2nd Edition

Materials
Degradation and its Control By Surface Engineering

A W Batchelor • Loh Nee Lam • Margam Chandrasekaran

Imperial College Press

Materials

Degradation and its Control by Surface Engineering

2nd Edition

This page is intentionally left blank

Materials

Degradation and its Control by Surface Engineering

2nd Edition

A. W. Batchelor
Monash University, Malaysia

Loh Nee Lam
Nanyang Technological University, Singapore

Margam Chandrasekaran
A Star Singapore Institute of Manufacturing Technology, Singapore

ICP Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

First published 2002
Reprinted 2006

MATERIALS DEGRADATION AND ITS CONTROL BY SURFACE ENGINEERING (2nd EDITION)

ISBN 1-86094-334-9

Printed in Singapore by Mainland Press

Dedicated to

Mrs. Valli Boobal Batchelor

Mrs. Yong Foon Loh

&

Sri. Comal Neelakantan Margam and Srimathi Rajeswari Margam.

PREFACE

In this second edition, the authors have included information about recent developments, such as the degradation of medical implant materials and provided some photographic illustrations of corrosion and surface coating technologies. The purpose of the book remains the same, i.e. to provide a comprehensive view of subjects that are traditionally treated as separate, i.e. the various forms of materials degradation and protective coating technologies. While the established subjects of aqueous electrode theory for corrosion and mechanical causes for wear are given adequate coverage in this book, less familiar problems such as liquid metal corrosion are also covered. An applied approach to the problem of materials degradation is offered to help engineering students to appreciate the significance of materials degradation. The wide variety of materials degradation that is found in practice, e.g. damage to polymer and ceramics is discussed here to enable students to recognize potential materials degradation problems in any given situation.

In industry, materials degradation of components and equipment comes in the form of wear caused by materials processing and handling, corrosion from process chemicals and natural water and fracture from fatigue and shock loading. In some industries, thermal and radiation damage may also be significant. In the absence of premature technical or economic obsolescence, the lifetime of any mechanical device is determined by which is the faster process, wear, corrosion or fracture. Engineering students therefore need a comprehensive view of corrosion, wear, fracture and their preventive technologies. In most courses on corrosion, electrode theory has been emphasized at the expense of detailed discussion of painting and other important coatings. Most practicing engineers need to know more about polymeric coatings and their effective use than the fine details of electrode theory. This book on materials degradation (or corrosion, wear and fracture) and its control is intended to help students to achieve a balanced knowledge of materials degradation and how to control it. They should then be able to either solve basic problems in materials or interact effectively with specialists. The lack of awareness of materials degradation by many engineers belies its importance. It has been known for an engineer to replace a zinc sacrificial anode with a stainless steel anode because the zinc anode required frequent replacement. As a consequence of this action, a large mineral processing tank became severely corroded necessitating replacement.

The authors aim to provide the reader with a readable and generously illustrated guide as to what is materials degradation and how it can be controlled. The nature of corrosion, wear and fracture is first described. Methods of materials conservation are then discussed and followed by examples of actual applications. The book is intended for final year students of an undergraduate mechanical or materials engineering course. Postgraduate students and professional engineers may also find the book useful.

ACKNOWLEDGMENTS

The authors would like to acknowledge with thanks the kind support of Monash University Malaysia, the School of Mechanical and Production Engineering, Nanyang Technological University, Singapore (NTU) and the A-Star Singapore Institute of Manufacturing Technology (Singapore) during the writing of this book. The authors also would like to acknowledge with thanks, Dr Asim Bag formerly of NTU for his help with the photomicrographs of carburized layers and Mr Aw Poh Koon for the photograph of thermal spraying. The help of Doxon Engineering Pte. Ltd Singapore in providing the photographs of induction hardening, plasma nitriding and plasma nitrided specimens is acknowledged with thanks. The efforts of Professor V.C. Venkatesh, originally of Tennessee State University, for reviewing this book are also much appreciated. The authors acknowledge with thanks Dr Tam Kam Chiu for his support and encouragement during writing of this book and Mr Richard Seng Wee Pin and Mr Ng Wei Kiet for proof reading the manuscript. The authors are also grateful to Bofors Ltd. for permission to reproduce illustrations from their handbook. The authors are also grateful to Mrs. Valli Boobal Batchelor, Mrs. Yong Foon Loh and Mrs. Lakshmi Chandrasekaran for their patience and help during the preparation of the 2nd edition.

This page is intentionally left blank

CONTENTS

CHAPTER 1

INTRODUCTION

Accumulated human experience over several millennia in the adaptation of natural materials for useful purposes has caused the notion of materials degradation to be firmly embedded in language and culture. The familiarity of such words as 'fracture', 'wear out', 'burn up' or 'rot' are evidence of the close relationship between human civilization and materials degradation. An instinctive appreciation of surface engineering is present in almost all civilized peoples where the quality of a surface controls much of the value of the object enclosed within. Common examples of this association of value with effective surface engineering are the emphasis given to the polishing or to the painting of mechanical equipment. A single scratch on some new article often causes depreciation in its financial value that is much greater than the cost of mechanical damage caused by the scratch.

Whereas lay persons usually give priority to the subjective significance of materials degradation and surface engineering, it is the task of technologists to evaluate the true costs of materials degradation and balance this against the costs of surface engineering necessary to prevent such materials degradation.

1.1 Definition of materials degradation

Before proceeding into a subject, which depends on an enormous variety of information from many diverse sources, some pertinent questions would have to be asked: What is materials degradation? How does it affect an engineering system? What is the significance of materials degradation in terms of engineering performance?

From the moment that any material is released from the point of production, it is in theory subjected to some form of materials degradation although such degradation may not be readily observed or measured. The rapid rusting of freshly machined steel surfaces is a common example of immediate materials degradation. Such damage continues throughout the lifetime and materials degradation of any component be a limiting factor. No known service environment provides perfect immunity from materials degradation. Under atmospheric conditions, electrochemical corrosion is the most significant degradation process in economic terms. In outer space, although the absence of atmosphere precludes corrosion, radiation damage can be severe. Corrosion, radiation damage and many other mechanisms all have a shared feature of reducing the performance of engineering materials to cause premature failure of components and devices. A simple definition of materials degradation is that it is the consequence of a wide range of physical processes; it is almost universal in occurrence and is a major engineering problem.

