[16]

Flaking is a common phenomenon in steel and weld deposits and can occur during the fabrication or subsequent use of a part. It is characterized by hairline or shatter cracks and is generally attributed to the combined action of residual tensile stresses and hydrogen embrittlement. Tensile stresses arising from volume increases during transformation, particularly in segregated areas, are more conducive to this type of failure than stresses of thermal origin.

Hydrogen alone has never been found definitely to cause cracks, but it is reflected in low ductility; also residual stresses alone appear unable to produce flaking in the absence of hydrogen.

Two of the most common incidences of flaking are those which occur during the cooling of large sections of alloy steels and in railroad rails. In the latter case, this phenomenon is often referred to as "shelling" and it is apparently due to high repetitive rolling contact stresses of the car wheels.

Grinding Cracks[35,36,37]

Sources

Grinding cracks are manifestations of high tensile stresses induced by grinding. The exact mechanism of stress generation during grinding is highly speculative. However, it is generally agreed that the removal of metal by the wheel causes a considerable heating of a small volume of metal immediately beneath, and this may be sufficient to raise its temperature to the plastic range. The metal expands, and, since the rigidity of the surrounding metal prevents lateral expansion, only vertical expansion takes place causing a bulge which is at once leveled by the wheel. The remainder of the heated volume, however, is permanently upset and, as the wheel passes on, the hot metal is chilled rapidly. It now tends to shrink and, because there is actually less metal to fill the volume than is required, the metal is stretched and remains highly stressed if not actually broken. The cracks, either those which appear at once or those appearing after pickling, always seem to be a right angles to the direction of grinding. This action may also account for the wavey appearance of some ground surfaces.

Another possible source of surface stress is the fact that any chip-forming operation leaves the freshly formed surface in a stressed condition purely as a result of the plastic deformation that must precede rupture of the metal. In addition, stresses may be set up as a result of metallurgical charges induced by heating such as overtempering, reheating, tempering, or transformation of retained austenite. Annealed steel capable of being heat treated can also be stressed by reheating caused by abrasive grinding.

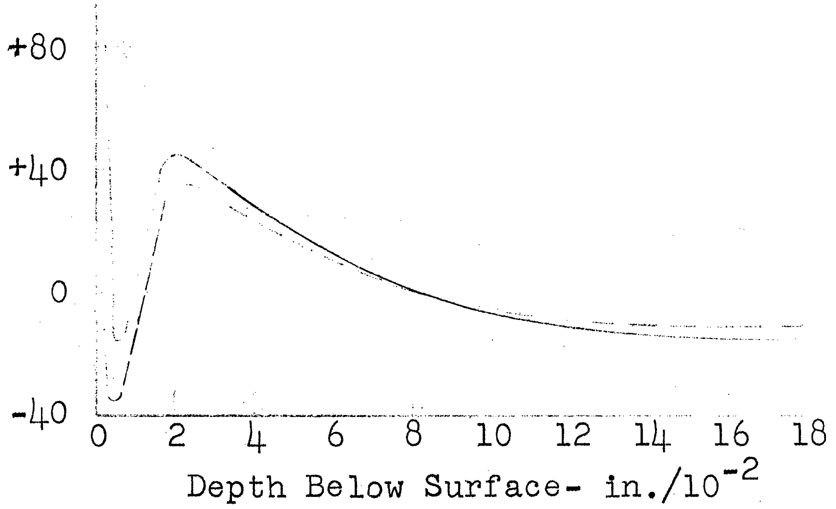
Appearance

The distinguishing feature of grinding cracks is that they are always fine,shallow, and bear some definite relationship to the grinding marks or scratches left by the abrasive grains of the wheel. By contrast, any cracks that may have been present in hardened steel prior to grinding, such as those occurring in heat treatment are likely to be deep and related to the general geometry of the piece rather than to the grinding pattern.

Grinding cracks in flat surfaces are always found to be primarily perpendicular to the grinding marks. Typical depths of such cracks are 0.010 to 0.020 inches due to the extremely shallow penetration of the induced residual stresses.(see Figure 20.)

If the crack pattern is highly pronounced, there may be cracks joining the perpendicular ones and forming a network as shown in Figure 21.

The same type of cracks may be found in curved surfaces, although here they may deviate more from the nominal direction.



Parallel to the direction of abrasive travel
Perpendicular to direction of abrasive travel

Figure 20

Residual stress distribution in annealed manganese oil-hardened tool steel due to surface grinding.

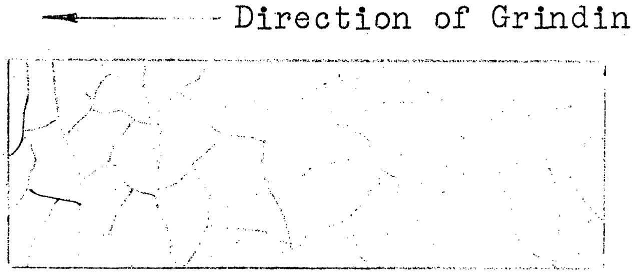


Figure 21

Network of Grinding Surface Cracks

Cracks parallel to the grinding marks can also be encountered especially when threads are ground.

A form of spalling may occur in carburized steels, primarily under conditions of broad grinding contact between wheel and work. The relationship of spalling type failures to residual stresses has been previously discussed.

Delayed Cracking[36,38]

This phenomemon can be attributed entirely to surface stresses. High stress concentrations tend to relax and thus become lower, but his very process of relaxation may change the stress pattern in such a way as to build up localized stresses elsewhere beyond the breaking point. If some of the microconstituents have been rendered less stable by machining, they may transform and upset the existing stress balance so as to cause delayed cracking.

Since stress relaxation causes delayed cracking, fracture will occur only after a time interval. Examples of such failures include the fracturing of a projectile tip some time after it has passed through armor plate; the cracking of a thick-walled steel tube on a radial plane at the bore sometime after being released from external pressure; the delayed explosion of steel ball bearings upon removal from a tumbling barrel; the spontaneous cracking of quenched and tempered steel helmets weeks after being hit by test bullets; and the spontaneous and violent aging rupture of armor-piercing shells after hardening.

Cracking of Metals in the Presence of Liquid Metals [16,38]

Liquid metals can crack solid metals that are residually stressed as shown in Figure 22. This cracking is due to intergranular penetration of the liquid metal into the cracking metal; however, without residual stress this cracking never occurs. For example, properly stress-relieved cast iron parts will hold liquid solder or tin for years without failure although they will fail in 24 hours if not stress-relieved. Aluminum solder will crack aluminum alloys if they contain residual stresses. Similarly during brazing operations intergranular penetration of the steel by molten brazing alloy has been known to occur and cause cracking. Steel will crack when exposed to molten tin, antimony, white metal, aluminum, copper, and other metals. Copper alloys, though immune when stress free, will crack when stressed if exposed to lead, tin, solder and mercury (at room temperature).

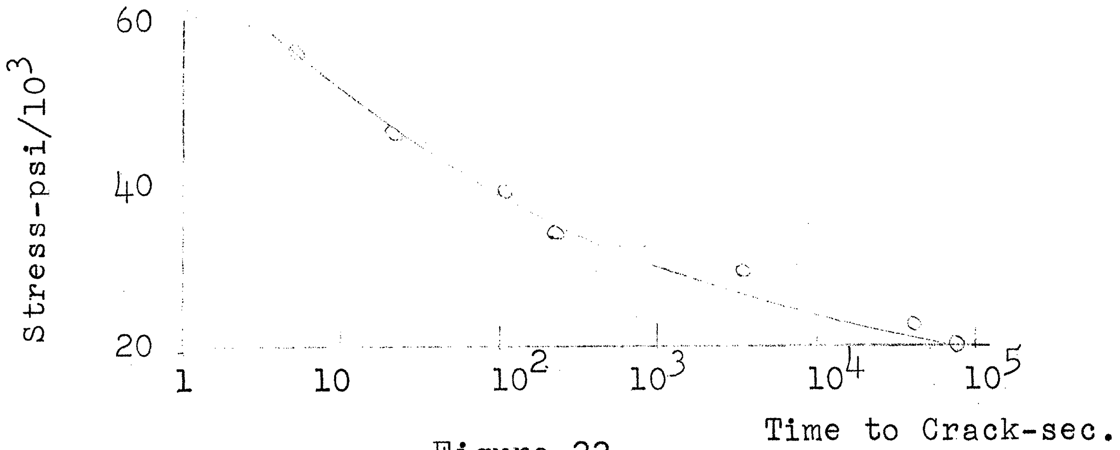


Figure 22

Cracking Time for a Stressed Cartridge Brass Immeresed in Mercurous Nitrate

Stress-Corrosion Cracking[31, 39]

Many metals in the presence of certain corroding media will develop cracks, if, in addition, the metal is stressed, this type of attack will be enormously accelerated even to the point of explosive violence. This phenomenon is termed stress-corrosion cracking.

If such cracking is to occur, there must be tensile stresses in the general neighborhood of the yield strength, but failures have occurred at much lower stresses. At these stress levels, a great deal of corrosive action would be needed to produce failure if it were not for an interaction between stress and corrosion.

Although the exact nature of the phenomenon is not known, stress-corrosion cracking has been observed in all metal systems. Yet for each metal, specific environments are required. No stress-corrosion cracking has been observed in a vacuum. The environment that induces cracking frequently attacks the metal only superficially if stresses are absent. The tensile stresses aid the replenishment of the corrosive agent within the crevice. Susceptibility to cracking increases with an increase in tensile stress as does the time to crack in a given corrosive media. The fractures are brittle, showing no localized deformation near the crack.

Cracking of iron and steel when exposed to hot caustic solutions; copper alloys when exposed to ammonia solutions; silver alloys to ferric chloride solutions; and aluminum, magnesium and even gold alloys to other corrosive media have been observed. In brasses, cracking rarely occurs with internal stresses less than 12,000 psi, but stresses of 12,000 to 20,000 psi produce cracking.

Heat Cracking

Internally stressed metals when heated can crack. While heat can relieve residual stresses, the process in not instantaneous and consequently residual stresses can operate in metals for finite times. This effect can be increased by the addition of thermal stresses when heating is sudden or rapid. The simplest explanation for such cracking is that residual stresses (in conjunction with thermal stresses) exceed the fracture stress of the metal (which falls ith rising temperature) and failure ensues. However, this explanation is somewhat oversimplified and does not hold for all forms of heat cracking.

Fire-Cracking

While thermal fatigue may at first glance seems like it isn't really a form of wear, it actually is, if wear may be defined as "deterioration due to use." Thus, thermal fatigue may be included in a discussion of wear mechanisms. Evidence of thermal fatigue may be found in the form of small cracks on the surface of objects which have alternately heated and cooled many times. Sometimes this type of cracking is called "fire-cracking."

Basically, there are two types of thermal failures which will be discussed in this section. The first is thermal shock failure, which occurs when a part shatters after it has been heated and then suddenly cooled. The sudden change in temperature is termed thermal shock, and if failure occurs after only one cycle, the failure is termed a thermal shock failure. Brittle materials, such as glass, are subject to failing in thermal shock, while ductile materials will deform, and seldom break.

However, if the heating and cooling cycling is continued, the part is subjected to thermal fatigue, which may or may not cause failure, depending upon how many times the cycling occurs, and on the temperature range over which it is cycled.

The mechanism of thermal fatigue is as follows: As the surface of a part is heated up, it tends to expand. But, this expansion is resisted by the cooler interior of the part, so that compressive stresses are built up. And, when the yield strength of the material is exceeded, there will be plastic flow in the hottest protion of the part. The amount of plastic flow will depend upon the magnitude of the compressive stress, which in turn is dependent upon the temperature difference between the exterior and interior of the part. As the temperature in the interior becomes nearer the temperature of the surface, the stress will decrease. However, there will be residual compressive stresses, due to the plastic flow which took place.

Upon cooling, these residual stresses decrease, due to the elastic recovery of the material, and eventually become zero for a short time. Then, the reverse of what happened during heating occurs. The exterior of the part will tend to contract, but is resisted by the warmer interior, causing tensile forces to be produced.

Since the warm interior of the part is weaker than the surface, it may possibly flow plastically. However, the greater size of the interior usually keeps the stress level down, so that the yield point of the surface may be reached first. Thus, the surface, which previously had been compressed, is now elongated. And, upon further cooling, high residual tensile stresses will be left in the part.

Thus, on further cycling, the residual tensile stresses must first be overcome before the compressive stresses formed upon heating can be built up. Therefore, it is entirely possible that the compressive stresses built up on the second and subsequent cycles will not reach the the yield point of the material, and thus no more plastic flow will take place. In this case, there will be no tendency for the part to fail during continued heat cycling.

However, if upon heating during the second cycle the compressive stresses do build up beyond the yield point and plastic flow again occurs, the alternating tension and compression stress pattern will continue. And, after a certain number of cycles, thermal fatigue will occur.

The factors which affect thermal fatigue are the temperature gradient invclved; the coefficient of thermal expasion, a low coefficient rendering a material less susceptible to thermal fatigue; the high temperature strength of the material;

and the ductility of the material at high and low temperatures, the low temperature ductility being more important.

To help alleviate the problem of thermal fatigue, several things can be done. Since thermal fatigue is due to stresses exceeding the yield point, parts subjected to heat cycling should be designed so as to lessen stress concentrations. In addition, materials with high yield strengths and endurance limits can be used. Also, materials with low coefficients of thermal expansion should be chosen. And, if possible, the temperature gradient should be as small as possible, although control of this factor is often impossible.

Firecracking is characterized by gross failures extending the length of a tube; across the width of a sheet; or a maze of cracks on a strip surface. The failures are brittle and occur rapidly. The presence of certain low-melting impurities (such as lead or tin in copper alloys) greatly increases the susceptibility of an alloy to fire-cracking.

Stress-Precipitation Cracking

Although the effects of this phenomenon are similar to, and often confused with, those of fire-cracking, there are important differences in their causes.

When a dispersed phase of an alloy is precipitated by heat treatment (as in aging) residual stresses are developed around each particle of precipitate because of differences in density of the precipitate and the matrix from which it formed. If the precipitate is more voluminous, the matrix will be placed in tension tangentially (the radial stresses being compressive) and the precipitate will be put in compression. Under these circumstances aging can cause cracking (especially when residual stresses are present from other sources); the fracture courses through the matrix that is under tension. Copper-chromium alloys belong to this class. If the precipitate is more dense, the matrix will be placed in compression and the precipitate will be in tension. In such alloys cracking does not run through the matrix. Because of aluminum's lightness, the matrix of aluminum alloys is usually thrown into compression, hence they are generally immune to cracking during aging.

Thermal Fatigue of Bearing Alloys[16]

Thin strips of tin-base alloys exhibit marked cracking when repeatedly heated and cooled through a temperature range of 85 to 300°F. Noncubic metals like tin, cadmium, and zinc alloys (99% pure) behave in a similar manner. Rates of heating and cooling were found to be of little significance. This phenomenon is termed thermal fatigue, and it is a direct result of residual stresses.

The cause of the residual stresses which lead to this phenomenon is explained on the basis that these metals possess marked anisotropy of thermal expansion, and repeated temperature changes produce plastic deformation; cubic metal lead exhibits no deformation or failure of this type. Metals which possess a degree of anisotropy of thermal expansion cannot be obtained in the strain-free condition at room temperature by casting or annealing. This point may have some practical explanation as regards service failures in some bearings.

Warping [38,39]

Cutting into a metal that has residual stresses will cause it to deform due to the release of residual stresses and subsequent elastic deformation. Machining the skin off cold drawn and quenched bars (as in cutting screw threads, milling slots or keyways or machining a shaft to close tolerances) will cause "snaking", especially when the amount of skin machined off is not symmetrical about the axis.

Warping can cause secondary effects. A drilled hole may close slightly as a result of the release of residual stresses; drill and reamers snap off because of the consequent bending; and the metal is blamed for poor machinability.

Breaking of Metals when Cut[38,39]

When a metal is parted, the pattern of residual stresses is changed throughout the rest of the body. This change may raise the stress above the tensile strength of the body at some point and failure will result.

Milling the notch into Charpy bars of high-strength, brittle steels often opens a split in advance of the notch. The original residual stress pattern and the change in it wrought by the milling of the notch lead to the breaking. The shattering of the pre-stressed glass of an automobile rear window when scratched is a familiar example of this phenomenon.

6. THE CONTROL OF RESIDUAL STRESSES

From the preceding section it is obvious that residual stress patterns can be deleterious to wear and service life of machine elements by causing warping cracking, spalling, flaking, thermal fatigue, fire-cracking, stress-corrosion, etc. Because of their diverse effects, the control of residual stresses is of great importance. Also, because of this variety of effects, any specific situation involving residual stresses will require experimentation before they can be kept under full control.

The term "control" is used here in a general sense. It relates to any measure that leads to or assists in arriving at the optimum combination of service properties that are affected by the presence of residual stresses. In general, their measures can be classified from the viewpoint of production as follows:

1. Selection of the most suitable material.
2. Development of processing and heat-treating methods that are conducive to the most favorable residual-stress conditions.
3. Close control of finishing operations, such as grinding and straightening.
4. Stress-relieve annealing.
5. Application of a final mechanical treatment which favorably affects the residual stress pattern. This is termed mechanical stress-relief.

The most important single factor in regard to intelligent control of residual stresses is that it should, and in many instances must, be applied not only to the last stage but also to earlier stages of the production of a part in order to achieve full success.[19]

The extent of using the various control measures depends largely on the particular application of the part (ie. the nature and combination of the properties required to obtain optimum service performance). Their properties are either entirely of a mechanical nature, a combination of mechanical and chemical resistance; or a physical property such as dimensional stability during machining operations.[19]

With these points in mind, each of the above mentioned classifications of control measures for residual stresses will now be discussed.

Selection of the Most Suitable Material

The most important consideration for the designer in this respect is a careful evaluation of the strength and shape requirements of the proposed machine element as related to the chemical and metallurgical factors which cause residual stresses in the various metals; that is, precipitation (aging characteristics), bulk transformations (cooling characteristcs), and chemical compatability or reactivity (with the environment and other mating materials). The effect of residual stresses relative to specific materials have been discussed previously in connection with delayed cracking; cracking in the presence of liquid metals; stress-corrosion cracking, stress-precipitation cracking; and thermal fatigue. Their relationships will not be discussed here, but there are many references in the literature on this subject. If should be mentioned however, to emphasize the magnitude of this problem, that four steels within one hardenability specification, of H band, can give different residual stresses.[40] Hence, for steel, although there may well be a significant relationship between hardenability and residual stresses, many other factors must be considered.

Development of Processing and Heat-Treating Methods

The variety of metal processing methods suggests that considerable variation is possible in the stresses trapped in the surface layers of a part. Table 1 illustrates that several factors are involved in the introduction of residual stresses during various processing operations.[35] Trapped stresses can be

Residual Stress Introduced or Altered By	Temp. Difference	Phase Transform	Cold Plastic Flow	Warpage	Altered Chemistry
Heat-treatment (quenching and tempering)	Yes	Yes	No	No	No
Case Hardening (carburizing and nitriding)	Yes	Yes	No	No	Yes
Induction hardening	Yes	Yes	No	No	No
Machining	Possibly	Unknown	Yes	Yes	No
Grinding	Yes	Unknown	Yes	Yes	No
Cold working (shot peening, surface rolling, tumbling, lapping, blast cleaning)	No	No	Yes	Yes	No
Straightening	No	No	Yes	No	No

Table 1

Types of Residual Stress Introduced by Processing Operations introduced by high mechanical stresses associated with the severing of a metal. In the case of machining operations, these are associated with the geometry of the cutting tool, speed of cutting, nature of cutting, etc. If sufficiently high temperatures are developed locally, thermal stresses or even transformation stresses might be introduced as previously discussed in connection with grinding. Also, in all metal removal processes, if the metal removed was originally stressed, some unexpected change in shape or size would occur, causing a change in the residual stress in the remainder.

In developing production methods to control residual stresses, many factors must be considered. For example, surface compressive residual stresses have been induced in test specimens by a "gentle grinding" process as opposed to the unfavorable tensile stresses which are usually induced by ordinary grinding.[41, 42] However, this "gentle grinding" cannot be practiced in commercial work. Commercial grinding demands methods that will not unduly hamper production. Grinding wheels of the softness (32A46-H8VBE) that is required for "gentle grinding" reduce production rates because increased wheel wear makes it necessary to dress the wheel more frequently. The shallow depth of cut that is necessary (downfeeds of from .0001 to .0005 as opposed to range of .001 to .005 for production grinding) in gentle grinding cannot be tolerated in high rate production shops because of the limitations on output that would be imposed.

To solve this problem the best that can be expected is a compromise in the grinding practice as applied to highly stressed parts. A compromise which is in extensive use, is to grind in such a manner as to reduce the induced residual tensile stress to values well below the spontaneous fracture (stress-corrosion cracking) point. The ground surface, as well as other surfaces, of the part may then be subjected to a prestressing treatment such as shot peening, rolling, or tumbling, whereby the harmful residual tensile stress is converted to beneficial residual compressive stress.*

The above procedure results in a fatigue strength of ground surfaces greater than is possible by the most gentle grinding as shown in Figure 23.