Materials degradation imposes a cost penalty on all engineering systems. For instance, a mechanical structure has to be constructed with extra metal or concrete to allow for corrosion-induced loss of strength. If such corrosion did not occur, the extra metal could be dispensed with and the structure could carry more weight. The corrosion ensures that a more expensive structure is required for any given load. Wear between pistons and cylinders inside an internal combustion engine causes leakage of combustion gases from inside the cylinders and the engine becomes less efficient. Piston-cylinder wear causes a vehicle to consume more fuel per distance travelled than would otherwise be the case. The risk of sudden fracture of an aluminum structure necessitates frequent and expensive preventive repair of aircraft fuselage. Further examples of loss of performance are discussed below in this chapter.

Materials degradation can be defined in terms of loss of performance of an engineering system. Loss of performance can relate to many parameters, e.g. loss of reflectivity in optical equipment caused by fungal attack. A more common example might be the loss in mechanical strength of a structural component exposed to a corrosive medium. For any item of equipment, there is a critical minimum level of performance, e.g. whether a useful image is obtained from the optical system or whether the mechanical structure will collapse due to corrosion damage. For the engine with worn cylinders, wear can increase the clearance between piston ring and cylinder to such an extent that there is no compression of combustion gases. In this case the engine can be considered to have failed, as it will no longer be able to pull the car or truck up the hill. A mechanical degradation proceeds at a rate that varies with local conditions and failure occurs if the performance declines to below the critical level. Loss of efficiency occurs if performance declines but remains above the critical level during the service lifetime. Referring to the previous examples of the optical system and mechanical structure, loss of efficiency could involve a progressive distortion of the image for the optical system and load restrictions for the mechanical structure. This view of mechanical degradation is illustrated schematically in Figure 1.1

Some loss of performance is inevitable unless very expensive control measures are implemented, e.g. constructing a car from stainless steel instead of plain steel. The objective of surface engineering is to ensure that performance remains far above the critical level for the entire service lifetime of the system. The scientific study of degradation of engineering materials can be summarized as predicting the rate of decline of performance; this is the gradient on the graph shown in Figure 1.1. Engineering analysis of materials degradation is directed at finding the factors controlling this gradient and how to reduce it.

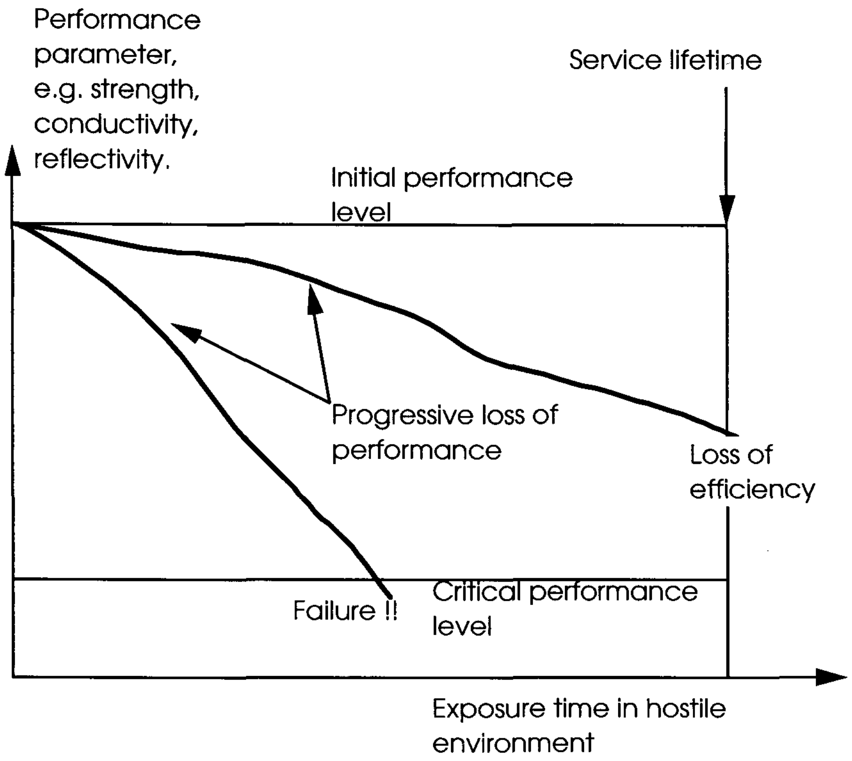


Figure 1.1. Graphical definition of materials degradation as loss of performance of an engineering system.

1.2 Definition and significance of surface engineering

Surface engineering is a comparatively recent term that refers to control of problems originating from the surface of engineering components. It is generally considered that the surface of a component is much more vulnerable to damage than the interior of the component and that surface originated damage will eventually destroy the component. Most types of materials degradation such as, wear, corrosion and fracture are usually located at the surface of a component. As a result of this concentration of damage processes at the surface of a component, surface engineering is essential to control these damage processes.

A time-honored solution to the problem of materials degradation is to try and shield the material from the hostile agent. For example, painting is a well-established example of this practice and serves to shield a metal (or wood) from water and oxygen. The range of techniques available to prevent materials degradation by shielding has expanded enormously in recent years. The principal of shielding by surface engineering is illustrated schematically in Figure 1.2.

Corrosion damage is illustrated in Figure 1.2 but the same principle applies to wear and surface-originated fracture.

Shielding is usually achieved by coating the material with another substance or by generating a surface modified layer, which is more durable than the original material. Another form of shielding is to inhibit the chemical processes that directly control degradation. Cathodic protection where an electrical potential is used to suppress electrochemical corrosion of metals is an example of this latter method. The vast majority of materials degradation problems are, however managed by the former method of coatings and surface modified layers. Most physical phenomena have been applied where possible for the development of coatings and surface modified layers. The power of an explosion has been used to directly clad steel with a corrosion resistant layer of aluminum. Each method of coating or surface modification has its own advantages and disadvantages. For instance, painting may be cheap but will it be effective when the corroding water is substituted by abrasive aqueous slurry? A comprehensive knowledge of the various coating and surface modification techniques is required before an optimum or near-optimum solution to a materials degradation problem can be obtained. A definition of surface engineering is an informed selection of the appropriate coating or surface modification technology and its effective application to prevent or delay one or more forms of materials degradation thereby improving the performance of components and devices.

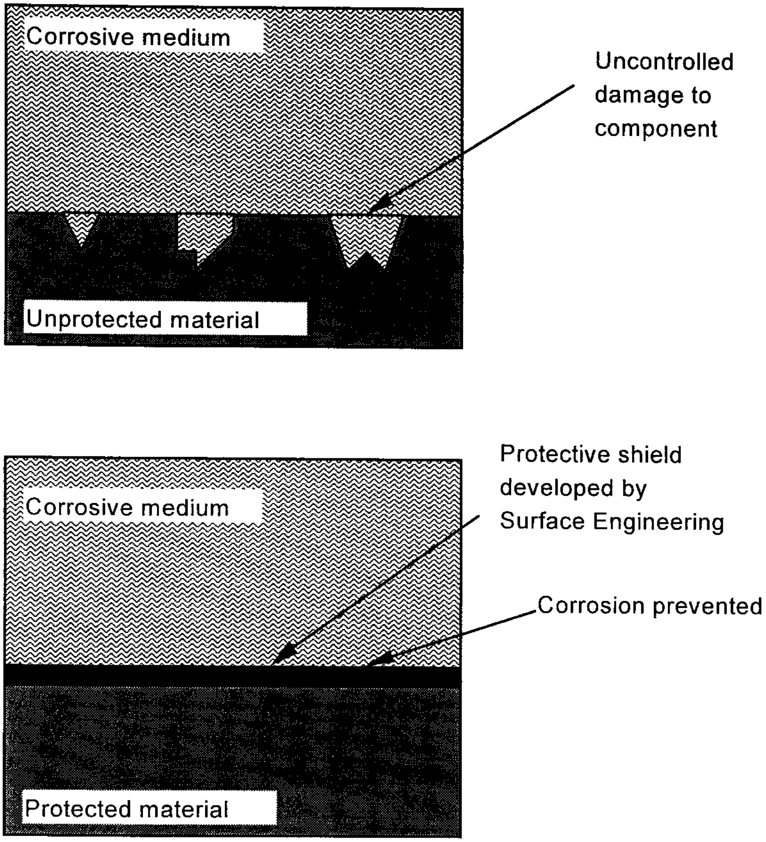


Figure 1.2 . Graphical definition of surface engineering to control materials degradation.

1.3 Classification of materials degradation by physical mechanism

Almost any known natural phenomenon can cause materials degradation. Heat, light, short-wavelength electromagnetic radiation, radioactive emissions, chemicals, mechanical stress and interactions with bacteria, fungi or other life forms; all these can damage materials. Classification of materials degradation according to its basic cause is a first step in any plan to protect a specific material in a specific situation. A classification of the known forms of materials degradation in as physical, chemical or biological phenomena is illustrated schematically in Figure 1.3 as a Venn diagram.