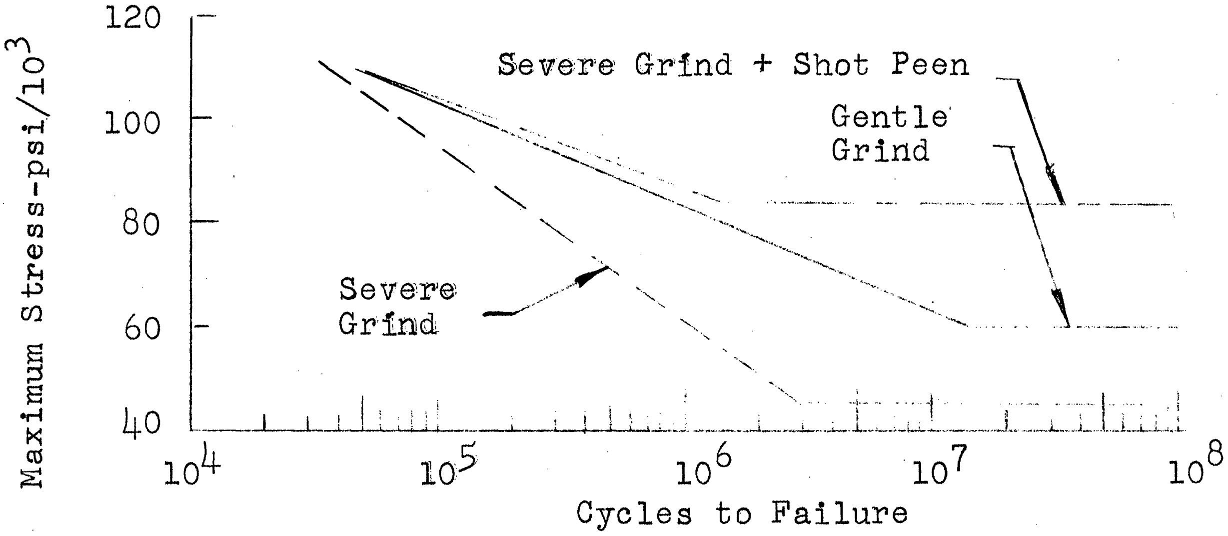


Figure 23

S-N Curves showing the beneficial effect of shot peening, as opposed to that of gentle grinding in improving fatigue strength. Steel with added vanadium-Rockwell Hardness C45.[41]

Control of Finishing Operations

After the above example of the favorable effect of residual compressive stresses by cold working (shot peening), it may be

---

* As described by Almen, J. O., in the discussion to Reference 41.

shown by the example of a gear how the control of finishing operations can create residual stresses which influence the strength of its teeth.

When loading a tooth under service conditions by a load on the tooth flank, a stress distribution arises, as shown in Figure 24a. The highest stress occurs on the tooth root and

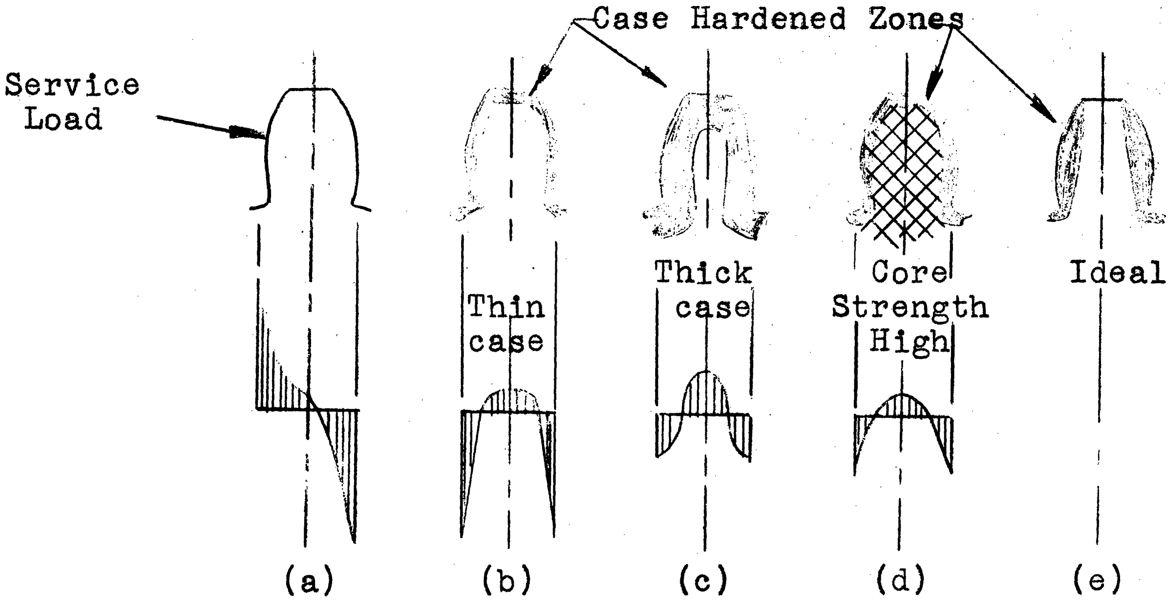


Figure 24
Load and Service Stresses in a Gear with Various Case Depths

exhibits a certain asymmetry because the tooth load possesses a radical component, whereby the compressive stress on the unloaded side of the tooth becomes a little higher than the tensile stress. To resist wear a tooth flank must be hardened. In case hardening steel gears, a surface layer of martensite is created, which causes a distribution of residual stresses, as shown in b.

The essential point here is that in the core of the tooth a sufficient part of the cross-section must remain unhardened, so that relatively low residual tensile stresses in the core can balance a high residual compressive stress on the surface. If, on the other hand, the hardened zone is excessively thick as in (c), residual stresses of only small amounts appear. Also, if the core itself hardens to a high degree martensitically, (favored either by a high content of carbon in the basic material - illustrates the importance of choice of material - or by too high a hardening temperature) only small residual compressive stresses are developed in the surface (d). From these considerations it may be concluded that the highest fatigue limit values in the tooth are obtained if the hardened layer is made relatively thin and, above all, the core strength is not too high.*

It is not advisable to keep the core layer as thin as possible for reasons of the fatigue limit in the tooth root, because in the tooth flank (especially with regard to pitting at the pitch line) a relatively thick hardened zone is required. Ideally, therefore, these contrary demands would be satisfied by a hardening process, as in (e). Such a contour hardened zone is not attainable with core-hardening, but it can be realized by inductive gash hardening. It should be mentioned, however, that the exact and economical manufacture by this hardening process is not yet completely developed. Furthermore, in the special case of the gear, induction hardening has the disadvantage that directly under the hardened zone there is a thin annealed layer which has a strength somewhat lower than the tempering strength of the core material. This may, in certain instances, unfavorably

---

* These assertions have been adequately proven in Reference 21 from where this discussion was abstracted.

Material	Control Measures
Stress-Corrosion Cracking	
Aluminum Alloy Steels heat-treated to high hardness	1. Stress relieve anneal at a temp. slighlty below tempering temp. 2. Mechanically surface treat (shot peening)
Stress Corrosion and Hydrogen Embrittlement	
Aircraft and other steels, heat-treated to high hardness	1. As above 2. As above 3. Apply 1 & 2 prior to exposure to hydrogen 4. Bake at 300-400°F. for several hours (removes hydrogen from metal)
Season Cracking and Related Failures	
Brass and Copper alloys subjected to drawing and spinning Carbon and low-alloy steels Stainless Steels Aluminum Alloys	1. Stress-relief anneal (does not adversely affect hardness 2. Apply finishing operations of mechanical nature (stretching, compressing, bending, reeling, polishing). 3. Proper selection of tool contour and reductions. 4. Proper combination of depth of draw and corner radius
Fire-Cracking	
Alloy Steels	1. Eliminate low-melting impurities (if not intentionally added constituents of the alloy) 2. Quench properly (See Table 3) 3. Anneal at a slow heating rate (reduces residual stresses and minimizes thermal stresses)
Stress Precipitation Cracking	
Age-Hardening Alloys	1. Mechanically surface treat (surface rolling, shot peening, straighten)

Table 2a

Typical Methods of Controlling the Effects of Residual Stresses

Finishing Operation	Control Measures
Grinding	1. Heat to low temperature before grinding (decreases stresses induced during grinding).
	2. Heat to low temperature after grinding (decreases stresses induced by grinding).
	3. Reheat to at least 300°F. after grinding for precision parts. Preferably done after rough and followed by careful finish grinding (helps maintain dimensional stability)
	4. Reduce wheel speed
	5. Reduce down feed
	6. Use grinding fluid a. Lubrication in grinding fluids is more important than cooling effect b. Air, solutions of rust inhibitors and soluble oils all have about the same effe c. Straight grinding oils have produced compressive surface residual stresses at optimum conditions (gentle grinding) d. Sulphurate oils at optimum conditions produce more desirable stress distributio than other fluids.
	7. Use "soft" grinding wheels
Welding	1. Use case in fitting
	2. Use backstop procedures (welding of short increments in the opposite direction to which welding is progressing and fuse each increment into the start of the preceding one)
	3. Use wandering procedures (welding of short increments at random throughout the structure)
	4. Partial restraint while welding
	5. Keep the welding evenly balanced both vertically and horizontally
	6. Shot peen after welding
	7. Preheat
	8. Post heat treat (anneal)
	9. Erection by subassembly.

Table 2b

Typical Methods of Reducing Residual Stresses due to Various Finishing Operations influence the fatigue limit in the tooth root as well as the development of pitting.

Other methods of controlling the effects of residual stresses by control of finishing operations are illustrated in Table 2.

Stress Relief Annealing[38]

General relief and diminution of residual stresses can be obtained by heating the parts. The residual stresses are eliminated in two ways viz:

1. Heating decreases the yield strength of the metal so that residual stresses deform the warm metal plastically and are thereby relieved. Stress relief by this method is immediate on heating but is limited in degree. The stresses can never be cut below the yield strength of the metal at the temperature used. Such relief of stresses is sometimes harmful, however, the plastic deformation that relieves the stresses is a form of warping that may be serious in parts requiring close dimensional control. Plastic deformation can also cause fracture, as noted in connection with fire-cracking.
2. Residual stresses theoretically will relax or die out at any temperature if given enough time. It should be remembered, however, that the creep of ferrous metals at room temperature is probably measured in terms of hundreds of thousands of years. Thus storing of castings for several months will not in general stress relieve them. Heating accelerates fading and brings the time for stress relief within reasonable production time. In contrast with the first method of stress relief by heating,stress relief by this method is not limited to any temperature, but neither is it immediate in effect. Parts that might crack if heated suddenly (fire-cracking) can frequently be stress relieved without trouble if heated slowly. It should be remembered that heat and time can anneal a metal in the normal sense-that is, soften the metal and eliminate the effects of cold work.

   It is frequently important to know whether annealing at a given temperature and for a given time will cause a metal to lose its temper. In some metals (such as brass) stresses may be relieved without softening. In other metals (for example, steels) some overlapping occurs and a compromise must be made; some softening must be tolerated to get complete stress relief or partial stress relief must be tolerated to avoid softening.

Table 3 lists temperatures and times commonly used in stress relieving of metals and alloys. Times shorter than those given afford less relief; longer times may bring on softening. Higher or lower temperatures than those given may be used, provided the time at temperature is shortened or lengthened accordingly. The general rule of thumb is to cut the time in half for each 10F increase in temperature contemplated or to double it for each 10F decrease contemplated. Instances of partial stress relief without softening (as in steels)are indicated by parentheses in Table 3.

Mechanical Stress Relief[38]

It is not always possible to eliminate residual stresses by heating. For example, where stress precipitation cracking is a factor, any attempt to stress relieve the metal thermally brings on the very precipitation that causes cracking; in cold drawing, annealing often involves temperatures at which there is some risk of deteriorating surface quality or reducing tensile properties.[43]

In other instances mere reduction of stress does not give the maximum benefits. Tensile residual stresses may be detrimental to fatigue life; relief of them may increase fatigue life; but substituting compressive residual stresses for tensile stresses can approximately double the benefits gained by relief (as in the grinding example previously discussed). For these reasons, residual stresses are frequently relieved or changed by mechanical working methods based on the principle that any cold working process will superimpose its constructive residual stress pattern on any that already exists in the metal.

Stretching a metal only slightly, but enough to put it into the plastic state, reduces any residual stresses.

al	Temperature, (a) degree Fahr.	Time at Temperature, (b)hr.
ay cast iron	800 to 1100	5 to 1/2
Carbon Steel		
ss than 0.35% C, less than 3/4 in.	Stress relief usually not required (c)	
ss than 0.35% C, 3/4 in. or greater	(1100 to 1250)	1
e than 0.35% C, less than 1/2 in.	Stress relief usually not required (c)	
e than 0.35% C, 1/2 in. or greater	(1100 to 1250)	1
ecially killed for service at low emperature	(1100 to 1250)	1
Carbon-Molybdenum Steel (all thicknesses)		
ss than 0.20% C	(1100 to 1250)	2
20 to 0.35% C	(1250 to 1400)	3 to 2
Chromium-Molybdenum Steel (all thicknesses)		
Cr, 0.5% Mo	1325 to 1375	2
25% Cr, 1% Mo and 5% Cr, 0.5% Mo	1350 to 1400	3
Cr, 1% Mo	1375 to 1425	3
Chromium Stainless Steel (all thicknesses)		
pes 410 and 430	(1425 to 1475)	2
pe 405	Stress relief usually not required for thicknesses less than 3/4 inc. (c)	
Chromium-Nickel Stainless Steel (d)		
pes 304, 321, 347	Stress relief usually not required for thicknesses less than 3/4 in. (c)	
pe 316, more than 3/4 in	1500	2
pes 309 and 310, more than 3/4 in.	1600	2
Welding Dissimilar Materials (d)		
-Mo steel to carbon steel or to ny C-Mo steel	1350 to 1400	3
pes 410 and 430 to any other steel	1350 to 1400	3
-Ni stainless steel to any other etal	As required for the steel to which Cr-Ni stainless steel is joined	

Table 3

Typical Stress-Relief Treatments

Metal	Temperature, (a) deg. F.	Time at Te perature,(
Copper Alloys		
Copper	300	1/2
90 Cu-10 Zn	400	1
80 Cu-20 Zn, 70 Cu-30 Zn	500	1
63 Cu-37 Zn	475	1
60 Cu-40 Zn	375	1/2
70 Cu-29 Zn-1 Sn	575	1
85 Cu-15 Ni, 70 Cu-30 Ni	475	1
95 Cu-5Sn, 90 Cu-10Sn	475	1
	375	1
Magnesium Alloys		
M-1, Hard rolled sheet, 1.5 Mn	400	1
M-1, extrusion, 1.5 Mn	500	1/4
AZ31X hard rolled sheet, 3 Al, 1 Zn, 0.3 Mn	300	1
AZ31X extrusions, 3Al, 1 Zn, 0.3 Mn	500	1/4
AZ51X hard rolled sheet, 5 Al, 1 An, 0.25 Mn	375	1
AZ61X extrusion, 6 Al, 1 Zn, 0.35 Mn	500	1/4
AZ80X extrusion, 8.5 Al, 0.5 Zn, 0.15 Mn	400	1
AZ80X HTA extrusions, 8.5 Al, 0.5 An, 0.15 Mn	600	1/4
Nickel Alloys		
Nickel and Monel	525 to 600	3 to 1
"K" Monel and "KR" Monel	525 to 600	3 to 1
80 Ni-20 Cr, Inconel, 60 Ni-25 Fe-15 Cr	700 to 900	3 to 1

(a) Temperatures in parentheses give partial relief of stresses

(b) The interval during which the entire piece is at temperature

(c) Unless necessary in order to obtain dimensional stability

(d) Rate of heating should be less than 200°F per hr. For the other Ferrous metals a heating rate of less than 400°F per hr. is recommended. All ferrous metals should be cooled from the stress-relieving temperature at a rate not greater than 200°F per hr.

Table 3 (cont.)

Typical Stress-Relief Treatments

Rolling, shot blasting, peening, and other surface working processes are used to induce compressive residual stresses at surfaces subjected to high fatigue stresses or stress corrosion.

These methods must be used judiciously, however. Stress precipitation cracking that occurs in drawn rods can be prevented if the residual stress pattern from drawing is added to that resulting from surface rolling. Since the two patterns are opposite in shape they can be combined so that residual stresses are essentially cancelled. On the other hand, experience shows that too light a surface rolling will cancel only those residual stresses near the surface which have resulted from the rod drawing, but fail to remove deeper residual stresses. Cracking can still occur below the surface, where it is more harmful because it escapes superficial inspection but can still play havoc in subsequent machining operations or fatigue performance.

7. MEASUREMENT OF RESIDUAL STRESSES

The ability to obtain accurate measurements on a repetitive basis is on of the primary requisites in any experimental investigation. Since a number of methods have been devised and used to measure the level and distribution of residual stresses, it appears that no single technique is wholly satisfactory. The more sensitive ones are tedious and time consuming, and when employed by different individuals, even these have yields variable answers. For these reasons, many findings are questionable and a single conclusive approach is still sought. A few methods will be presented in brief, but for a wider and more detailed coverage, references,2,12,13 and 14 should be consulted.

1. Sach's Boring and Turing Method

Wire resistance strain gages are adapted to the inside or outside diameter of a cylindrical test specimen. As successive layers are turned off or bored out, strain measurements are obtained and used in pertinent equations to calculate the magnitude of longitudinal and tangential stresses.

2. Layer Removal Method

From an initially flat test specimen, successive layers of metal are etched and measurements of the change in curvature of the specimen are made. These values are used in pertinent equations and calculations of biaxial stresses are obtained.

3. X-Ray Diffraction Method

An X-Ray source of known wave length is exposed to the work surface and measurements of the diffraction angle are obtained. By utilizing Bragg's Law, $\lambda = 2d \sin \theta$, the interplannar spacing is determined, and with appropriate equations, residual stresses are calculated.

4. Acoustoelasticity

This is very recent approach that utilizes polarized ultrasonic waves which are emitted to the test specimen. With proper equipment, any changes in the angle of polarization are measured. These changes are then employed to determine the magnitude and sense of internal residual stresses. Due to its recent development, this method has not been widely used, but at this time appears to possess great potential.

BIBLIOGRAPHY

1. Cold Working of Metals, The American Society for Metals, Cleveland, Ohio, 1949, pp. 34,41,48.
2. Metals Handbook,1955 Supplement, pp. 89 96, 120.
3. Letner, H. R., "Residual Grinding Stresses in Hardened Steel", ASME Trans., Vol. 77, No. 7, pp. 1089-1098, October, 1955.
4. Heriksen, E. K., "Residual Stresses in Machined Surfaces", ASME Preprint No. 50-SA-27, April 24, 1950.
5. Colwell, L. V., "Machinability Studies of X4340 Steel at High Strength", The University of Michigan, ERI Report, No. 2133-1-F, January, 1954.
6. Colwell, L. V., "Residual Stresses in Metal Cutting", Lectures given at the University of Michigan, 1958 Summer Conference Course, Published by the Industry Progam of the College of Engineering under the title "Engineering Approach to Surface Damage".
7. Tarasov, L. P., Hyler, W. S., and Letner, H. R., "Effect of Grinding Conditions and Resultant Residual Stresses on the Fatigue Strength of Hardened Steel", Reprint from the Proceedings, ASTM, Vol. 57, 1957, Philadelphia.
8. Letner, H. R., "Grinding and Lapping Stresses in Manganese Oil-Hardening Tool Steel", Trans., ASME, Vol. 75, 1953.
9. Colwell, L. V., Sinnott, M. J., and Tobin, J. C., "The Determination of Residual Stresses in Hardened Ground Steel", Trans., ASME, Vol. 77, pp. 1099-1105, 1955.
10. Halverstadt, R. D., "Analysis of Residual Stresses in Ground Surfaces of High Temperature Alloys", ASME Preprint No. 57-SA-62, Presented at San Francisco, June 1957.
11. Clorite, P. A., and Reed, E. C., "Influence of Various Grinding Conditions Upon Residual Stresses in Titanium", ASME Preprint No. 56-A-44, Presented at New York, December, 1956.
12. "Evaluation of Methods for Measurement of Residual Stress", General Information Report No. TR-147, Prepared by Division 4, Society of Automotive Engineers, Inc., September, 1957.
13. Letner, H. R., "Application of Optical Interfence to the Study of Residual Surface Stresses", Reprint from the Proceedings of The Soceity of Experimental Stress Analysis, Vol.X, No. 2, 1952.
14. Benson, R. W., and Raelson, V. J., "Acoustoelasticity", Product Engineering, July 20, 1959.
15. ASM Committee on Residual Stresses, "Residual Stresses", Metal Progress, August 15, 1955.

16. Horger, O. J., "Residual Stresses", Handbook of Experimental Stress Analysis by Hetenzi, M., London, Wiley and Sons, Inc., 1950.