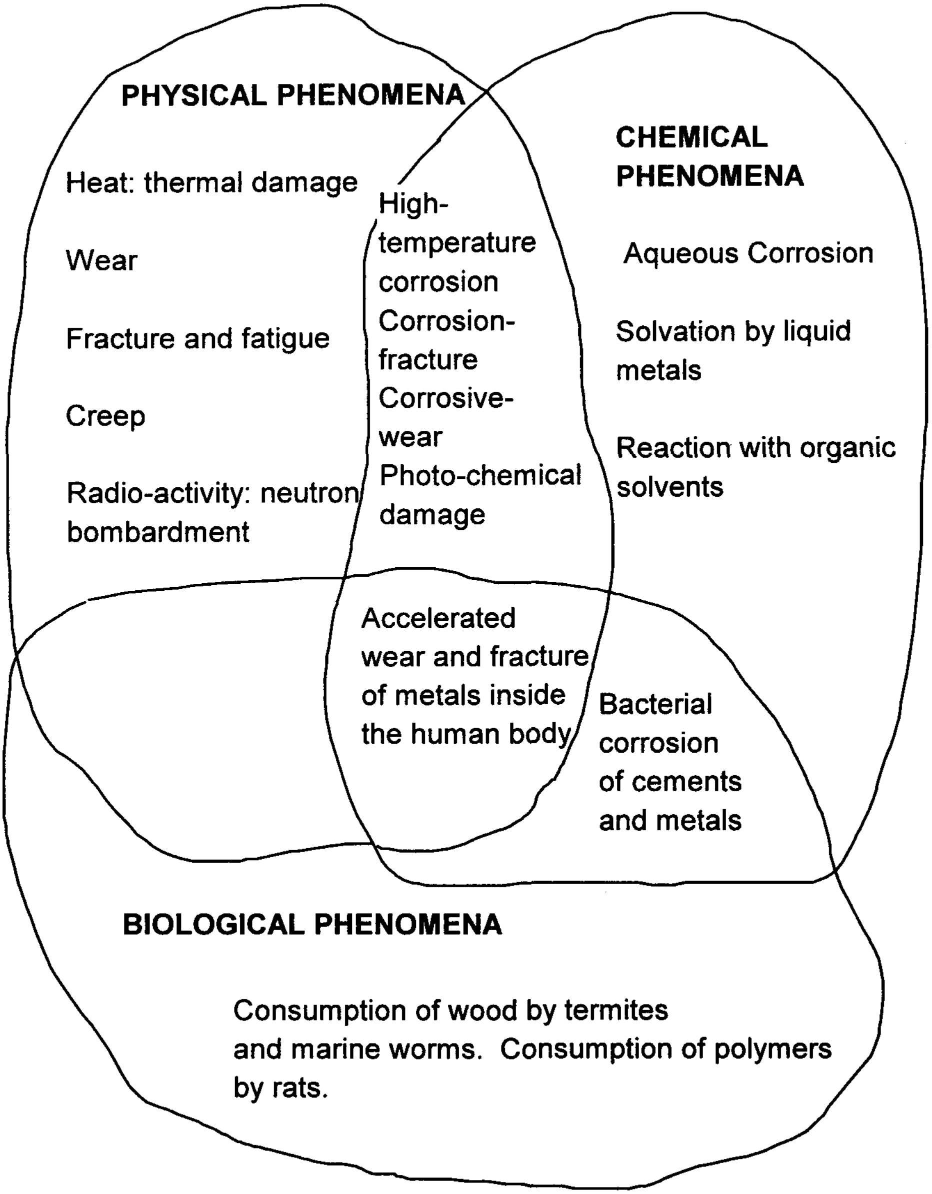


Figure 1.3. Classification of the various types of materials degradation in terms of physical, chemical and biological phenomena.

There are three basic categories of materials degradation, physical, chemical or biological in origin. Physical origin refers to the effect of force, heat and radiation. Chemical origin relates to destructive reactions between the material and chemicals that contact it.

Biological origin includes all interactions between life forms and engineering materials.

Materials degradation such as thermal damage or chemical reactions, which are either entirely physical or chemical in nature, coexists with combined forms of materials degradation such as corrosive-wear. Materials degradation is an uncontrolled process without restriction on interactions between seemingly unrelated events. Environmental conditions also exert a strong effect on materials degradation with the result that any given materials degradation problem is effectively dependent on locality. As will be shown in later chapters, there is a wide range of possible interactions or synergy between degradation processes and considerable care is needed to predict which of these interactions are significant to any given situation.

An important example of interactions between degradation processes is Corrosive-wear, which is mechanical wear that is accelerated by chemical damage to the worn material. Biological phenomena are either found in a pure form, i.e. organisms eating artificial materials or in a complex form such as bacterial corrosion where the waste products of bacteria are destructive to materials. Biological, chemical and physical phenomena all interact in the human body to impose severe stresses on any metal component implanted in the human body.

Damage that depends on just one type of phenomena, e.g. physical is easier to recognize than multi-phenomena modes of damage such as those occurring inside the human body. Protective measures are usually ineffective when a multi-phenomena mode of damage is mistakenly concluded to a single-phenomenon mode of damage. In this book, single-phenomena modes of materials degradation are described before more complex forms of degradation are discussed. An example of this approach is the prior discussion of abrasive wear and aqueous corrosion before corrosive-abrasive wear is introduced.

1.4 Economic and technical significance of materials degradation

How much will materials degradation cost the company or project? What are the technical problems caused by materials degradation? Practicing engineers and technologists need to know the answers to these questions to ensure the success of any engineering venture. An idea of the scale of costs incurred by materials degradation can be obtained by trying to estimate the value of all the engineering equipment that is currently in use in the world. A figure reaching perhaps trillions of dollars would not be surprising. Suppose now that materials degradation causes an effective depreciation of 5% a year in capital value, which is a very conservative estimate. The cost of materials degradation would then be in the range of $100 billion to $1 trillion a year. Expert studies have pursued this method of estimation of costs in far greater detail and arrived at similarly high cost estimates. Friction and wear losses alone cost about 1% of the GNP of Germany or about 10 billion DM per annum in 1975 prices [1]. Corrosion is estimated to impose a similar level of costs on advanced economies with one estimate being 5% of GNP. Admittedly, the idea of a direct cost imposed by materials degradation is simplistic in economic terms as much economic activity is generated by replacement of corroded or otherwise degraded components. However, even with the beneficial effect of extra economic activity, materials degradation represents an undesirable financial burden, which is the task of engineers to try and reduce as much as possible.

As discussed above, before a component or device must be replaced because of materials degradation, some loss of performance would have already occurred. Rusting and corrosion can reduce the factor of safety in mechanical strength so that a structure or component can fail unexpectedly during severe loading. In 1989, a Boeing 747 aircraft suffered an in-flight structural failure, which is believed to be due to localized corrosion and metal fatigue. Fortunately, the structural failure was limited to the loss of a door and the plane was able to land safely. Corrosion products of pipe materials can easily become contaminants of the fluid conveyed by the pipe. Dissolved metal emanating from the stirrer blades of a casting system can compromise the metallurgical content of the cast metal. Wear of screw threads and gears causes a progressively increasing backlash in mechanical transmission. This will prevent accurate location of a cutting tool or milling cutter. Wear of cutting tools causes increased roughness and dimensional inaccuracy on the cut material. This corruption of mechanical performance by materials degradation must be considered when devising reliable engineering systems.

Before the reader is persuaded that materials degradation is entirely destructive, it should be noted that the same phenomena, which in uncontrolled form cause damage, could also be controlled to achieve useful results. For example, the interaction between corrosion and abrasive wear by small hard particles is found to polish silicon wafers much more efficiently than pure abrasion by the hard particles alone. In this instance, the acceleration of abrasive wear by corrosion reduces the time required to remove machining marks from a silicon wafer surface and generate an extremely smooth surface on the wafer. The key factor is control, which is very costly to implement and cannot be readily applied on a routine basis to any engineering system. This is a major reason why materials degradation is suppressed instead of being arranged to occur at some pre-set time. It would be economically advantageous if a large metal structure such as a ship could be planned to self-destruct into small manageable lumps of metal at the end of its service life. At present, obsolete ships are manually dismantled by large teams of workers in countries with low labor costs. It is probable that in the future, all countries will reach similar labor costs with no cheap labor countries. Ship breaking will then become more expensive and impose an extra cost on ship construction unless an alternative method is found.