17. Boas, W., and Honeycombe, R. W., "Thermal Fatigue of Metals", Nature, Vol. 153, April 22, 1944.

18. Sachs, S. and Van Horn, K. R., "Practical Metallurgy," Pub. American Society of Metals.

19. Sachs, S., "Control of Residual Stresses in Practice"., SAE Trans. Vol. 63, 1955.

20. Barrett, C. S., "Internal Stresses", Metals and Alloys, Vol. 5, October 1934.

21. Sigwart, H., "Influence of Residual Stresses on the Fatigue Limit", Reprint of the Institute of Mechanical Engineers, September 1956.

22. Lipson, C. and Thiel, D. H., Applications of Stress Analysis, Ann Arbor, University of Michigan, 1959.

23. Rosenthal, D., and Sines, G., "Effect of Residual Stresses on the Fatigue Strength of Notched Specimens", Reprint of the American Society for Testing Material, 1951.

24. Kudryavtsev, I. V., "The Influence of Internal Stresses on the Fatigue Endurance of Steel", Reprint of the Institute of Mechanical Engineers, September 1956.

25. Buhler, H., and Buckholly, H., "The Effect of Residual Stress on the Dynamic Bending Strength", Mitteitungen Forschungs Insitute, Dortmund, Vol. 3, No. 8, September 1933.

26. Nachtman, E. S., "Effects of Residual Stresses in Cold-Finishing Steel Bars", Product Engineering, Design Digest Issue, Mid-October, 1956

27. Almen, J. o., "Fatigue Weakness of Surfaces", Product Engineering, November 1950.

28. Setty, S. K., Lapsley, J. T., and Thomsen, E. G., "Stress Alter Hardness", Trans. ASME, Paper No. 57-A-77, 1957.

29. Lipson, C., Engineering Approach to Surface Damage, The University of Michigan, 1958.

30. Polushkin, E. P., Engineering Foundation Report No. 6, 1927.

31. Horger, O. J., "Residual Stresses", Metals Engineering, ASME Handbook, McGraw-Hill Book Co., 1953.

32. Moore, N. F., "Shot Peening", American Wheelabrator and Equipment Corp.

33. Lipson, C., "Failure of Tractor Wheels", an internal report of the John Deere Waterloo Tractor Works, December 1952.

34. Roland, E. S., "Resistance of Materials to Rolling Loads", Engineering Approach to Surface Damage, University of Michigan, 1958.

35. Mattson, R. L., "Effects of Residual Stress on Fatigue Life of Metals", SAE Trans., Vol. 62, 1954.

36. Tarasov, L. P.. "Some Metallurgical Aspects of Grinding", Reprint from Machining Theory and Practice, American Society of Metals, 1951.

37. Brophy, G. R., "Stresses and Cracks in Hardened and Ground Steel", Trans. ASST, 1930.

38. A.S.M. Committee on Residual Stresses, "Residual Stresses", Metal Progress, August 15, 1955.

39. Baldwin, W. M., "Residual Stresses in Metals", ASTM Edgar Marburg Lecture, 1949.

40. Boegchold, A. L., "Some Effects of Quenching and Tempering on Residual Stresses in Steel", Metal Progess, Vol. 57, February 1950.

41. Tarasov, T., and Grover, H. J., "Effects of Grinding and Other Finishing Processes on the Fatigue Strength of Hardened Steel," a reprint of the ASTM, June 1950.

42. Halvestadt, R. D., "Analysis of Residual Stress in Ground Surfaces of High-Temperature Alloys,Trans. ASME,Paper No. 57-SA-62, August 1957.

43. "Reducing Residual Stresses" from "Engineering Abstracts", Product Engineering, November 1951.

44. Cellers, M. Q., "Control of Residual Stresses Through Welding Sequence", The Iron Age, September 2, 1943.

CHAPTER 18

WEAR RESISTANT MATERIALS

1. METALLURGY

General Description

There are many useful types and compositions of cast iron, and these vary widely in their hardness, strength, and in their plasticity properties. Cast irons can be classified according to the type of furnace used in their manufacture, and according to composition and purity. However, the most common and enlightening basis for their classification and discussion is provided by their distinctive metallographic structures.

The basic differences between steels and cast irons arise principally from a significant difference in their carbon content. All steels contain less than about 1.7%, and so, at sometime in their history, they exist entirely as solid solutions. Cast irons, on the other hand, always contain more carbon than can dissolve in austenite, even at the eutectic temperature. As they complete their freezing, therefore, all cast irons contain a certain proportion of the iron-carbon eutectic in addition to the saturated primary crystals of austenite previously formed. It is the presence of a eutectic mixture that is chiefly responsible for the differences in properties which distinguish cast irons from steels, and is primarily the form which the eutectic mixture assumes that distinguishes the various types of cast iron from one another. Any iron carbon alloy containing more than 1.7% carbon but less than 6.7% carbon satisfies the above definition and is a cast iron. However, the great majority of cast irons fall in the hypoeutectic region of composition; usually between 2.5 and 3.75 carbon.

Within this range of composition, variations in carbon content have an important influence upon the properties of cast iron, but the major effect is not the obvious one of altering the proportion of eutectic present. Far more important is the fact that the carbon content is one of the four principle factors which determine whether carbon not dissolved in the solid iron exists in the finished casting as cementite, graphite, or as a mixture of the two. It also determines the shape and distribution of these carbon or carbon rich particles, whatever their chemical nature may be. The other three of these fundamental factors are alloy and impurity content, cooling rate during and after freezing, and heat treatment subsequent to casting. Together with carbon content, these variables can be adjusted to produce cast irons of any of the following types, which are here classified according to their metallographic structure:

1. White cast irons in which all the carbon not dissolved in the iron itself is present as cementite (Fe_3C-combined carbon)
2. Gray cast irons, in which some, most, or all of the undissolved carbon exists as flakes of graphite (free carbon) contained in a matrix whose microstructure is that of an impure steel.
3. Chilled cast irons, in which a white cast iron layer appears on some or all of the surfaces of a casting whose interior is gray cast iron.
4. Mottled cast irons, in which small areas having the gray iron microstructure, are contained in a white cast iron matrix. (except by accident the mottled iron structure is normally produced only as the transition layer between the white iron surface and the gray iron center of a chilled iron casting.)
5. Malleable cast irons, which contain graphite, as do the ordinary gray irons, but in the form of irregular spheroids (temper carbon) instead of flakes. Malleable cast irons are produced by the prolonged heat treatment of special grades of white cast iron.

6. Nodular cast irons in which, through close foundry control and special alloy additions, the graphite of gray cast iron is caused to appear as spheroids that somewhat resemble the temper carbon of a malleable iron, but are more compact and more regular in shape. In a nodular iron, this graphite structure forms during freezing and does not require subsequent heat treatment of the casting.
7. Alloy cast irons, which may be of any of the above types, but are significantly modified in structure, or in properties, by the addition of alloying elements that are not normally present in irons of the same type.

It is chiefly to a discussion of the properties resulting from these microstructures, and of the principles involved in producing them, that the rest of this paper will be devoted.

Detailed Discussion

White Cast Irons

Solidification of a white cast iron and the solubility and eutectoid changes it undergoes during cooling, are exactly those indicated on the Fe - C diagram. At room temperature, therefore, a white iron exists as a mechanical mixture of ferrite and cementite crystals, plus minor amounts of impurity constituents. When broken, the fresh fracture surfaces of a white iron are clean and bright, like those of steel and are quite different in appearance from the dull gray fracture surfaces of an iron that contains graphite. It is the distinctive appearance of its fracture that has given white cast iron its traditional name.

All white cast irons are hypoeutectic in carbon content and most of them fall toward the lower carbon end of the cast iron composition range, in the vicinity of 3% or less. At room temperature, the typical microstructure of white case iron consists of course, often columnar, dendrites of transformed austenite, (i.e. of pearlite-imbedded in a matrix of transformed ledeburite, i.e., of the eutectic structure of ledeburite and pearlite, in which the pearlite replaces the eutectic austenite,) formed during freezing of the eutectic.

On the basis of this typical microstructure, the characteristic properties of a white cast iron are easy to predict and explain. Cementite is the hardest and strongest phase normally encountered in alloys of iron and carbon, but it is very brittle. Appearing in white iron in large amounts, and as a continous interdendritic network, it makes the cast iron very hard, strong and resistant to abrasive wear, but so brittle and so difficult to machine that its usefulness is distinctly limited. White cast iron is still widely used for grinding mill balls, and liners, screw flights for conveying abrasive solids and pulps, dies and extrusion nozzles, wearing plates, etc. However, where impact, as well as abrasive wear, may be encountered, white cast iron is usually used only as a surface layer on chilled iron casting.

The hardness and wear resistance of a white cast iron can be further increased to some degree by accelerated cooling after casting, or by heat treating the finished casting in order to convert the austenite in its microstructure to martensite, or bainite instead of pearlite. Such a treatment is not often feasible because of the inherent brittleness of white iron and the danger that it will crack during the hardening treatment.

A similar hardening effect can be achieved by including in the composition of the iron, alloying elements of a type and in an amount that will drastically increase its hardenability, so that normal cooling rates in the mold will still exceed the critical cooling rate in the casting. Alloy white irons in which this has been done, are the hardest, and most abrasion wear resistant cast irons available.

Malleable Cast Irons

In the discussion of white cast irons, cementite has been treated as if it were stable phase; actually it is a metastable phase. Under the service conditions to which steels and cast irons are normally exposed, cementite exists indefinitely in its original form. Occasionally, however, in service at elevated temperatures, it is found to decompose spontaneously into iron and graphite, sometimes with serious consequences. Its inherent tendency to decompose in this way is, perhaps, best exemplified by the sequence of events involved in the production of the commercial product known as malleable cast iron.

Malleable cast iron is always produced by heat treatment of white iron casting. The casting is reheated very slowly (to avoid cracking) to a temperature in the neighborhood of $1550^{\circ}F$. Graphitization occurs by the precipitation and growth of graphite nuclei which grow at approximately equal rates in all directions, and commonly appear as rough ragged spheroids usually identified as temper carbon. The austenite formed at the same time simply adds to the austenite crystals already present. Graphitization of cementite is a relatively slow process even at 1550°F., and so the cast iron must be soaked at the malleablizing temperature for many hours, usually twenty-four hours or more. At room temperature, the product of the cycle is a malleable cast iron, distinguished from the original white cast iron, by the presence of graphite rather than massive cementite, and from other types of cast iron by the characteristic shape of the temper carbon particles.

The product of this long cycle of heating, holding, and cooling is completely graphitized iron consisting of temper carbon and ferrite and is called "ferritic malleable". Graphite has almost zero hardness and strength and the mechanical properties of such iron are essentially those of its ferrite, reduced by interruptions in its structure, which amount to graphite-filled holes. Ferritic malleable, therefore, is soft and relatively weak, but because of the plasticity of its ferrite matrix and the fact that the numerous graphite inclusions are rounded and well distributed, it does have some degree of plasticity and toughness. It is not actually malleable in the ordinary sense, and it cannot be forged, rolled, etc. However, it is fairly easy to cast into complex shapes. Its machinability is excellent, and it has plasticity and toughness sufficient to conform to eccentric loads and to withstand a reasonable amount of impact in service. Therefore, ferritic malleable has been widely used in pipe fittings, valves, farm equipment, automotive parts, and other machinery applications, in which service is not severe and high wear resistance is unnecessary, but where ordinary cast irons fail because of excessive brittleness.

The ferritic type is, by far, the most common of the malleable cast irons, but several other modifications are occasionally encountered. An increase in the cooling rate over that required to produce ferritic malleable results in a pearlitic malleable cast iron. Here all the proeutectoid cementite which forms at a relatively high temperature is graphitized as it forms, but the eutectoid cementite is not. The resulting microstructure, at room temperature, is temper carbon in a pearlite matrix. In this condition the iron is harder and stronger than if it were a ferritic malleable, slightly more brittle and somewhat more difficult to machine (although still much easier to machine than most steels). Pearlitic malleable has been quite widely used in ordinance parts, in particular, because it combines good machinability with considerably greater toughness than can be expected with ordinary gray irons. It is sometimes called "partially malleablized cast iron", in recognition of the fact that a fully malleabilized cast iron contains only ferrite and temper carbon.

Further acceleration of the cooling after the malleabilizing anneal retains to room temperature some, or all, of the the proeutectoid cementite, as well as that present as pearlite, and would ultimately produce bainite or martinsite instead of pearlite. With any of these structures, the matrix of the iron is so brittle that its toughness, which is the major advantage of malleable iron, is lost. Consequently, cooling rates higher than those needed to retain a completely pearlitic matrix are never purposely used in heat treating the malleable cast irons.

Frictional wear resistance is a composite of several material requirements: (1) High controllable hardness, (2) Innocuous wear debris, (3) Structural embeddibility and (4) Lubricant retention. If we compare the structure of pearlitic malleable with the above ideal properties, there is an impressive match. The hardness of matrix can be adjusted as high as 67 Rockwell "C". This hard, high carbon matrix yields a brittle wear debris, likely to be a harmless micron size powder rather than a damaging ductile particle. In cases of extreme metal-to-metal engagement, a gall streak would have difficulty in propagating to a great length, because of the thousands of sites for debris imbedment. The temper carbon nodules, themselves being "microlubricants" also serve for lubricant retention and storage.

Usage of pearlitic malleable has shown that it is free from interconnected porosity that would otherwise interfere with the maintenance of pressure necessary for supporting hydrodynamic lubrication films. At the same time its microstructural constitution is such that the functional surface acquires a texture that is good for the lubricant retention needed for successful boundary lubrication. This optimum combination of properties usually is obtained only by expensive power metal bearing materials, but in pearlitic malleable these ideal wear resisting features are intrinsically in the ordinary product.

Gray Cast Iron

The gray cast irons are the common cast irons of engineering and industry. In their manufacture the natural tendency of cementite to decompose into graphite plus austenite or ferrite is encouraged by adjusting alloy composition and cooling rates so that a controlled proportion of the cementite formed during and after freezing of the iron, will graphitize as it forms. The result at room temperature is a microstructure in which graphite is present as a mechanical mixture with ferrite and usually some undecomposed cementite.

Cast irons differ from malleable cast iron in two principal ways. First, the graphite in a gray iron is formed during freezing and subsequent cooling of the cast iron so that no later heat treatment is needed; second, the graphite in a gray iron exists as thin curved plates or flakes, instead of relatively compact particles. Because these flakes interrupt its microstructure so seriously, a gray iron is relatively weak in structure and it is almost completely brittle. To compensate for this, gray iron is the cheapest of the ferrous alloys, and the easiest to cast. Its machinability is excellent, its dampening capacity very high, its notch sensitivity very low, and its compressive strength and bearing properties are good. This combination of properties combined with its low cost makes gray cast iron one of the most useful and widely used alloys of iron.

Certain factors make gray irons valuable for wear resistance, namely:

1. The graphite in the iron tends to add lubrication. Galling is thus minimized.
2. Where the danger of surface indentation does not exist, gray irons are the economical to use.
3. Gray irons are machinable. Thus, parts requiring close fits and tolerances can be produced economically.

The extensive use of gray irons for moving machinery parts is evidence enough that they have excellent wear resistant qualities. However, many conditions, outside the type of iron used, such as type and quality of lubrication, corrosion conditions, operating temperature, surface condition, pressures involved and speed of operation influence the choice of gray iron for any given service. On the other hand, the wide choice of alloyed cast irons and their resulting microstructures and strengths make it possible for more than one type of gray iron to be satisfactory for a given job. Therefore, there exist considerable differences of opinion as to the proper iron to use for wear resistance. Factors such as, conditions of molding, rate of cooling, shape, size and design of casting, insofar as they effect the microstructure and strength, definitely influence wear resistance. An iron possessing a pearlite matrix will resist wear better than one which contains some free ferrite. Free ferrite, due to its inherent softness will score and gall. Cementite, in quantities not sufficient to produce machining difficulties is beneficial to wear, and random graphite flake distribution is best.

The influence of hardness on wear resistance is very controversial, and it is believed that the influence of microstructure and alloying have far more influence on wear. In special cases the wear resistance of a gray iron may be increased by heat treatment, such as, quenching and tempering, flame hardening, induction hardening, and nitriding. Castings subject to corrosion along with wear are usually made of high chromium, high nickel and high copper material.

Chilled Cast Irons

Because of their brittleness, completely white cast irons are limited to applications that do not involve shock, loading or misalignment in service. It is possible, however, by chill-casting techniques to produce irons having the high surface hardness and abrasion resistance, of a white iron case, backed by a softer, less brittle, gray iron core. This possibility greatly extends the field of usefulness of cast irons in general.

Although all cast irons have a tendency to graphitize, this tendency can be suppressed, either by maintaining a low carbon silicon content in the iron, or by accelerating cooling during freezing. In the production of chilled cast irons, conditions are adjusted so that within the same section a surface layer freezes as white iron over a core that freezes as gray iron. The entire section must be cast at once and from the same melt of iron, therefore, this case-core structure cannot be produced by local adjustments of the composition within the casting. It can be made only by precise adjustment of the overall alloy composition in accordance with local cooling rates, which may also be adjusted. The technique used may take either of two forms depending on whether it is intended that all or part of the surface of the casting shall be white.

The surfaces of a sand casting are never smooth and precise dimensional tolerance cannot be maintained with iron castings made in sand molds. Except for relatively rough applications, e.g., wearing plates and griding mill balls-machining and finishing of some surfaces of the casting are usually necessary. White iron is so hard that it cannot be penetrated by ordinary cutting tools or machined by ordinary methods; in general it must be ground to size. The removal of metal by grinding is expensive. Therefore, it is preferable to produce chilled iron castings, in which only those surfaces are white iron that must actually resist wear in service, and the rest of the casting is gray iron, in the interest of machinability. Selectivity in this respect, can be exercised by the use of either chills or insulators built into the mold in which the casting is to be made.

A major deficiency in chilled cast irons results from the fact that the white iron surface layer is separated from the gray iron core by a transition layer of mottled iron. Typically, the mottled layer contains graphite rosettes in a white iron matrix and it is always relatively weak and very brittle. Its brittleness can be reduced a little by spheroidizing the cementite in this layer. To induce spheroidization, high quality chilled iron castings are sometimes cooled very slowly after casting to some temperature safely below that of the eutectoid transformation. At best, however, the presence of the mottled layer is undesirable and every effort should be made to minimize its thickness.

Chilled cast irons, with very high properties and precisely controlled chill depths are now produced in large daily tonnages, and are used in many critical applications. One of the most important of these uses, both economically and socially is in chilled iron railway car wheels, where the wear resistances of a white iron tread is combined with the high damping capacity, low notch sensitivity and excellent machinability of a gray iron plate and hub. Crushing rolls, stamp shoes and dies, plow shares, and many other heavy duty machinery parts are also commonly cast as chilled iron.

Nodular Cast Irons

The advantages that appear when the free carbon in a cast iron exists as spheroids instead of as plates, have already been indicated by comparisons of the ordinary gray cast irons with the malleable irons. Compact spheroids interrupt the microstructure of the iron much less seriously than flakes of equal mass, and so an iron in which graphite appears as small well distributed spheroids is markedly superior in strength and toughness to an otherwise similar iron which contains graphite as flakes.

Until six or seven years ago the only known means of producing spheroidal graphite in a cast iron was to cast the iron as white iron and then malleablize it by a long and expensive heat treatment. Within the past several years, however. techniques have been developed that make it possible to produce gray iron castings which in the as-cast condition and without any heat treatment contain all of their graphite as small, well distributed spheroids. Because they often show tensile elongations as high as 20% in advance of fracture, these are often called ductile cast irons. They have also been identified as spherulitic and spheroidal irons, but their most common title is nodular cast irons.

Nodular irons can now be made from almost any cast iron composition that would normally solidify gray, by adding to it a small precisely controlled addition of any of several special alloying elements. The nodule forming addition is usually made in the ladle shortly before the molten iron is to be poured into the mold, and is accompanied or immediately followed by the addition of some graphitizing agent. All elements known to have a nodularizing effect on gray iron also have a strong affinity for sulphur and seem to scavenge all the sulphur from the molten bath prior to producing nodular graphite structures. The additions are expensive and in removing large amounts of sulphur they are being used inefficiently. Therefore, the iron composition to which the nodulizing addition is made, must be one in which the sulphur content is low and consistent. Unfortunately, the close foundry control required to produce such a base iron makes it distinctly more expensive than an ordinary gray iron.

In addition to the expense and inconvenience of the alloy additions, and of composition control, nodular cast irons suffer from, at least, two other major disadvantages. First, the nodulizing and graphitizing additions must be made in the ladle not more than a few minutes before the iron is to be cast and the molten iron must be stirred thoroughly to dissolve and distribute them. As a result, either of the stirring, or of a gas evolving reaction, within the melt, a relatively large amount of gas is dissolved by or evolved from the iron and entrained in it. This gas is liberated from the iron while it freezes in the mold, is trapped in the solidifying casting and appears as blow holes in the finished product. Blowholes are a serious defect in any cast metal and they are especially difficult to avoid in casting nodular iron. A second serious disadvantage of nodular iron is that its contraction during freezing is considerably greater than that of ordinary gray iron. This increases both the difficulty of mold design and the expense of the casting operation, since elaborate precautions are required to avoid shrinkage cavities (pipes) in the finished casting.

The amount of ferrite in the as-cast matrix depends on the composition and rate of cooling, and may approach 100% which gives maximum ductility, toughness, and machinability. If the cooling rate in the mold is too rapid the matrix can be completely ferritized by heating to about 1300^{O}F. and holding for one to five hours, followed by cooling at any convenient rate. This treatment will convert the entire matrix to ferrite producing the softest type of nodular iron capable of elongations up to 25%. If higher hardness and strength are required, nodular iron may be heat-treated to produce a pearlitic or martensitic matrix.

The mechanical properties and wear resistance of any of the forms of nodualr iron are in general similar to those previously cited for the malleable irons.

Alloy Cast Irons

The term "alloy cast iron", is used to describe those irons to which nickel, chromium, copper, vanadium, molybdenum, etc. have been purposely added. Additions of these elements alter both the microstructure and mechanical properties of the iron. The following discussion will deal mainly with the changes in mechanical properties and only briefly touch upon the metallurgy. It is not to be assumed that the metallurgy is of lesser importance because for one to understand why these changes occur, a thorough knowledge of the metallurgical aspects is necessary. Microstructure more than any other singular factor determines the mechanical properties of steel and non-ferrous materials as well as cast iron.

Although elements such as carbon, silicon, phosphorous, sulfur, and manganese are present in varying amounts in all cast irons and are not alloys in the same sense as nickel, chromium etc., a rather brief discussion of their effects will be included in this section.

Influence of Silicon

Silicon is a graphitizing element, and tends to break down the cementitic carbon. It is the primary element in determining the relative proportions of combined and graphitic carbons. Thus with a given carbon content the foundryman can control the properties of his iron largely by silicon adjustment. Silicon dissolves in the iron forming a solid solution which increases in hardness as silicon content increases (up to 4%). This dissolved silicon in the matrix also adds to its strength.

Larger amounts of silicon (13 to 17%) are added to produce acid and corrosion resistant alloys. The major disadvantage of these alloys is their extreme brittleness. However, alloys containg 7 to 8% silicon combined with nickel and chromium are tough and heat resistant.

Influence of Manganese

The manganese content of most commercial cast irons is between 0.4 and 1.0%. Variations within this range have little effect on the properties of the iron if the silicon content is ample.

In low carbon, low silicon irons manganese, above 10%, causes the metal to become austenitic. As the manganese content of the cast iron is increased the matrix is successively pearlitic,sorbitic, martensitic, and austenitic. Manganese in excess of 1.25% increases strength, hardness, and resistance to wear, but decreases machinability. In amounts less than 1% manganese has little effect on the mechanical properties.

Influence of Sulphur

Sulphur contents as high as 0.12% are not noticeably harmful providing sufficient manganese is present. The resulting inclusions, while not beneficial, occur in such small percentages by volume, that their effect on the physical properties is completely hidden by the effect of the major inclusions in the irons. If sufficient manganese is not present, sulphur stabilizes carbides markedly, thus inhibiting graphitization.

Influence of Phosphorus

Phosphorus content varies between 0.15 and 0.9% in all cast irons. The structural component steadite is found in the irons in which the phosphorus content is on the higher side of the given limits. Mechanical properties are not noticeably effected if the phorphorus content remains below 0.3%. However, shock resistance decreases slowly while hardness and stiffness rise with increasing phosphorus.

Influence of Chromium

Chromium is a powerful carbide stabilizer and also enters into solution in the iron matrix. The complex iron chromium carbides are stable even under high temperature conditions. Small amounts of chromium cause the ferritic structures to be replaced by the stronger pearlitic types, the graphite flake size is made more uniform, and the grain size is made finer. Because its effect on the structure, small amounts of chromium alone increase the strength, hardness, depth of chill, resistance to heat, wear, and corrosion, and decrease machinability (.015 to .090%). If maximum resistance to oxidation and growth is desired, the amount of chromium may be as high as 2%. In the higher chromium irons additional graphitizing alloys must be added.

Influence of Copper

Copper functions primarily as a graphite forming element, but this ability is about 1/10th that of silicon. Copper in excess of its limits of solid solubility is dispersed as microscopic or submicroscopic particles in the iron. It tends to make the structure more dense, though the microstructure is not materially changed. The mechanical properties, particularly the tensile strength and hardness increase almost linearly with copper content; the improvement being greatest in the low carbon irons. Copper increases wear resistance, anti-friction properties, and shock-resistance where there is sliding friction, as in break drums. It also increases the resistance to acid and atmospheric corrosion.

Influence of Molybdenum

Molybdenum decreases the rate of decompositon of austenite, thus producing a fine to an acicular pearlite depending on the amount present. It is because of this (solid solution), that molybdenum exerts its beneficial effect. Like other alloying elements its greatest use is in combination with other elements such as nickel, copper, chromium, etc. Molybdenum improves fatigue characteristics and elevated temperature strengths.

Influence of Nickel

Nickel is a graphitizer, and small additions (under 1%) are effective in refining the grain and size of graphite flakes. It reduces chill and increases machinability. It may or may not increase strength depending upon whether or not adjustments are made to compensate for its graphitizing power. If this composition is made, nickel increases the strength, hardness, and resiliency. Although nickel is beneficial when used alone, it is more often used with other alloying elements.

Influence of Titanium

Titanium is called a reducing agent because of its affinity for oxygen and nitrogen. It also decreases the size of the graphite flakes and reduces the amount of combined carbon. Titanium improves both tensile and traverse strength, reduces chill, and may increase corrosion resistance, wear resistance, and machinability, depending on the over-all alloy composition.

Influence of Vanadium

Vanadium is the most powerful carbide forming element and serves primarily to stabilize cementite and restrain graphitization. Secondarily, it will react with oxygen and nitrogen or their compounds if such action has not already been effected by other means. The most important contribution of Vanadium is its effect upon the uniformity of graphite flake size and distribution in alloy cast irons. Vanadium increases tensile and traverse strength, hardness, wear resistance, and depth of chill. It decreases linear expansion and if added in large amounts decreases machinability.

2. PHYSICAL PROPERTIES

Cast iron is widely used in engineering and allied industries because of the ease with which it may be cast, its moderate cost, and its wide range of useful physical properties. It is one of the best known and least understood of the common materials of construction. One reason for this is that "cast iron" like "steel," is a general term used to designate not one,but a class of materials.

Graphite Types

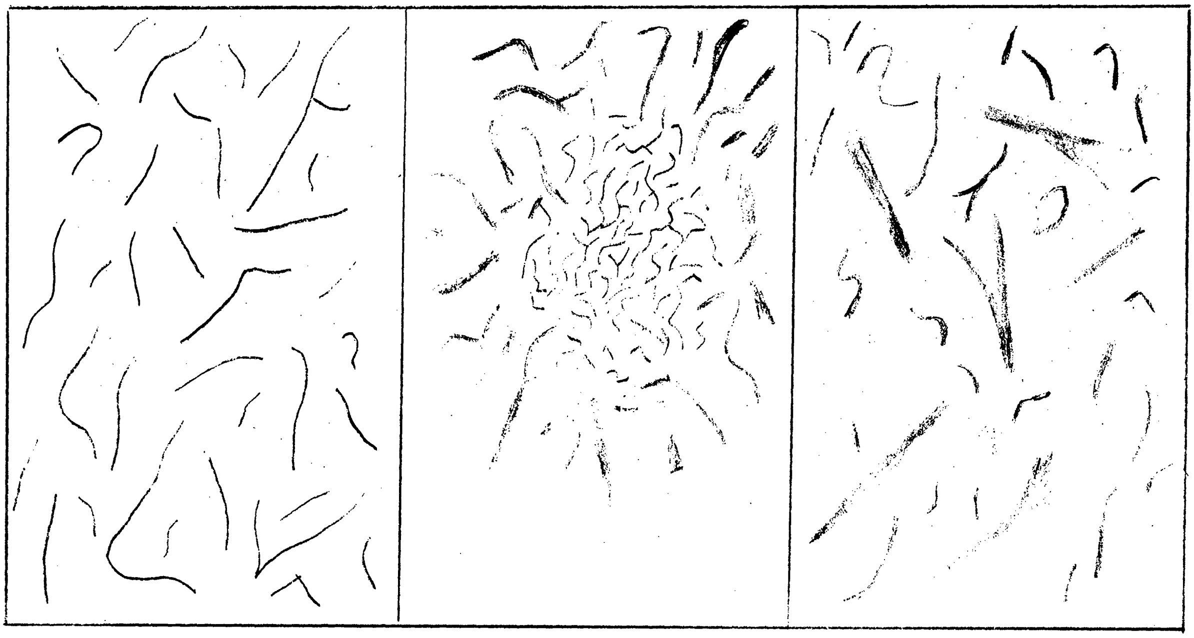
Figure 1 shows the five types of graphite classifications of cast iron.

Tensile Strength

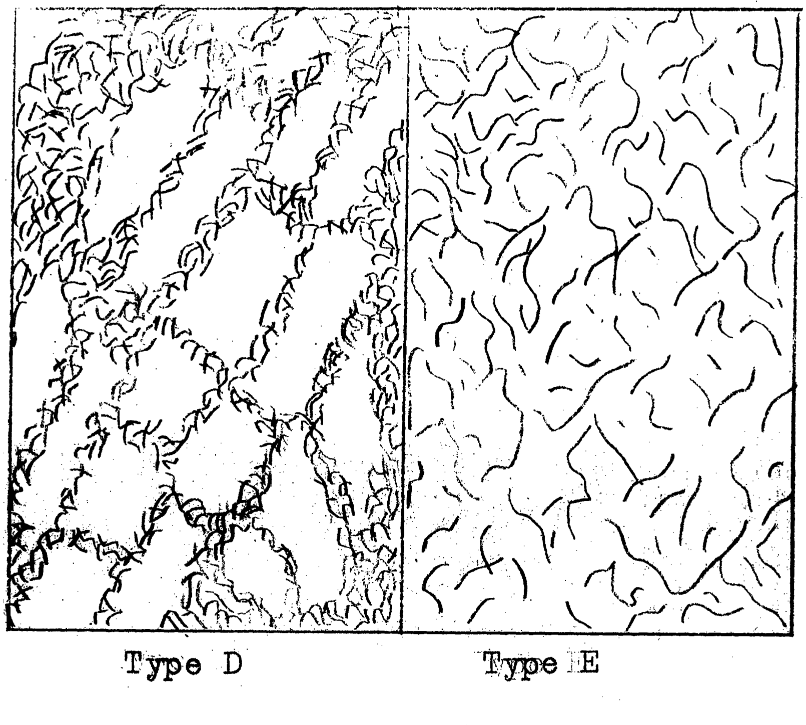
There are a great many different specifications for cast iron, but the A.S.T.M. specification A48 covering the engineering grades of cast iron is by far the most widely accepted. This specification recognizes seven classes of gray iron having tensile strengths from 20,000 to 60,000 psi.* Most gray iron is produced in the classes 20 to 50 inclusive. In most cases, the latter class requires reasonably large additions of alloys.

---

* Figures 3,5 and 8 are graphs of the tensile strengths as a function of class, for gray as well as malleable and nodular cast irons.



Type A Type B Type C



Type D Type E

Figure 1

Graphite Types

However, Class 60 iron usually requires a considerable amount of alloy plus a very close control of the carbon and silicon content.

It has been generally acknowledged that most of the mechanical properties of gray iron can best be co-ordinated with its tensile strength.

It is a well known fact that the tensile strength as well as a number of other physical properties of gray iron depends fundamentally on its chemical composition, its cooling rate from molten metal to room temperature, and its processing in the ladle.

The tensile strength as well as other related properties of all ferrous materials will vary with the section thickness or size. Gray irons are particularly sensitive in this respect. Figure 10 shows a typical variation in tensile strength and Brinell hardness with section thickness. It will be noted that the thinner sections are more affected by this condition than the heavier sections.

Compressive Strength

The compressive strength of cast iron is one of its most valuable commercial properties but references are scarce as compared to the number of tensile tests reported. Figures 2 and 5 give the compressive strengths for malleable, gray and nodular irons.

Perhaps the most important point to note in the figures is that the compressive strength is about four times the tensile strength.

	Gray	Malleable	Nodular
Elastic Modulus	1.6 - 19.5 x 10^6	22 - 27 x 10^6	25 - 28 x 10^6
Shear Modulus	5.4 - 7.11 x 10^6	12.5 - 18 x 10^6	14.5 - 20 x 10^6
Tensile Strength	20 - 60,000 psi	50 - 110,000 psi	60 - 120,000 psi
Yield Strength	18000 x psi	35 - 90,000 psi	45 - 90,000 psi
Shear Strength	32 - 66,000 psi	48 - 93,000 psi	50 - 95,000 psi
Impact Strength	6 - 1 ft. lbs.	10 - 3 ft. lbs.	12 - 5 ft. lbs
Compressive Strength	88 - 180,000 psi	190 -512,000 psi	190 - 525,000 psi
Endurance Limit	18 - 32,000 psi	25 - 39,000 psi	29 - 48,000 psi
Notch Sensitivity-K	1.0 - 1.25	1.2 - 1.45	1.2 - 1.45
Elongation - %	.5 - 0	20 - 2	25 - 5
Hardness	130 - 280 BHN	145 - 308 BHN	150 - 360 BHN
Machinability	Excellent	Excellent-Good	Excellent-Good
Damping Cap.	Excellent	Good	Good

Figure 2

Physical Properties of Cast Irons (Non Alloy)