1.5 Summary

Human culture has recognized the importance of materials degradation since the beginning of civilization. Many methods of protecting materials have been devised based on either traditional practice or more recently on scientific knowledge. There are many forms of materials degradation, which require a specific mode of protection, which may not be applicable to another form of material damage. Materials degradation causes such a high level of costs to industry and individuals that effective counter-measures are essential. Damage to materials can lead to either complete failure of an engineering system or else its loss of performance, efficiency or accuracy.

RECOMMENDED READING

1. Research Report (T76-38) Tribologie (Code BMFT-FBT76-38), Bundesministerium fur Forschung und Technologie (Federal Ministry for Research and Technology), West Germany, 1976.

This page is intentionally left blank

PART 1: MECHANISMS OF MATERIALS DEGRADATION

Overview

In Part 1 Mechanisms of materials degradation, the characteristics of the economically significant forms of damage to materials is discussed in terms of the physical and chemical processes involved. There is a wide range of materials degradation mechanisms but all can be classified into four basic categories; (i) by direct mechanical action, (ii) by heat or radiation, (iii) by the presence of chemical reagents and (iv) where two or more of (i), (ii) and (iii) combine synergistically. The economically significant forms of materials degradation are discussed sequentially in the following chapters according to this classification.

This page is intentionally left blank

CHAPTER 2

MECHANICAL CAUSES OF MATERIALS DEGRADATION

2.1 Introduction

When the concept of mechanical causes of materials degradation is invoked, mechanical forces are necessarily involved. With the exception of a stationary object in a weightless environment, there are few examples of a mechanical structure, which does not sustain any force. Most of these forces, e.g. the weight force are not destructive but some forces and their attendant stresses can cause rapid material degradation and failure. Principal examples of these destructive forces and stresses are oscillating forces and stresses associated with fatigue, extremely localized areas of very high stress at contact areas between solid bodies that cause wear, internal stresses that accumulate within a material to initiate delamination and peeling. Degradation mechanisms that are dependent on mechanical forces, i.e. wear, fatigue and fracture are discussed below in this chapter.

2.2 Wear

Direct experience of wear and the importance of controlling it are probably as old as the invention of the wheel and axle. Ancient civilizations were familiar with failure of axles by wear and devised rudimentary lubricants to suppress the wear. Prehistoric cavemen in northern countries probably resented the fact that their primitive foot coverings of un-tanned animal skin were rapidly destroyed by wear. An adequate description of what wear is, i.e. its fundamental causes and mechanisms are in contrast only a recent discovery. A primary reason for this historical delay in understanding, is that the collection of experimental evidence of wear mechanisms relies on sophisticated microscopes and surface profilometry, which were only developed not so long ago.

Wear appears to be an inevitable consequence of relative motion between solid bodies and is fundamentally related to the nature of contact between solid bodies. Profilometry studies have revealed that almost all surfaces of solid bodies are rough where perhaps the only known exception are perfectly cleaved faces of single crystals. Artificial smoothing and polishing of a surface only reduces the scale of roughness and cannot create a truly smooth or planar surface. Surfaces of real objects are composed of an irregular series of peaks and troughs and when two solid bodies are pressed together, true contact is limited to the peaks of surface roughness which are known as *asperities.* This model of localized contact between rough surfaces is illustrated schematically in Figure 2.1.

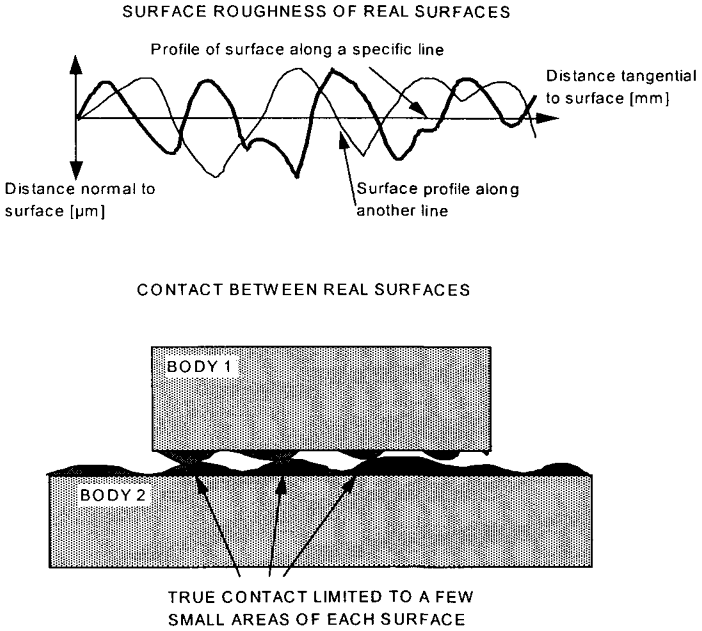


Figure 2.1. Nature of contact between rough surfaces

The scale of roughness is very small, the standard deviations of peaks and troughs from an imaginary central plane are only in the range of 0.1 to 10 μm for most machined surfaces but this is sufficient to dramatically limit the area of true solid contact. Contact stresses at the interface between two solid bodies are therefore much higher than is simply predicted by dividing the contact force by the external dimensions of the contact. Early theories predicted that the contact stress was equal to the compressive yield stress of the weaker contacting body, it is now believed that elastic contact stresses are more common but the values of stress are still large. The true contact area is typically 1 to 0.1% of the nominal contact area, which means that the true contact stress is 100 to 1000 times greater than the nominal contact stress (load/nominal contact area). The division of true contact area into numerous small areas of contact between asperities further raises the contact stress. This is because each asperity has a small radius of curvature, which raises the contact stress according to the Hertz equations of contact stress. Even these high levels of contact stress are relatively harmless if there is no relative motion. Wear begins when the asperities have to accommodate sliding motion or to a lesser extent rolling motion and impact. The combination of high stress and constrained motion initiates material deformation and other processes at contacting asperities, which lead to wear. Sliding contact generates heat through the frictional dissipation of energy and this heat has a strong influence on wear. The relationship between adhesion, deformation, frictional heating and the various forms of wear are illustrated schematically in Figure 2.2. Adhesion leads to adhesive wear and in extreme cases, frictional seizure. Plastic deformation and fracture promote a very common form of wear based on subsurface or surface crack growth. A variant of wear by deformation or fracture is abrasive wear and the related erosive wear. In this instance, the intervention of external hard particles controls the wear process. When the sliding amplitude is very small, fretting occurs which differs from typical sliding wear because wear debris is trapped inside the contact. If there is high levels of frictional heating or process heating then melting wear and diffusive wear becomes significant. Diffusive wear involves transfer of material as atoms by diffusion from one contacting body to another. All of these different modes of wear are discussed in this chapter.