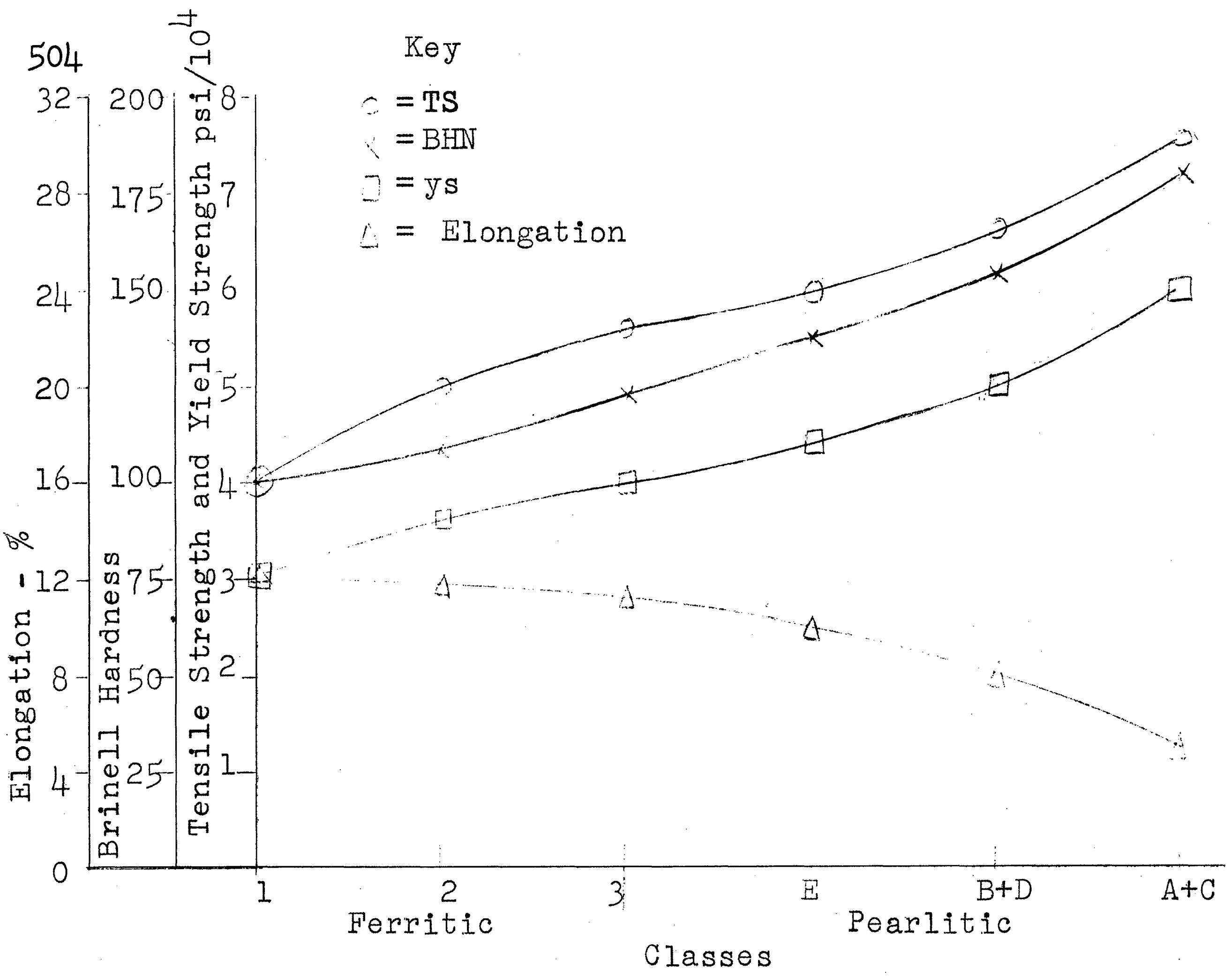


Figure 3

Physical Properties of Malleable Irons

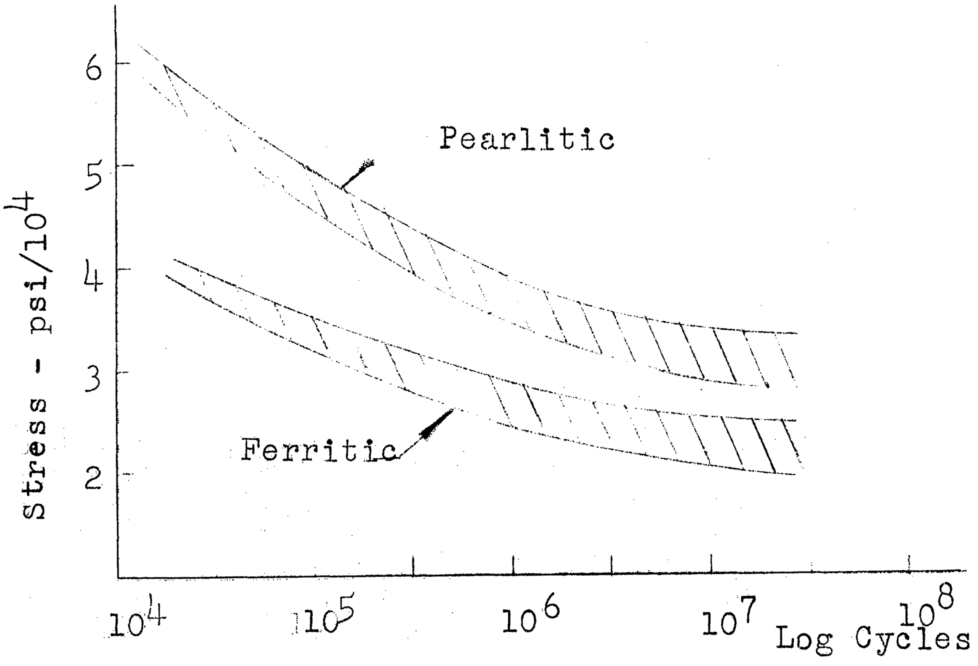


Figure 4

Endurance Limit Curves for Malleable Irons

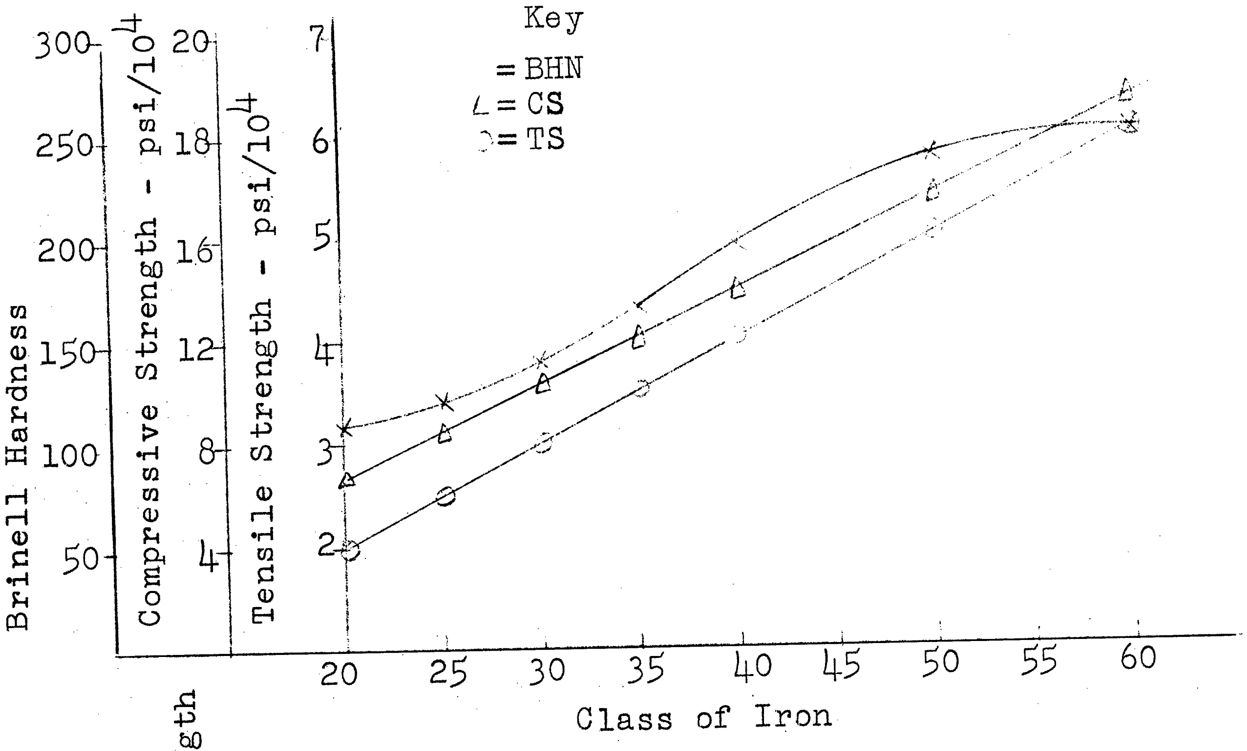


Figure 5

Physical Properties of Gray Cast Irons

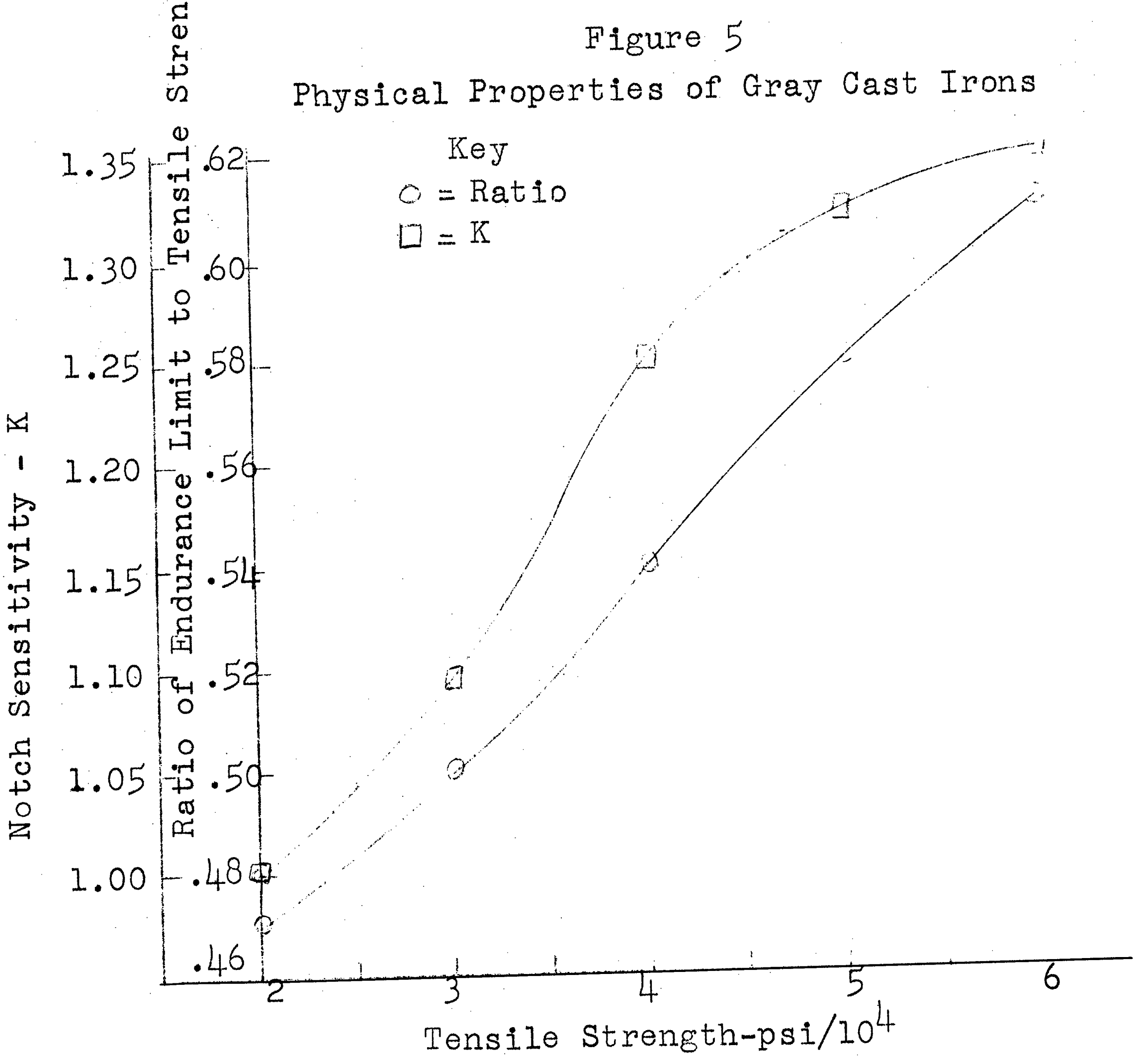


Figure 6

Endurance Limit and Notch Sensitivity of Gray Cast Irons

Elasticity, Plasticity, and Stiffness

Gray iron does not follow Hooke's law exactly. The effective modulus of elasticity at 25% of the ultimate strength ranges from 12 x 10^6 psi for the weaker irons to 18 x 10^6 psi for the stronger types of gray irons. Actually, there is no well defined elastic limit. For certain special and alloyed cast irons, the modulus of elasticity approaches and in some instances equals 30 x 10^6 psi. It should be noted however that this would definitely not be a grade used very often in engineering practice because of its high cost.

In Figure 11 it can be noticed that the modulus of elasticity increases rather proportionally until the higher tensile strengths are reached. The reason that the modulus of elasticity drops off at the higher tensile strengths is that its extreme brittleness at these higher strengths make the modulus rather difficult to determine.

Shear Strength

The shear strength of cast iron may be measured by either of two methods: 1). direct shear as by cutting off the end of a bar, and 2). torsion or twisting.

Direct Shear Strength

In general, the shear strength will vary from 1 to 1.6 times the tensile strength, with the lower tensile strengths having the higher ratios of shear to tensile strength. The results of many investigators who tested many different cast

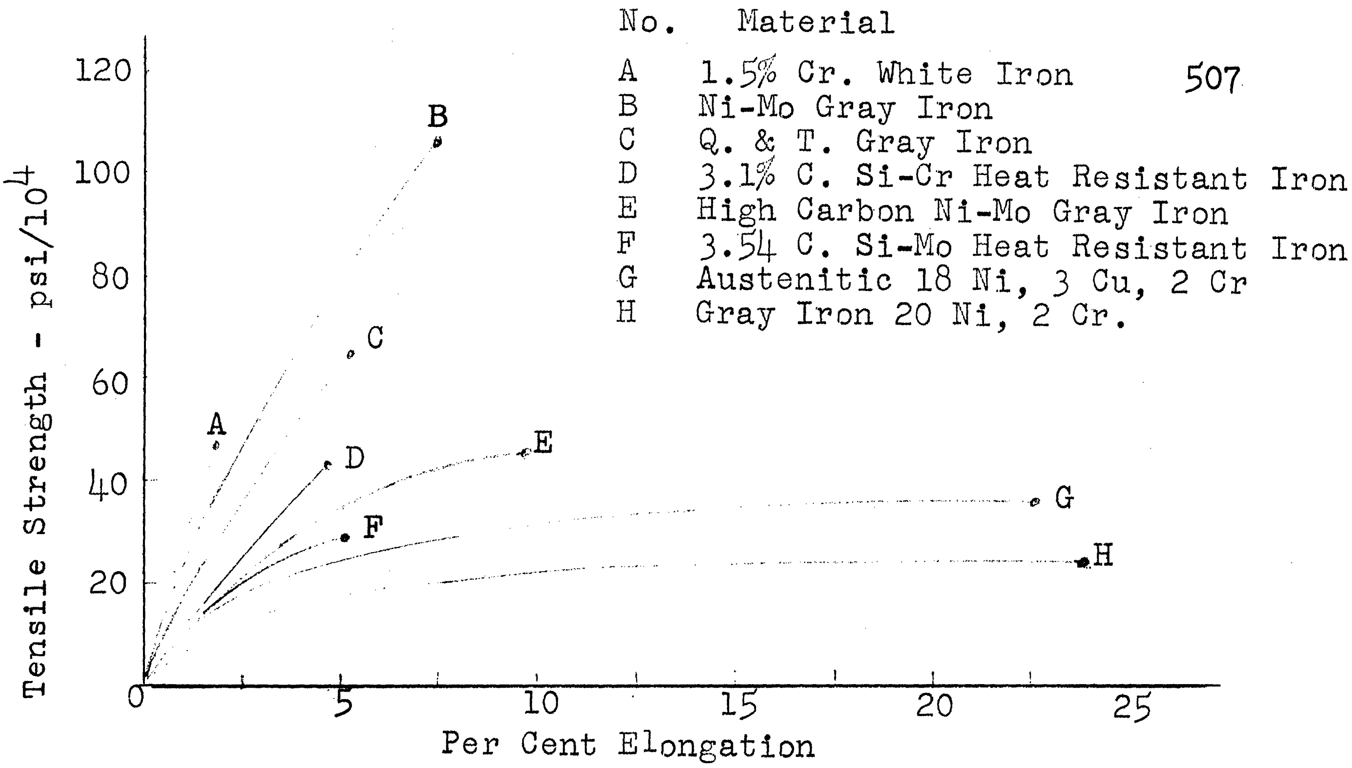


Figure 7
Properties of High Strength Grey Cast Irons

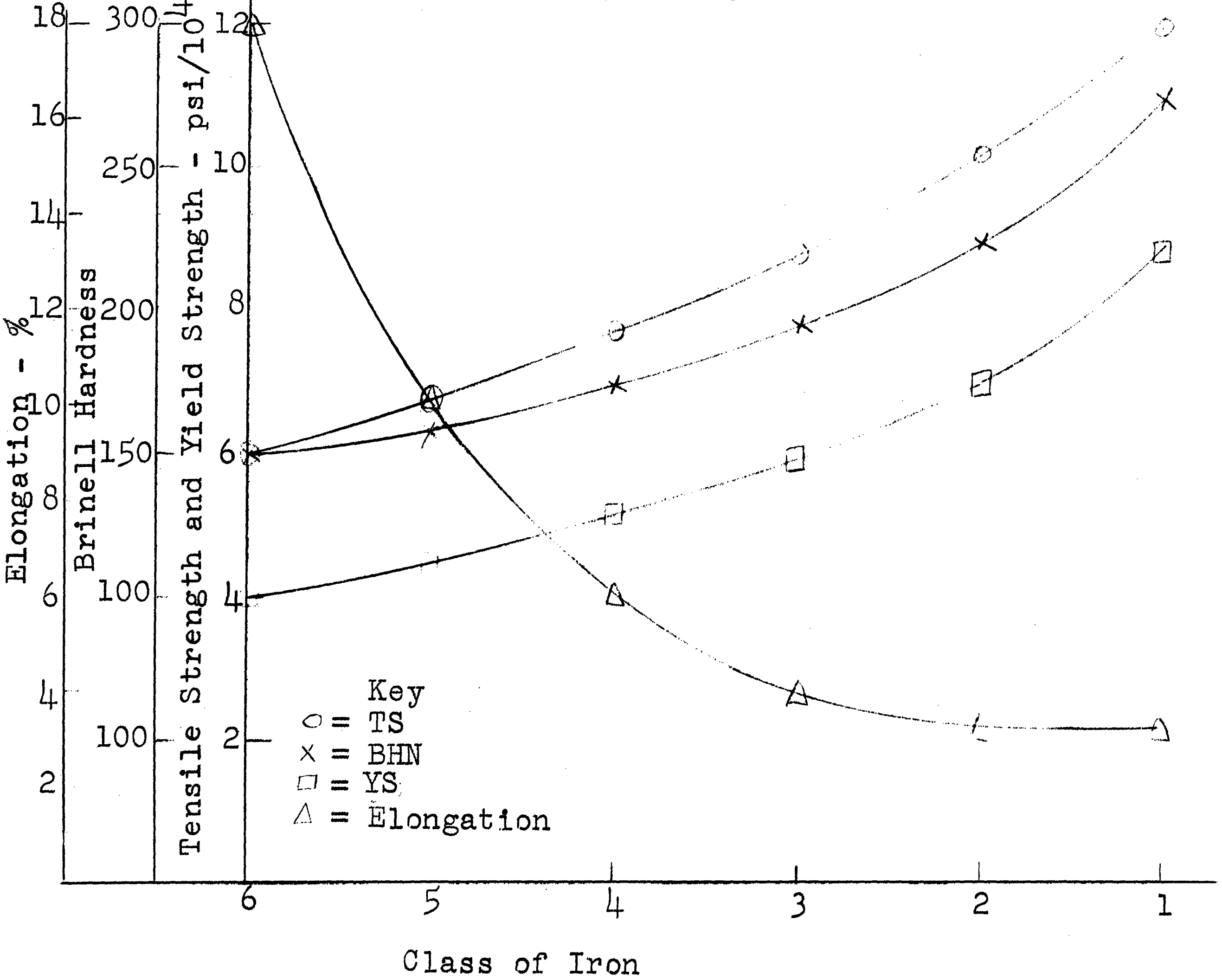


Figure 8
Physical Properties of Nodular Iron

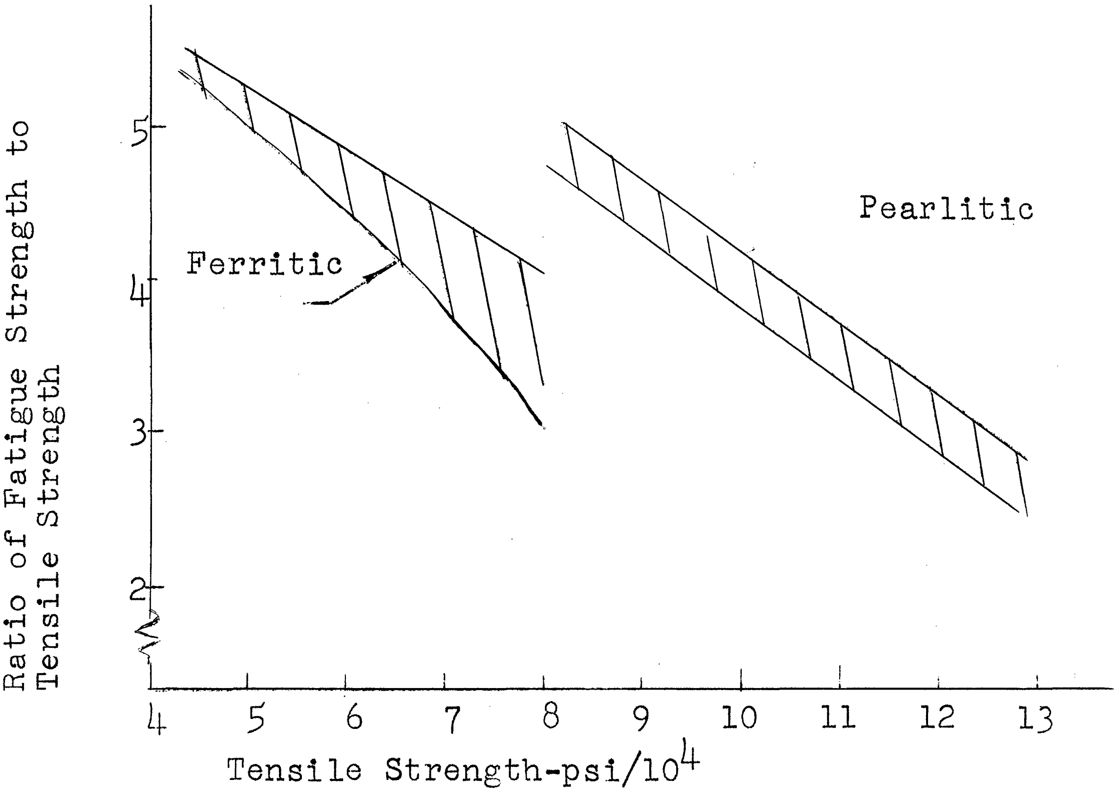


Figure 9
Fatigue Strength of Nodular Irons

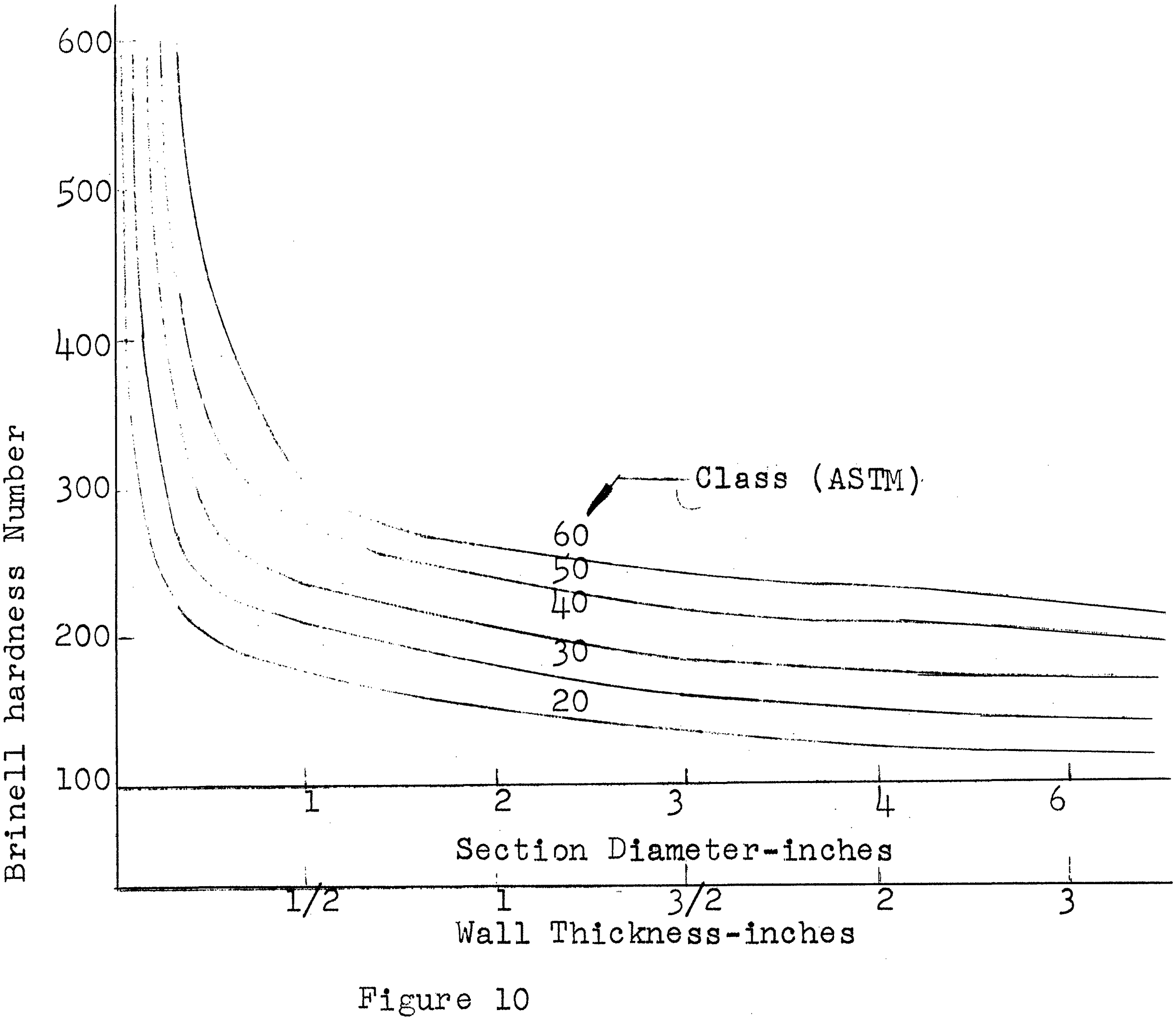


Figure 10
Variation of Tensile Strength and Brinell hardness with section thickness for Gray Cast Iron in sand molds.

irons are shown in Figure 12. It should be noted that the trend in each of the curves is the same, namely that the ratio between shear and tensile strengths decreases as the strength of the metal increases. Also notice that the upper part of the band width refers to higher strength cast irons and the lower part refers to lower strength cast irons.

Torsional Shear Strength

Neither the torsional strength nor the modulus of elasticity in torsion have been investigated extensively but enough has been done to indicate that there is a linear relationship between torsional and tensile strength. It appears that the ratio of torsional strength to tensile strength, for solid specimens in torsion, will vary from 1.20 to 1.40. The fractures of the torsional specimens were in all cases in the form of a helix, the break being in tension, as would be expected since the strength in tension is far less than in compression.

Figures 13 and 14 correlate the torsion strength and the modulus in torsion to tensile strength.

Fatigue and Endurance Limit

Investigations have shown that the endurance limit (fatigue strength) of cast iron apparently varies linearly with the tensile strength. The ratio of the endurance limit for completely reversed cycles of bending stress to the static tensile strength, usually varies from about 0.40 to 0.55. For the more common gray irons this means the endurance limits would vary

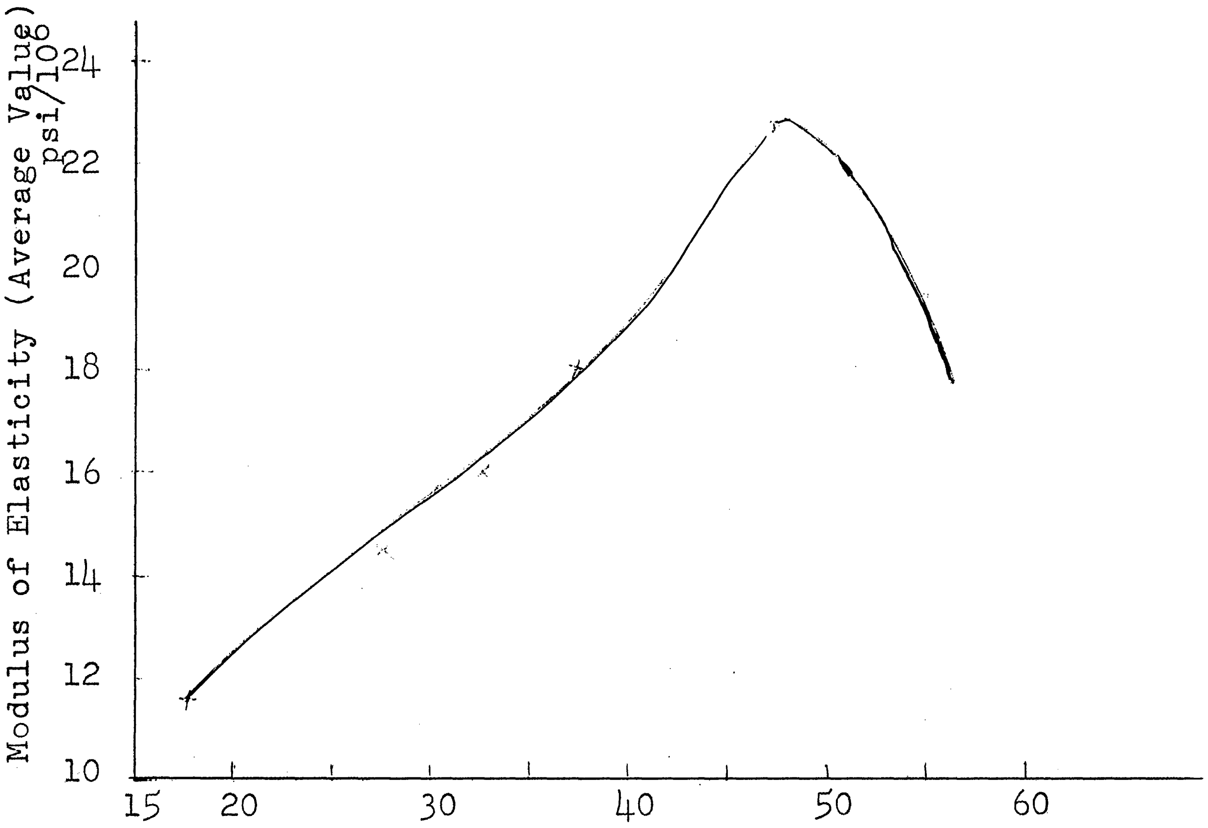


Figure 11

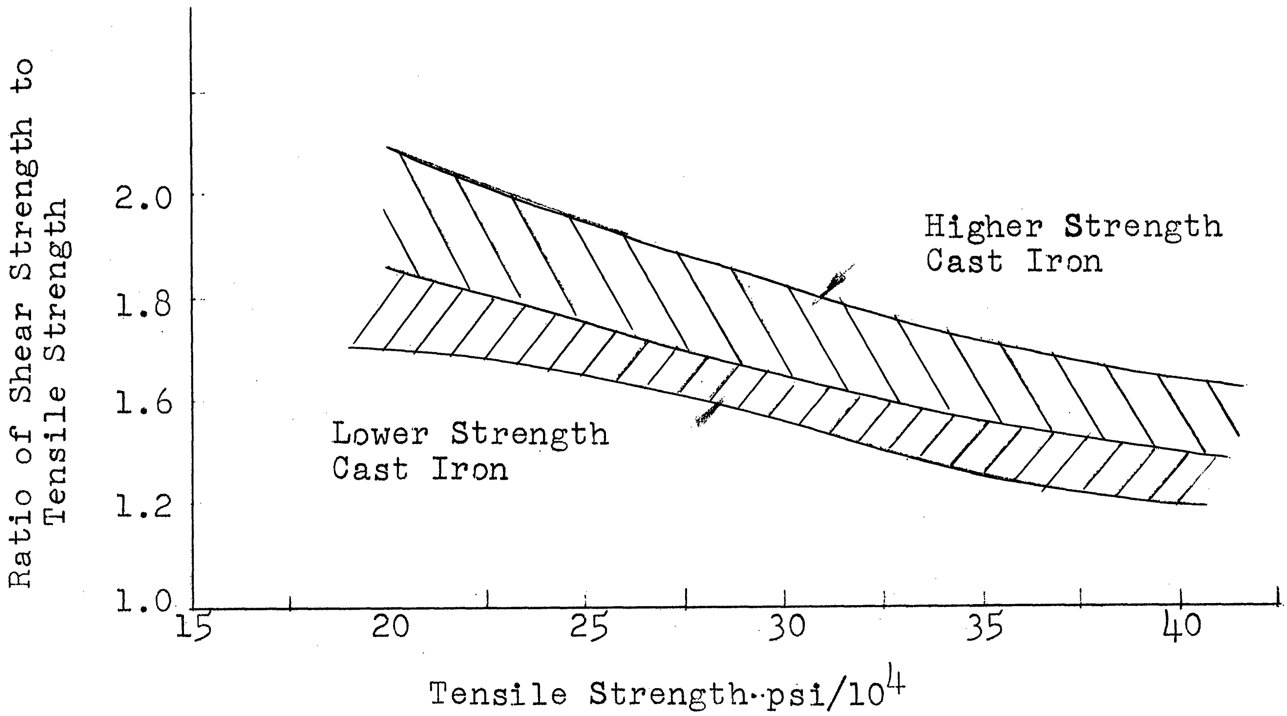


Figure 12

Relation between shear and tensile strengths that exists in cast irons of various tensile strengths (Determined by several investigators.)

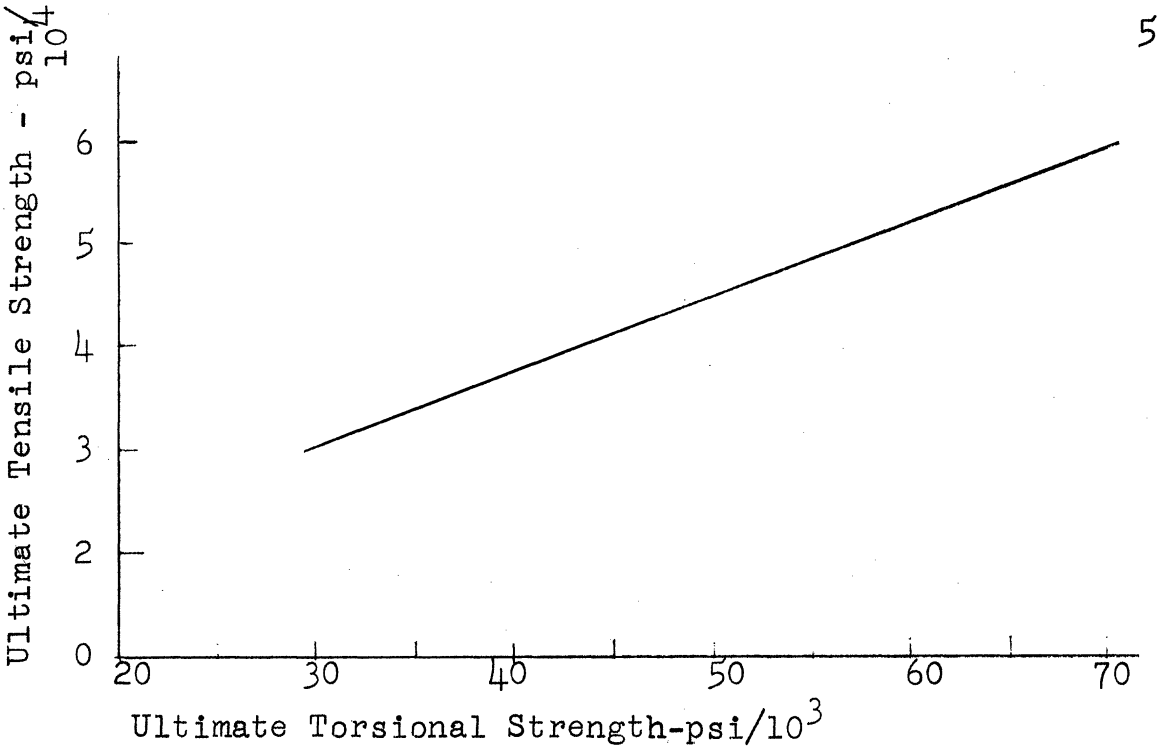


Figure 13
Relation Between Ultimate Torsional Strength and Ultimate Tensile Strength

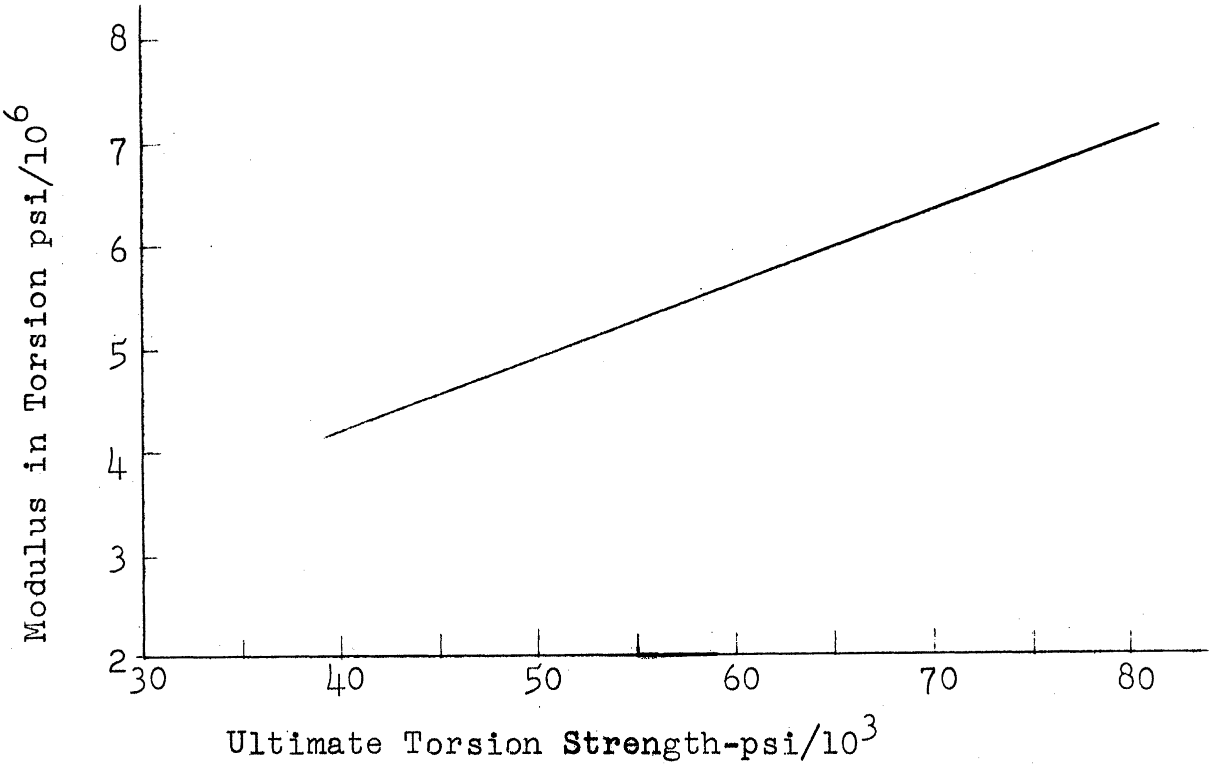


Figure 14
Torsional Modulus vs. Ultimate Torsional Strength

from about 9,000 to 25,000 psi. Special high-strength irons may have values of endurance limit that are somewhat higher. Figures 2,4,6 and 9 give endurance data for malleable, gray and nodular cast irons.

First notice that the line representing the variable stress is not on the conservative side as it was in the case of steel. The second important point which needs to be clarified concerns the compressive side of the diagram. Actually very little work has been done with the compressive side of the diagram but that which has been done indicates that the variable stress limit line on the compressive side holds true for a considerable distance into the compressive range. Certainly this line holds true far enough into the compressive range to be used for all ordinary engineering applications such as for the rods in an internal combustion engine. Notice also that since there is a 4 to 1 ratio between compressive strength and tensile strength that there will be a scale difference between the endurance limit in tension and in compression.

Effect of Notches and Grooves on Endurance Limit

Figures 2 and 6 give the notch sensitivity factor K for the three cast irons.

The important thing to notice here is that for low strength cast irons, the damaging effects of grooves and notches are very much less than that found for most other metals. An explanation for this difference is that cast iron inherently has so many discontinuities due to the presence of graphite flakes, that the addition of a mechanical notch can have little additional effect on the strength of the material.

A specimen of cast iron with a radial hole, which theoretically should have reduced the endurance limit 67%, actually made a reduction of only 13%. Specimens of another iron were tested with two different types of filleted grooves which were expected theoretically to reduce the endurance limit 74 and 67%, actually reduced the endurance limit only 0 and 8% respectively.

Hardness

The hardness of cast iron varies from 100 BHN for soft, easily machined gray iron, to over 600 BHN for chilled and white irons. Because of the fact that gray iron has a considerable amount of flake graphite in it, hardness determinations can be misleading. Therefore, for most gray iron testing up to approximately 500 BHN, the Brinell Hardness determination is recommended. Higher hardness determinations should be made with Rockwell Hardness Testors.

Figures 3,5 and 8 give the hardness for the three materials previously mentioned.

Many different investigators have come up with different relationships between the Brinell Hardness No. and tensile strength, but the easiest to work with a very good relationship is as follows: (for ASTM classifications of cast iron)

T. S. = 300 x BHN------------for classes 20-30
T. S. = 400 x BHN------------for classes 30-50
T. S. = 500 x BHN------------for class 50

Figure 16 shows the relationship between tensile strength and Brinell hardness number.

Effect of Overstressing and Understressing

It has been found that overstressing usually caused some damage, or reduction in endurance limit. The amount of damage is a function of the amount of overstress and the cycle ratio. The cycle ratio is the ratio of number of cycles of overstress to the number of cycles of overstress required to produce failure. One experiment found that an overstress of 30% for a cycle ratio of 80% reduced the endurance limit by 18%. This example for cast iron was found to be quite analogous to certain carbon steels.

Understressing on the other hand was found to have a remarkable beneficial effect on the endurance limit. Understressing is, of course, stressing below the endurance limit. The amount of benefit, or increase in the endurance limit, is a function of the magnitude of the stress and the number of cycles. Maximum strengthening of unnotched specimens was obtained for stresses just below the original endurance limit, and for a number of cycles equal to about 15×10^6 cycles. For an iron having a static tensile strength of 30,000 psi, the maximum increase in endurance limit was found to be 31%. The increase in endurance limit decreases as the magnitude of the understress decreases and as the number of cycles of understress decreases from about 15×10^6 cycles.

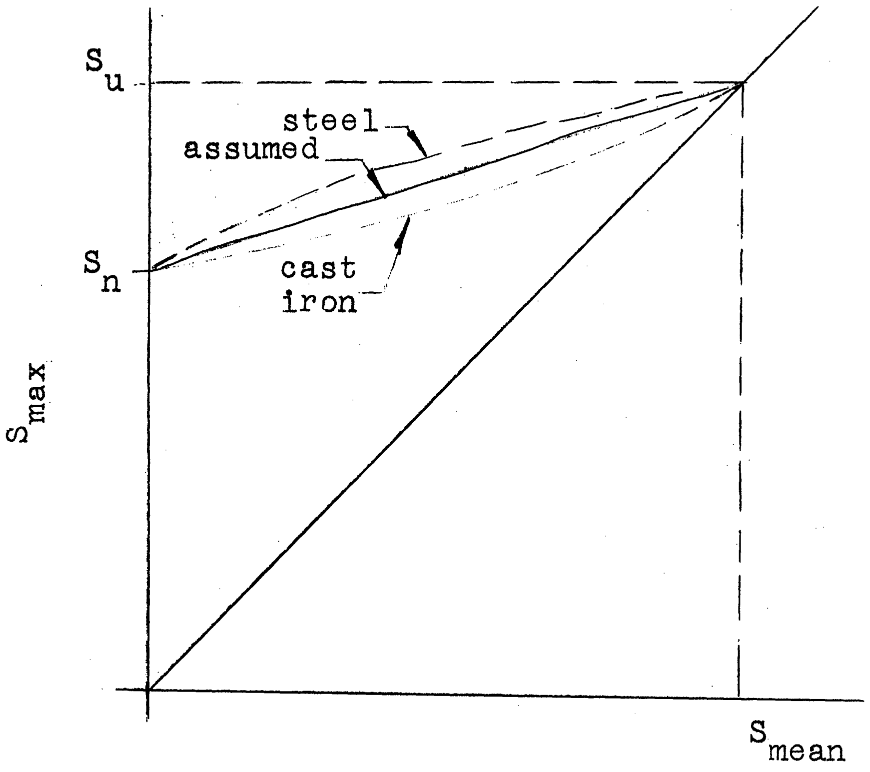


Figure 15

Modified Goodman-Diagram

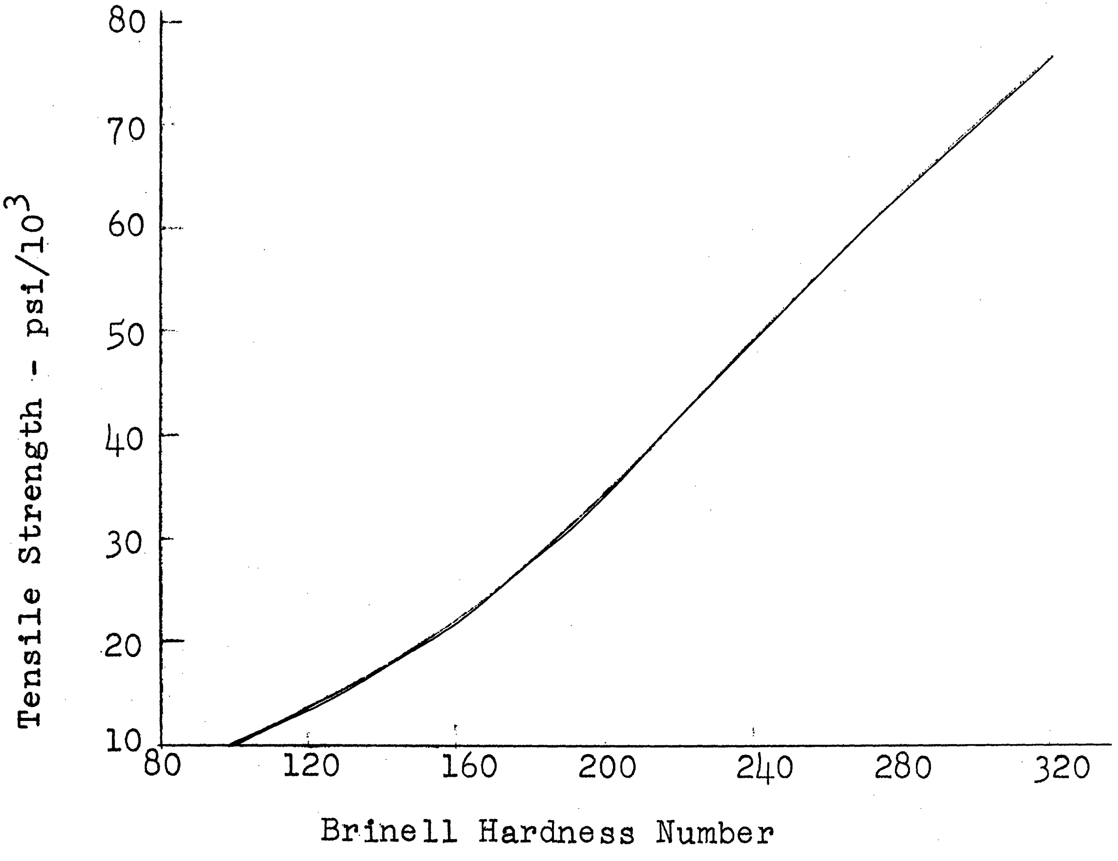


Figure 16

Tensile Strength vs. Brinell Hardness Number

3. WEAR

For our discussion we shall classify wear into three distinct categories: 1. Scuffing, 2. Normal wear, 3. Corrosive wear. Scuffing and normal wear are both mechanical types of wear.