The only way to prevent wear is to separate solid surfaces so that although they appear to be in contact they are in fact separated by a thin film of lubricant or coating material that is perhaps 1 μm thick. This is a basic function of lubricating oils and can be extremely effective. Another method is to separate the sliding surfaces by magnetic levitation but this involves expensive super-conducting magnets and so is limited to specialized applications.

2.2.1 *Adhesive wear*

When wear begins as a result of localized stress and relative motion, the question is what is actually worn away? At the beginning of wear, the underlying material, e.g. steel of a steel component is almost never worn away.

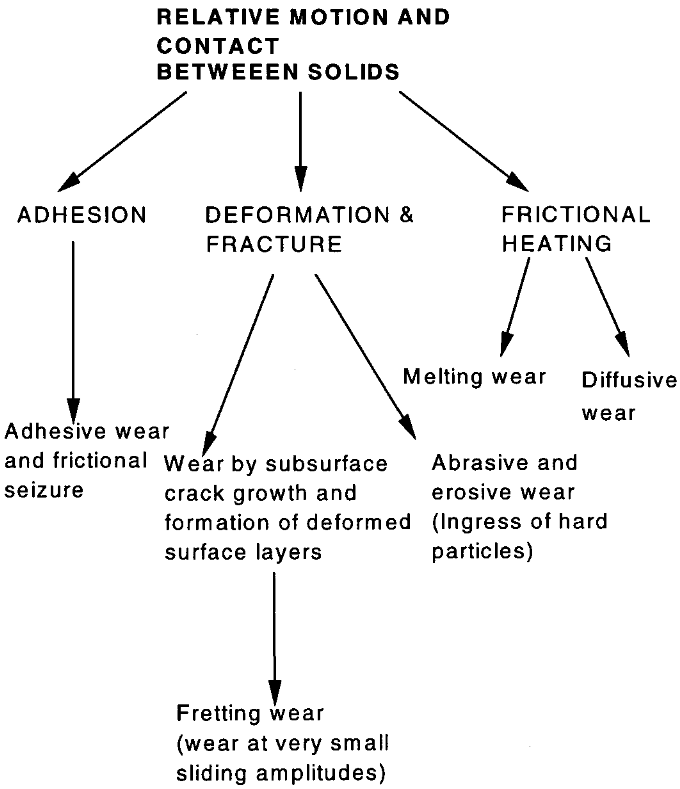


Figure 2.2 The relationship between different wear mechanisms and their physical causes.

Instead a layer of contaminants and for metals an oxide layer is usually removed. It is very difficult to maintain a surface perfectly or very clean and clean surfaces can only be reliably produced under conditions of high vacuum with cleaning by ionic bombardment. Surface contaminants in the form of water and oils always accumulate on surfaces under atmospheric conditions. The atmosphere contains oxygen, which aggressively reacts with metals to form an oxide layer on each surface. Even though a metal component may appear shiny and metallic, it is in fact being viewed through a transparent layer of oxide, which is typically 5 nm thick. At the beginning of wear, it is the contaminants and oxide films, which are worn away to only later expose the metal surfaces. What happens when the metal surfaces are finally exposed? Clean exposed metal surfaces are entities that are rarely seen and have unusual properties. A major characteristic is electron emission by metal surfaces and exchange of electrons with contacting metal surfaces. Electron exchange between contacting metal bodies exerts a fundamental influence on friction and wear. The mechanism of free electron transfer from a metal to a metal or non-metal is shown schematically in Figure 2.3.

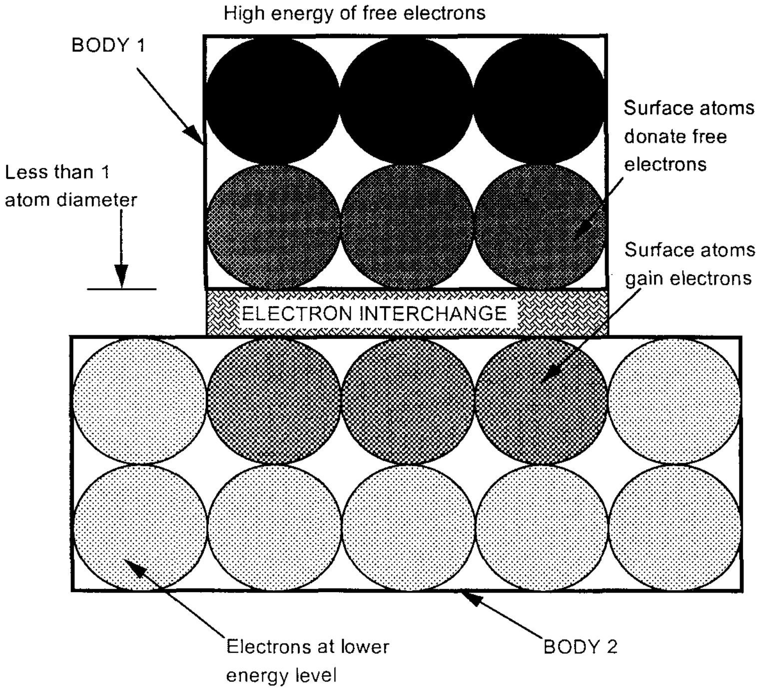


Figure 2.3 Mechanism of inter-metallic adhesion by transfer of free electrons.

Metals contain electrons, which are not bound to a specific atom but are free to move between atoms. The free movement of these electrons enables efficient conduction of electricity and heat in metals. When two clean metal surfaces approach each other within a distance that is less than the diameter of one atom, electrons from the high-energy metal transfer to the low energy metal. A strong bond between the two metal bodies is produced by this electron transfer, where the bonding is as strong as inter-atomic bonding within each metal body. The transferred electrons remain free to move between atoms so that there is a permanent interchange of electrons between the contacting bodies. Contact in this context implies approach to within one atom diameter. A similar form of adhesion is also possible when a metal contacts a non-metal such as a ceramic or polymer. In this case, the character of the electron bonding assumes a more chemical character with strong bonding being observed between metals and a reactive non-metal such as fluorine or oxygen. An important example of bonding between fluorine and metal is provided by contact between iron and Polytetrafluoroethylene (PTFE).

Adhesion based on electron transfer is immediate, spontaneous and not restricted to high temperatures. Another form of adhesion based on diffusion is also observed to occur where adhesion develops from diffusion driven mixing of the contacting bodies. This mechanism of adhesion is extremely slow and restricted to high temperatures, time periods as long as several weeks with heating to several hundred degrees Celsius are required before noticeable adhesion between metals is generated by diffusion. Sliding contact and wear are characterized by short periods of contact between asperity that last only milliseconds or even microseconds which effectively prevent diffusion based adhesion.

If the two contacting bodies illustrated in Figure 2.3 are then forced to slide past each other, a severe form of wear known as adhesive wear will be caused by the free electron bonding. In simple terms, the opposing metal bodies tear small pieces from each other while sliding. The process of contaminant and oxide layer removal followed by metal-to-metal bonding and adhesive wear is illustrated schematically in Figure 2.4.

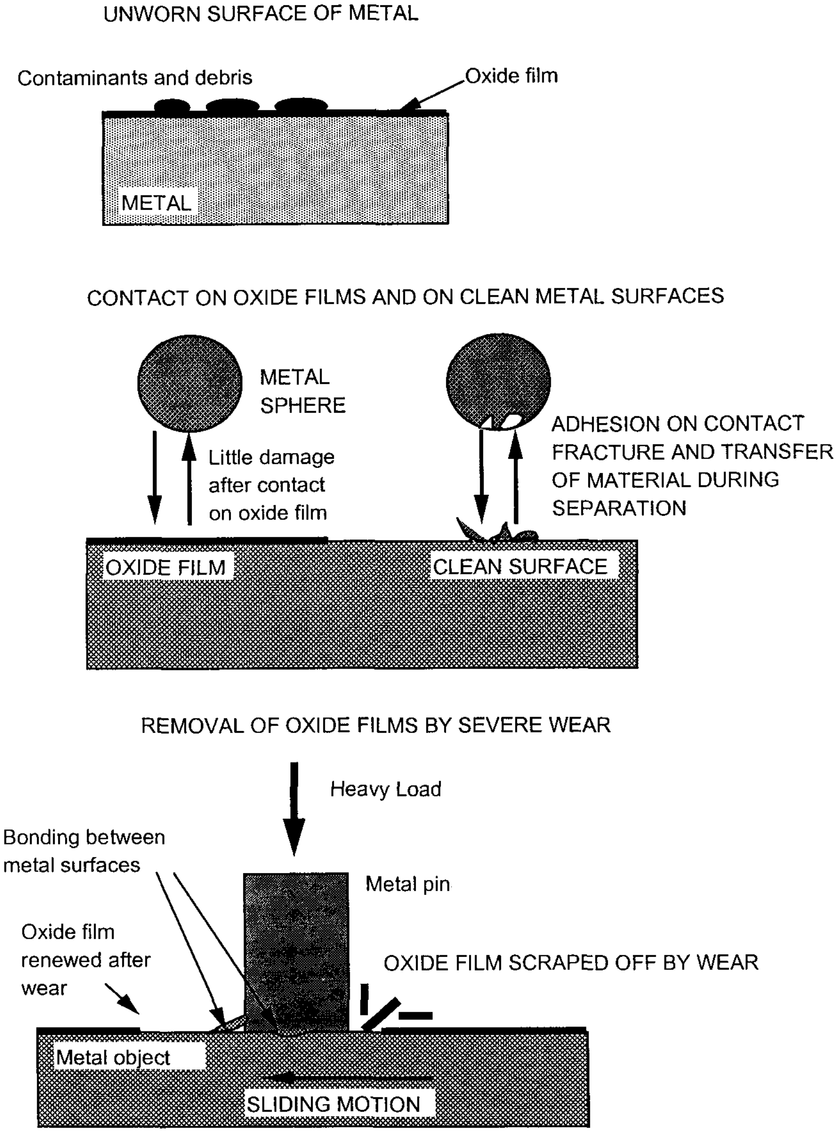


Figure 2.4 Contaminant and oxide layer removal on metals by wearing contact to initiate adhesive wear.