In order to explain what we mean by scuffing or normal wear let us consider Figure 18. This is the wear pattern observed in an automotive cylinder. We notice that at the top and bottom dead center the wear is considerably higher than that at the center. This is because at these places we have high pressure intensities, low velocities, and low lubrication film. These conditions give rise to scuffing or that type of wear where considerable amounts of metal are removed by welding at the asperities.

In the central part of the same cylinder we find that the conditions are not as bad and lubrication is adequate hence we notice that wear occurs at a lower rate and is called normal wear. As the velocities are high enough there is little or no welding encountered at the asperities.

Scuffing

Coming back to our discusssion of scuffing let us look at Figure 17. In this figure are plotted two schematic curves of wear rates which could be encountered in an engine as we operate it from the time it is assembled to the time it is fully worn out. We notice that both the curves have nearly

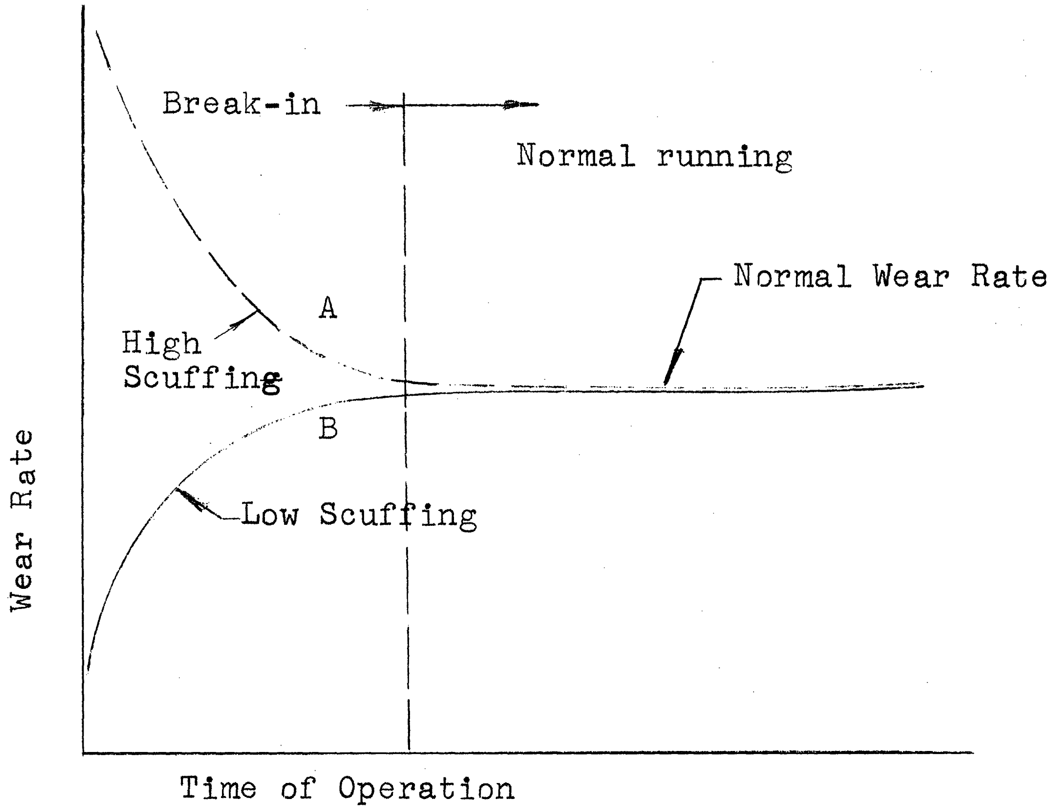


Figure 17
Schematic wear rates for different conditions of break-in.

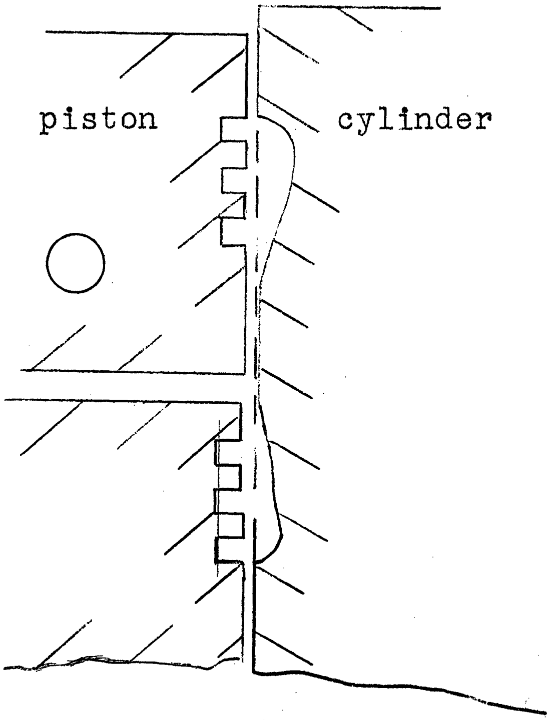


Figure 18
Wear Pattern Suggested by Jominy

the same wear rate after break in but during break in the wear rates can differ very widely. This is mainly due to the fact that, during break in, wear due to scuffing is very potent, and in Engine A this scuffing is higher than in B.

This helps us establish the causes for scuffing. Let us look at the pressure intensities in the automotive cyclinder and the corresponding wear. In Figure 19 we find that Bundy[15] has plotted the wear and pressure intensities in a Studebaker engine; we note that they follow each other very closely. This establishes our first cause for scuffing which is high pressure intensity. We also find that rubbing velocity is very low and according to Bisson this is another cause of wear. Finally due to the high temperatures encountered we find that very little lubrication is present.

Theoretical Approach

According to Davies,[24] metal to metal contact produces welding at the asperities. Now if these welds are stronger than the base material the base material gets sheared and a lot of metal can be transferred to the mating surface. This is exactly what happens in scoring. This is again aggravated by the fact that the two metals are readily soluble in each other and hence welds are more readily formed.

In order to prevent this kind of wear according to Bisson[24] 1. we should have a lubricating film with low shear strength and a backing material of high flow pressure 2. high cohesion or strength, and low adhesion or weldability.

Gray cast iron seems to meet these specifications in a very good manner. It has a hard matrix which constitutes high flow pressure. It also contains graphite which is a good anti-welding medium and a dry lubricant.

More than that it has microstructure which has small pockets or reservoirs that hold oil which provide lubrication when it is needed the most during starting and stopping periods. This has been proved to be very useful. In fact, a lot of journal bearing inserts are indented in order to hold small reservoirs of lubricant for the same purpose and they have proved very effective.

In order to evaluate each of the factors influencing scuff resistance let us evaluate them separately whilst we keep others fixed. We will observe the effects of microstructure in graphite and matrix, then we will observe the effects of chemical compositions, and finally we will evaluate the effect of surface finish.

But before we do that let us review the test procedure followed.

Test Procedure

Most of the former testing in this field was carried out by using a drum specimen which was made to rub on the mating specimen. This kind of test, which is known as an accelerated wear test, deserves a great deal of criticism because it fails to depict the true conditons encountered.

The Detroit Diesel testing program tested all the specimens in the actual environment of a 90 H.P. diesel engine. The metal specimen was used in form of a cylinder sleeve and the engine was run at 2000 RPM for 30 minutes. The cylinders were then taken apart and examined for scuffing and then reassembled and run again for a period of 30 minutes at a higher horsepower of 95. This was repeated for higher horsepowers until scuffing was observed. Then the wear resistance was noted as the ratio of the minimum H.P. needed to produce scuffing to 90 H.P. (normal). Now we know that the pressure intensity along the cylinder wall is directly proportional to the H.P. Hence what we obtain as a figure for wear resistance is actually a relative measure of pressure intensity required for scuffing to take place.

Effect of Graphite

There were five different types of specimens used to evaluate scuffing. The results are given in Table 1 and are also plotted for convenience in the histogram form in Figure 21. We find that type A cast iron has the best wear resistance especially when it has larger and randomly distributed flakes of graphite. Steel has the least wear resistance hence we observe that the oil pockets of gray iron discussed before do play an important part in giving the needed lubrication.

Effect of Matrix

A different series of tests were conducted in order to evaluate the effect of matrix. Type of graphite was kept

Test No.	Type of Graphite (a)	Total Carbon, %	Resistance to Scuffing (b)
1	None	(5150 steel)	1(c)
2	100% Type D centrifugally cast	3.25 (avg)	1.11
3	Type A, size 4 to 6, some Type B, centrifugally cast	3.28	1.33
4	Same as 3 except cast in sand mold	3.35 (avg)	1.30
5	Type A, size 3 to 4, some Type C, sand cast	4.00	1.45(d)

Table 1

Effect of Graphitic Microstructure on Resistance to Scuffing.

(a) Different chemical compositions were tested in two of the four types of iron. See Table 3 for compositions. Matrix of all specimens was tempered martensite.

(b) Expressed as the ratio of horsepower to produce scuffing divided by normal horsepower.

(c) All the steel sleeves scuffed below normal horsepower.

(d) Maximum available engine horsepower produced no scuffing.

constant at type D and the findings are tabulated in Table 2 and plotted in Figure 21.

It was found that the softest and the hardest matrix had the poorest wear resistance. This can be explained by the help of Davies, (or Bowden and Tabor's) theory that during rubbing, welding takes place at the asperities. When this happens, if the base metal is weaker (as in the case of the soft matrix) the base metal is sheared off and removed; hence scuffing is inevitable. In case of hard metals scuffing takes place due to pure abrasion. This has also been observed by Paul Lane.[20]

Again in order to avoid this kind of welding at the asperities, according to Davies we must avoid metals which are readily soluble. When free ferrite is present this welding is more potent. Hence free ferrite must be avoided and a very hard and abrasive martensitic structure should also be avoided. A good pearlitic or tempered martensitic structure is advisable.

Effect of Chemical Composition

LaBelle[24] carried out a series of tests which concluded that chemical composition had no effect whatsoever on scuff resistance. But it should be borne in mind that chemical composition does have an effect on microstructure and hence can affect scuff resistance indirectly. This has been confirmed by Shuck and Phillips.[19,23]

Test No.	Microstructure of Matrix	Hardness	Resistance to Scuffing
1	Pearlite with ferrite occurring in areas to Type D graphite	196 to 227 Brinell	1(a)
2	Martensite tempered at 400°F	Rockwell C 53 to 56	1.06
3	Martensite tempered at 800°F	Rockwell C 44 to 47	1.22
4	Martensite tempered at 950°F	Rockwell C 39 to 41	1.39

Table 2

Effect of Matrix Microstructure on Resistance to Scuffing

(a) All sleeves scuffed below normal operating range

Test No.	Type of Graphite	Total Carbon, %	Resistance to Scuffing (a)	Sleeve Wear (b)	Ring Wear (c)
1	100% Type D	3.10 to 3.40	1.11	0.003	0.020
2	Type A, size 4 to 6 some type B	3.25 to 3.50	1.30	0.002	0.027
3	Type A, size 3 to 4 some type C	4.00	1.45	0.0035	0.085

Table 3

Effect of Type of Graphite on Wear Resistance

(a) See Table 1

(b) Inches per 1000 hr.

(c) Gap-increase, in. per 1000 hr.

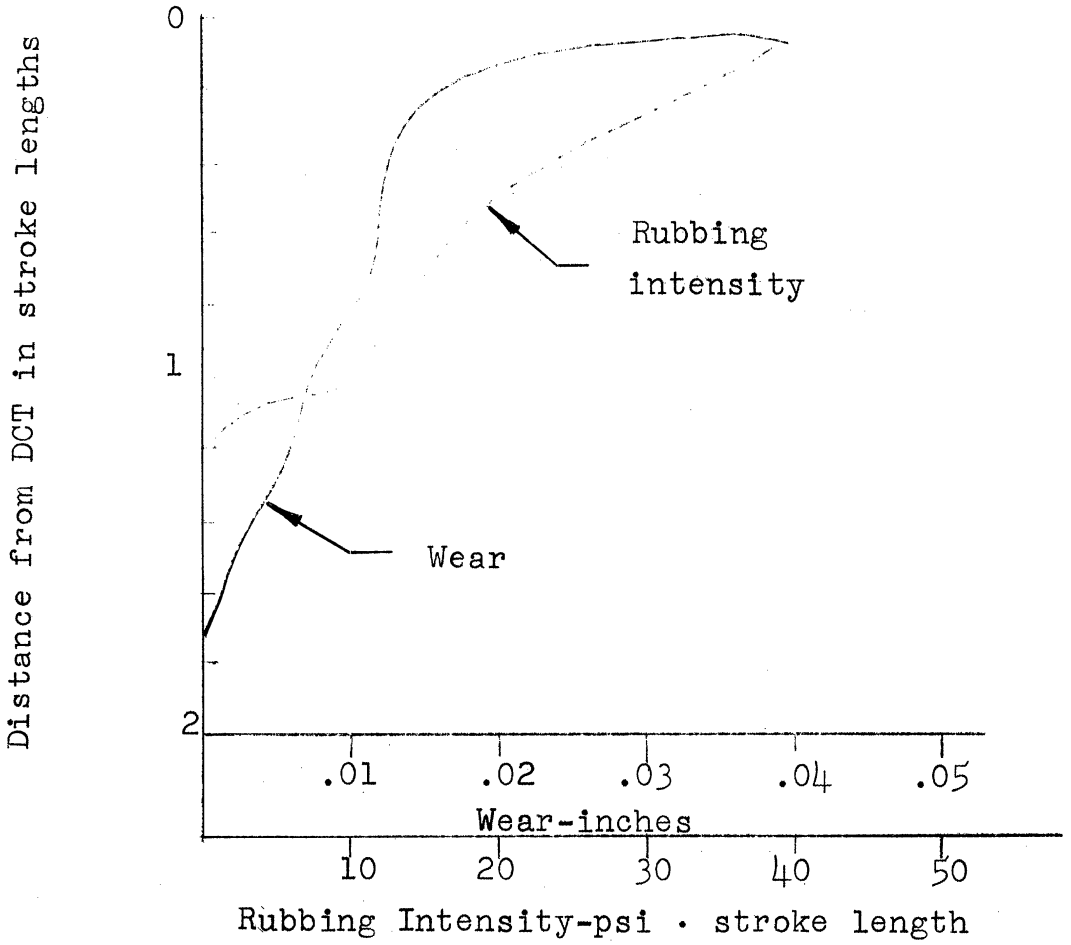


Figure 19

Relation between pressure intensity due to ring; and actual wear (as observed in 6 cylinder Studebaker engine).

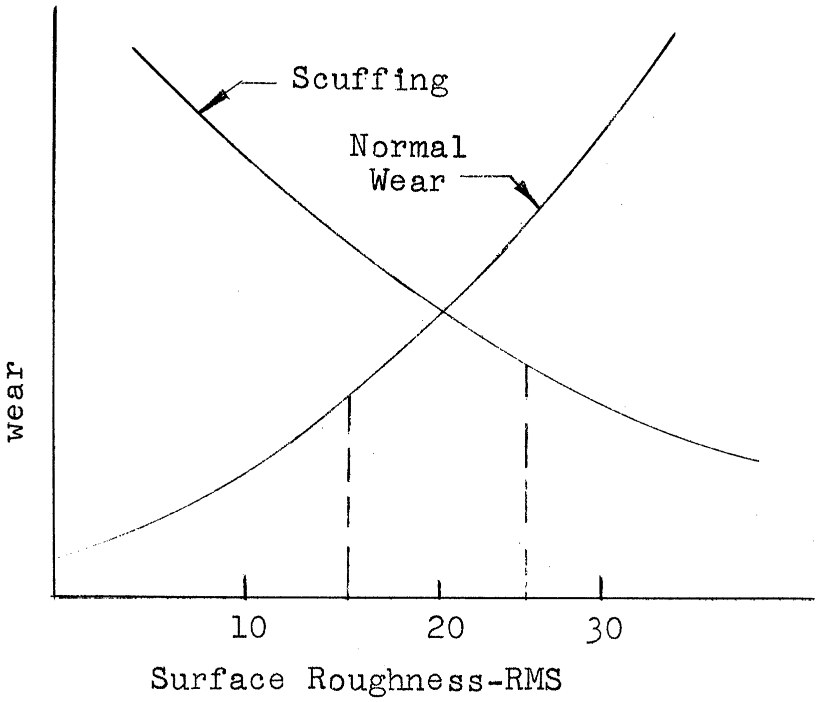


Figure 20

Schematic representation of effect of surface finish on wear

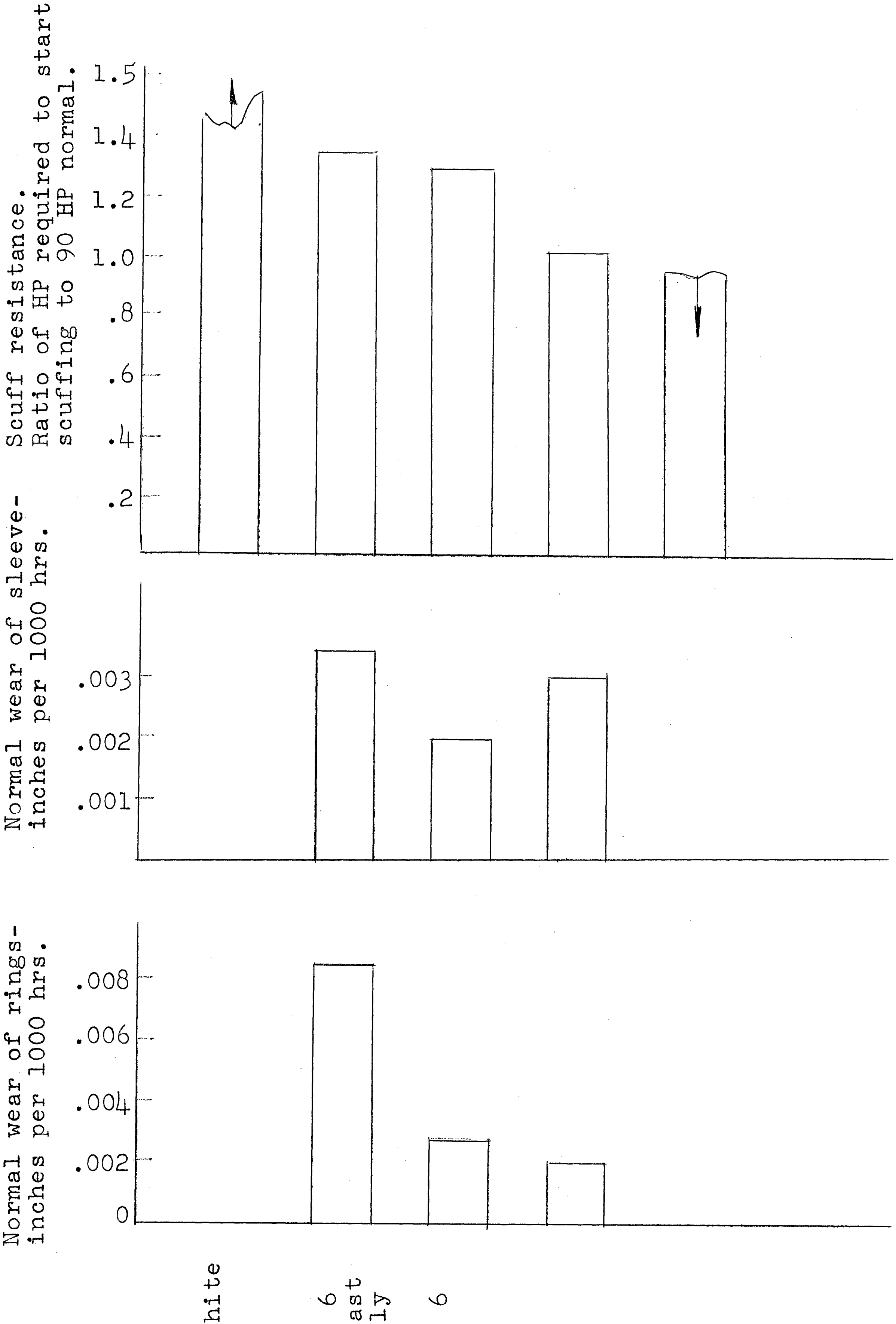

Scuff resistance.
Ratio of HP required to start scuffing to 90 HP normal.
1.5
1.4
1.2
1.0
.8
.6
.4
.2
Normal wear of sleeve-inches per 1000 hrs.
.003
.002
.001
Normal wear of rings-inches per 1000 hrs.
.008
.006
.004
.002
0
Type A Graphite No. 3 to 4
Type A 4 to 6 + Type B. Cast Centrifugally
Type A 4 to 6 + Type B. Sandcast
Type D
Steel 5150

Figure 21

Effect of Surface Finish

Referring to Figure 20, we find that as the surface roughness is increased, we increase the amount of oil pockets and hence increase the scuff resistance. But in increasing the surface roughness we find that normal wear increases tremendously. Hence we compromise at a good optimum value of 15 to 25 RMS finish.

Types of Surfaces

As long as we are discussing the effect of surface finish a few of the different types of surfaces used for prevention of scuffing are worth mentioning.

1. Inherent Surface Finish of Type A Graphite

If we observe the surface finishes of different types of metals all polished in the same manner, we find the following results:

Steel			1.5 RMS
Cast Iron			
	Fine	Type D	4.5 RMS
	Medium	Type A	9-11 RMS
	Coarse	Type A	28.30 RMS

hence we find that no matter how we finish type A graphite we cannot reduce its surface to a finer finish in order to obtain scuffing.

2. Artificially Roughened Surfaces

LaBelle[24] mentioned a case of bad valve guides of Type D graphite. In this case the type D graphite with a smooth surface caused scuffing problems. A resort to a special type of Reamed and Tapped finish alleviated the scuff resistance.

Lane mentions the type 17-6 engine developed by Ethyl Corporation. This engine had an artificially roughened surface called a Knurled Surface which showed no scuffing at all. This was because of the useful small oil pockets provided by knurling. When these pockets were clogged by impurities we found that the scuffing started because of the absence of the lubricating effect. When these impurities were removed scuffing stopped and even healing took place.

3. Chemically Prepared Surface

This surface is better known as the Parco Lubrite process. In this process the cylinder block is dipped into phosphoric acid for 15 to 20 minutes and then taken out and put into operation. The acid reacts and forms a layer of $FePo_4$ and $MnPo_4$. This again is useful as a solid lubricant and facilitates the engine's break in process; the period during which scuffing is very high.

This process also etches and relieves any working strains or residual stresses in the part.

Normal Wear of Cast Iron

Now that we have discussed scuffing in adequate detail let us consider normal wear. This kind of wear differs from scuffing in both its causes and effects. Scuffing is caused under low lubrication, high pressure intensities, and low rubbing velocities. Normal wear on the other hand is caused under adequate lubrication, lower pressure intensity and higher rubbing velocity.

In scuffing we obtain a large amount of welding at the asperities. In normal wear we have a lot of abrasion as an effect. More significant is the fact that normal wear will occur at any H.P. pressure, or velocity, whilst scuffing will occur only if a certain H.P. is exceeded. Hence we find that normal wear cannot be measured in the units of H.P. which we used for scuffing, but will be measured in inches of metal removed.

The tests were run on the same diesel engines rated to 90 H.P. but in this case the engines were run for 1000 hours to minimize the effect of break in. Piston rings play an important part in the overall picture and these were kept identical in the sense that they were all chromium plated.

Effect of Graphite

This is shown in Table 3 and Figure 21. It shows that as graphite gets coarser the normal wear increases, because as the graphite size gets larger the load carrying surface of the matrix is reduced.

This also correlates with the theory of Bisson because as the amount of graphite is increased the flow pressure (compressive strength) is reduced, thus increasing the wear since

$$\text{Wear} \quad \frac{\text{shear strength of film}}{\text{flow pressure of backing surface}}$$

Also the ring wear increases tremendously because the surface gets rougher, and hence the abrasion rate is higher.

Effect of Matrix

It was found that a soft matrix is bad because it tends to scuff or smear. But it is also noted that the very hardened matrix is bad because it causes abrasion in normal wear conditions hence a microstructure of composition with no ferrite and a hardness of around 250 BHN is found to be ideal. Ferrite is avoided because of its binding properties. Here the effect of phosphorus needs mention. It was found that phosphorus containing cast iron has higher normal wear resistance. This is because phosphorus forms a eutectic called steadite which has a low melting point and low hardness and acts as a lubricant which has low shear strength. This principle is made use of in materials like Babbitt used for journal bearings.

Effect of Surface Finish

Higher degree of surface roughness increases the abrasiveness of cast iron, and hence increases normal wear, therefore, a lower surface roughness is recommended, but this again causes higher scuffing; hence a value of 15 to 25 RMS is a good criterion.

Corrosive Wear

This is one aspect of wear in cast iron which was completely ignored by LaBelle of Detroit Diesel in his evaluation of cast irons.

Two people who deserve mention in this field are Cattaneo and Starkman[13] of Shell Development Company.

Testing Procedure

The Shell Oil Company had a research program in which they used an Otto cycle engine with a water cooled jacket. They evaluated wear as the weight lost in piston ring per 40 hours of test runs. Now it can be argued that the wear in piston rings changes with the types of cast iron used. But here we should note that since the cylinder cast iron and the ring materials were kept the same, the wear on piston rings is a good indication of the wear on the cast iron of the cylinder walls. The wear data was a statistical evaluation of a number of tests.

Causes of Corrosive Wear

Two major causes are a result of the fuels used. These are the CO_2 and sulphur products of combustion. Another cause is the halides used as scavenging agents and contained in the leaded fuels used to avoid preignition.

Effect of Carbon Dioxide

$$CO_2 + H_2O = H_2CO_3$$

H_2CO_3 is carbonic acid. This is a very potent reactant in liquid form but is not so harmful in gaseous form. Hence, if we increase the temperature of the jacket coolant we can decrease the CO_2 wear considerably. This is shown in Figure 23. If we cannot increase the jacket temperature conveniently, we can use an additive of an organic salt of calcium on the base lubricant and thereby neutralize the effect of the H_2CO_3 acid. This effect is shown in Figure 23.

Effect of Sulphur

Sulphur forms various reactants in the combustion process. The most potent of these is SO_3 and the effect of this compound is shown in Figure 22. Wear due to sulphur can be controlled by various factors. (a) Organic calcium salt can be added to reduce the acidity and this effect is found to be linear as in Figure 22. (b) We can also use organic salt of sodium and the effect of this is shown in Figure 24. (c) Lime pellet filters can also be beneficial and the effect of this is shown in Figure 25.

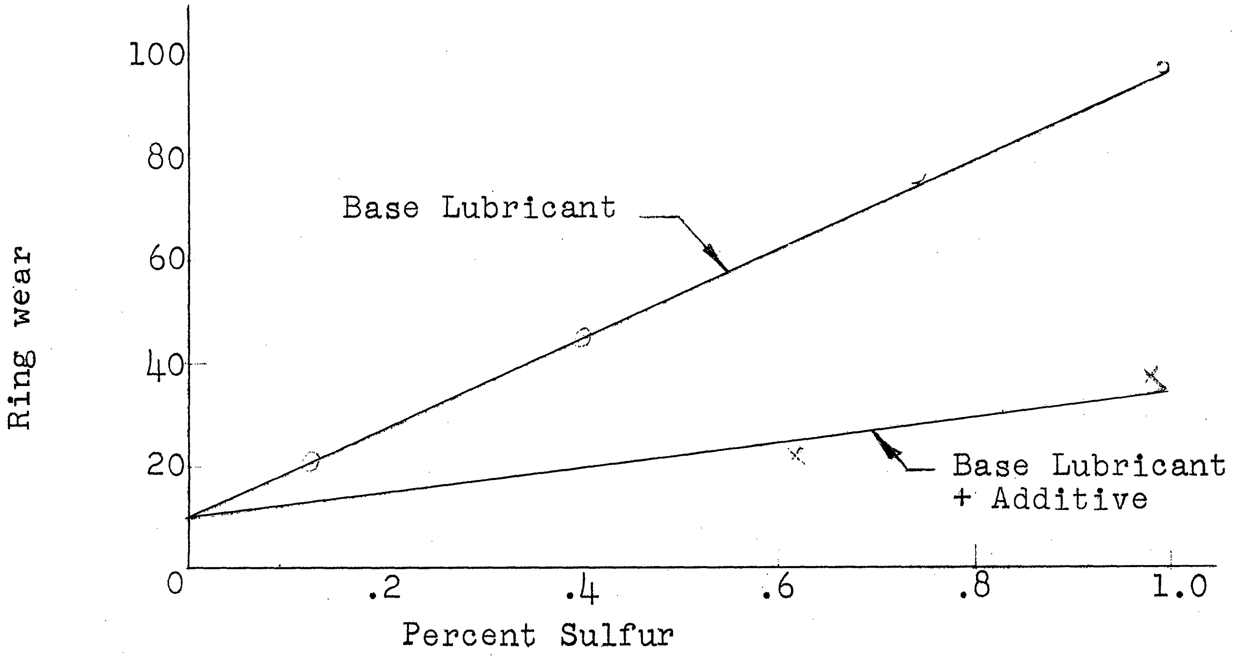


Figure 22

Effect of adding organic calcium salt to neutralize wear and deposits formed by sulphur in diesel fuels. Wear expressed as % of wear with 1% sulphur fuel and base lubricant

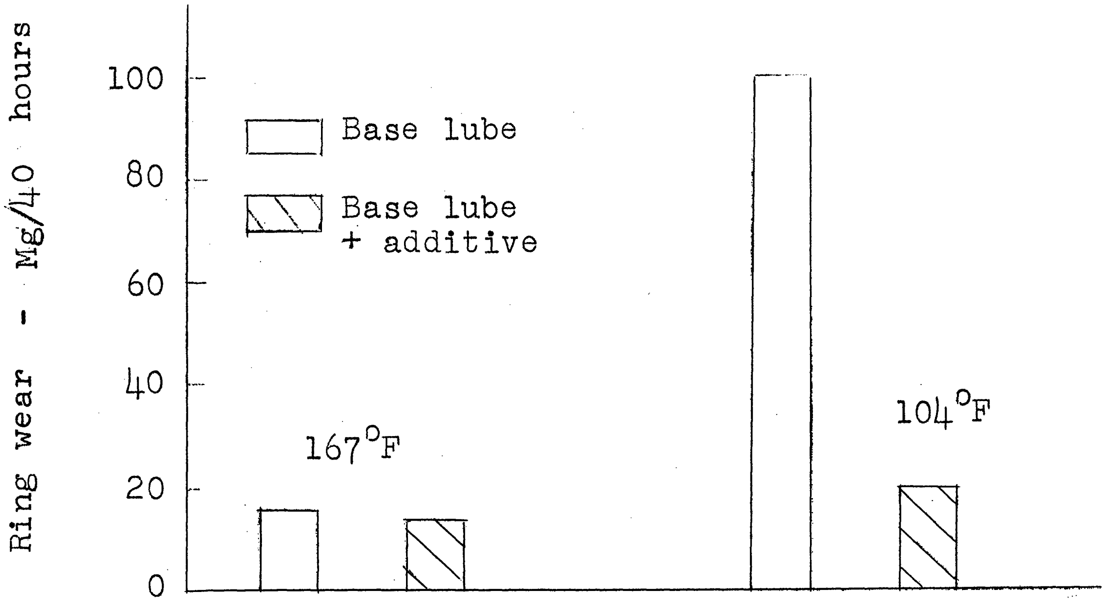


Figure 23

Basic additives added to lubricating oil to reduce corrosive wear at low temperature due to CO_2 and condensed water, but have no advantage at higher operating temperatures.

It is important to note that as more and more sulphur is contained in fuel the additives become depleted and the effect of this is shown in Figure 25.

Effect of Leaded Fuels and Their Scavenging Agents

In gasoline engines we find that in order to prevent pre-ignition in the engine, lead is added to the fuel. To remove this lead from the combustion zone, a lot of scavenging agents such asethyldene dibromide are used. It has been found that this scavenging agent is responsible for corrosive wear. Hence we have to use the proper additives, and the effect of additives is shown in Figure 26. When we see that the corrosive wear is reduced to 1/5 of the original with a base lubricant.

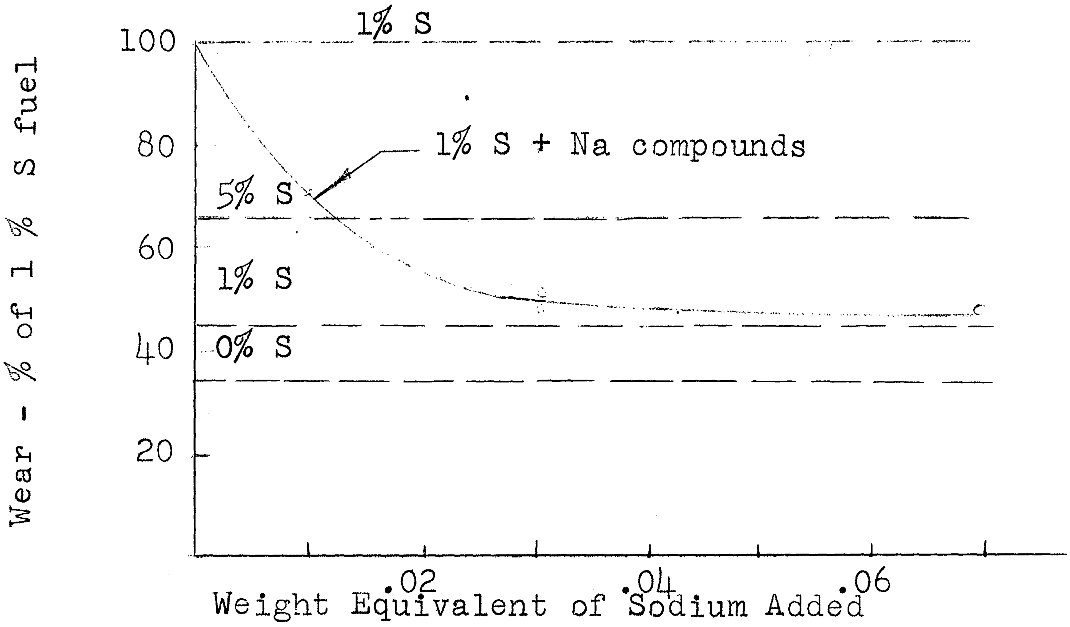


Figure 24

Effect of organic salt of sodium on wear level of diesel engine using fuel containing 1% Sulphur. Wear expressed as % of wear with 1% Sulphur fuel.

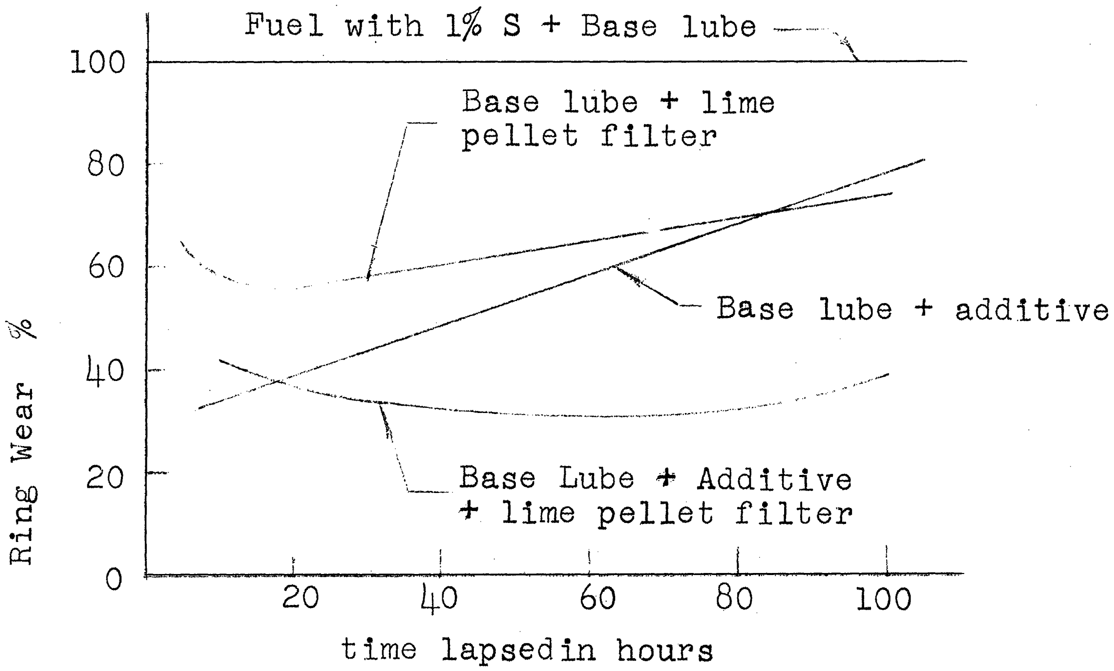


Figure 25

Gradual exhaustion of additives due to high sulphur fuel. Wear expressed as % of wear with 1% S Fuel.

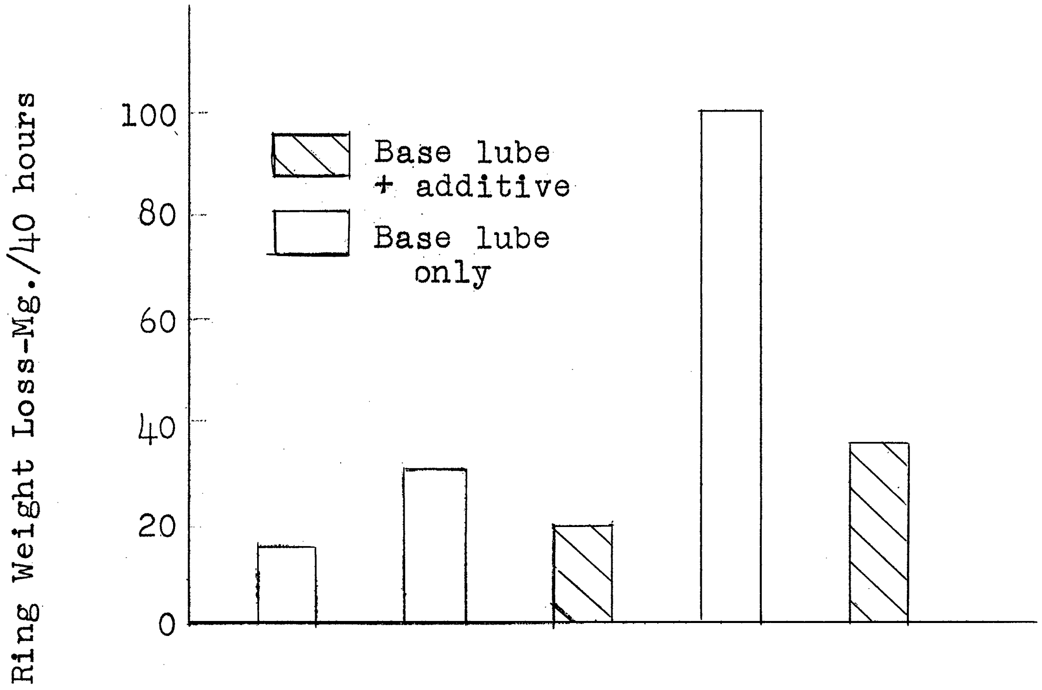


Figure 26

Effect of lubricant additives on wear due to lead and scavenging agents in gasoline engines.

BIBLIOGRAPHY

1. Metals Handbook, American Society for Metals, 1948 Edition.
2. Metals Handbook, American Society for Metals, 1955 Supplement.
3. Cast Metals Handbook, American Foundrymen's Association, 1944 Edition.
4. Modern Pearlitic Malleable Castings Handbook, Malleable Research and Development Foundation, 1958 Edition.
5. The Making, Shaping and Treating of Steel, United States Steel Corporation, 1957.
6. "Nodular Cast Iron," ASM Committee on Nodular Iron, Metal Progress, Vol. 66, No.1, July 1954.
7. Marshall, L. C., and Pearson, D. A., "Properties of Malleable Iron at Elevated Temperatures, Metal Progress, Vol. 77, No. 3, March 1960.
8. Guy, A. G., Elements of Physical Metallurgy, Addison-Wesley Publishing Co., 1959.
9. Handbook of Cupola Operation, American Foundrymen's Association, 1946.
10. Gray Iron Founders' Society Inc., "The Engineering Properties of Gray Cast Iron," Battelle Memorial Institute, 1939.
11. American Foundrymen's Association, "Symposium on Cast Iron," American Foundrymen's Association, Chicago, Ill., 1933.
12. Rosen, C. G. A., "Wear in Diesel Engines," American Society for Metals, 1950.
13. Cattaneo, A. G., and E. S. Starkman, "Fuel and Lubrication Factors in Piston Ring and Cylinder Wear," American Society for Metals, 1950
14. Boyer, R. L., Design Aspects of Cylinder and Ring Wear, Cooper-Bessemer Corporation, 1948.
15. Bundy, F. P., Theory of Function and Wear of Piston Rings and Cylinder Walls, Cooper-Bessemer Corporation, 1948.
16. Burwell, J. T., Wear Tests and Service Performance Interpretations of Tests and Correlation with Service, American Society for Metals, 1951.
17. Gray Iron Castings Handbook, 1958, Gray Iron Founders Society.
18. Bowden, F. P., and Tabor, D., Mechanical Wear, American Society for Metals, 1950.
19. Shuck, A. B., "A Laboratory Evaluation of Some Automotive Cast Irons," Trans. American Foundrymen's Society, 1948.

20. Lane, P.,"Bore Wear from the Viewpoint of Materials," SAE Journal, Vol. 45, No. 4, 1939.

21. Sefing, F. G., On Wear Resistance of Gray Iron, ASME Petroleum Engineering Conference, Tulsa, Oklahoma, October, 1946.

22. Eagan, T. E., Wear as Applied Particualarlyto Cylinders and Piston Rings, Cooper-Bessemer, 1948.

23. Phillips, G. P., "Hardened Gray Iron, " Foundry, Jan.,1952.

24. Lipson, C., and Colwell, L. V., Engineering Approach to Surface Damage, University of Michigan Press, Ann Arbor, Michigan, 1958.