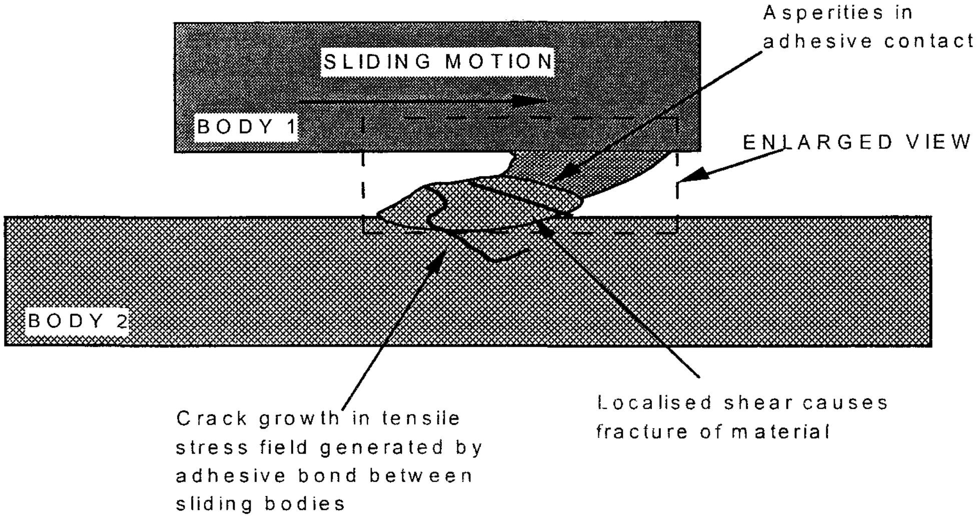
Adhesion between metals is a form of cold welding and in its most severe form can bind the two wearing metals to prevent further relative motion. Atmospheric oxygen fortunately limits adhesive wear considerably but for space vehicles it is a major problem. In one incident the docking assemblies of two spacecrafts that joined in space could not be later separated because of adhesive bonding.

It appeared that the impact of docking was sufficient to cause perforation of residual oxide films on the metal surfaces of the docking assembly. Direct metal contact and unwanted adhesion then resulted.

Under atmospheric conditions, it is sliding not rolling or impact, which initiates adhesive wear. This is because this mode of relative motion is most effective at removing oxide films. Adhesion between sliding bodies has two consequences; severe wear and friction and the formation of transfer layers. Severe wear (i.e. rapid wear with high friction) occurs because sliding motion is only possible during adhesive wear when asperity contacts and adjacent material are torn out of the worn surfaces. A high friction coefficient exceeding unity has also been observed. Wear particles once formed by rupture of an asperity contact have a tendency to partly roll and partly plough their way through the sliding contact. These wear particles also grow in size by the accumulation of further material to cause extensive grooving in the worn surface. In the context of a journal and bush, growth of one single wear particle in the annular space between journal and bush may cause frictional seizure. Transfer layers are composed of wear debris that may or may not have been formed by adhesive wear, which subsequently adheres to the wearing surfaces. When adhesion is absent, wear debris is rapidly expelled from a sliding contact. Transfer layers effectively change the composition of materials in sliding contact and the friction and wear coefficients will also change as a result. Transfer layers are noticeable when a hydrocarbon polymer is made to slide repetitively on a metal surface. The transfer layer is believed to reduce the friction and wear coefficients from high initial value to usefully low values. The processes of severe wear by adhesion and the formation of transfer layers is illustrated schematically in Figure 2.5

The greatest likelihood of adhesive wear occurs between similar metals in dry sliding. Substitution by a dissimilar combination of metals reduces friction and wear. For example, sliding between steel test specimens is more likely to cause adhesive wear than sliding between steel and brass or bronze test specimens. A complex microstructure is also unfavorable to adhesive wear, so that the likelihood of adhesive wear is less with steel than with

ADHESIVE WEAR MECHANISM



FORMATION OF TRANSFER LAYER AND FIELD BELOW WORN SURFACE

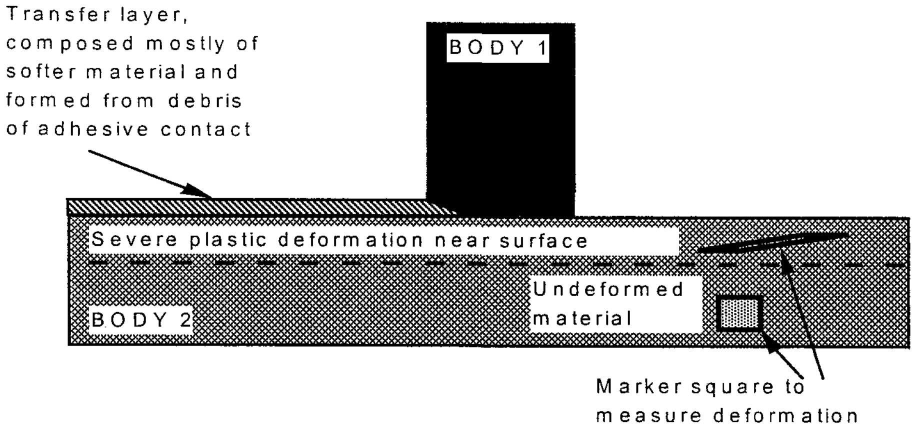


Figure 2.5 Mechanism of severe adhesive wear, asperity deformation and gouging by wear particles. The formation of transfer layers.

pure iron. The reason for the lessening of adhesive wear with increasing complexity of microstructure is that the probability of two identical micro-structural phases coinciding at a contact decreases with the number of phases present in a material. Adhesion tends to be strongest between identical phases so that this probability controls the severity of adhesive wear. Some phases in metal alloys have a non-metallic character when compared to the base metal. For instance, the low adhesion tendency of non-metallic phases such as interstitial carbide or carbon compared to the metallic phases of steel, martensite, austenite, ferrite is a major reason why cast iron is more suitable for dry sliding than plain steel. Micro-structural effects on adhesive wear are illustrated schematically in Figure 2.6.

Non-metals such as ceramics and polymers do not initiate electron exchange so that adhesive wear is far less likely provided that melting temperatures are not reached by frictional heat. This condition imposes severe restrictions on polymers. If surface melting is caused by frictional heating, another form of wear occurs, melting wear, which is discussed later.

High hardness suppresses true contact on a microscopic scale between atoms of opposing surfaces, as the contacting surfaces are unable to deform sufficiently to make contact. Hard materials are therefore less likely to suffer adhesive wear provided that the process of hardening does not change other characteristics of material structure to predispose adhesive wear. Brittleness is also a favorable property since if the material is ductile, severe plastic deformation and the formation of large wear particles occurs during asperity contact. High tensile ductile materials such as nickel base alloys and stainless steels are prone to severe friction and wear for this reason.

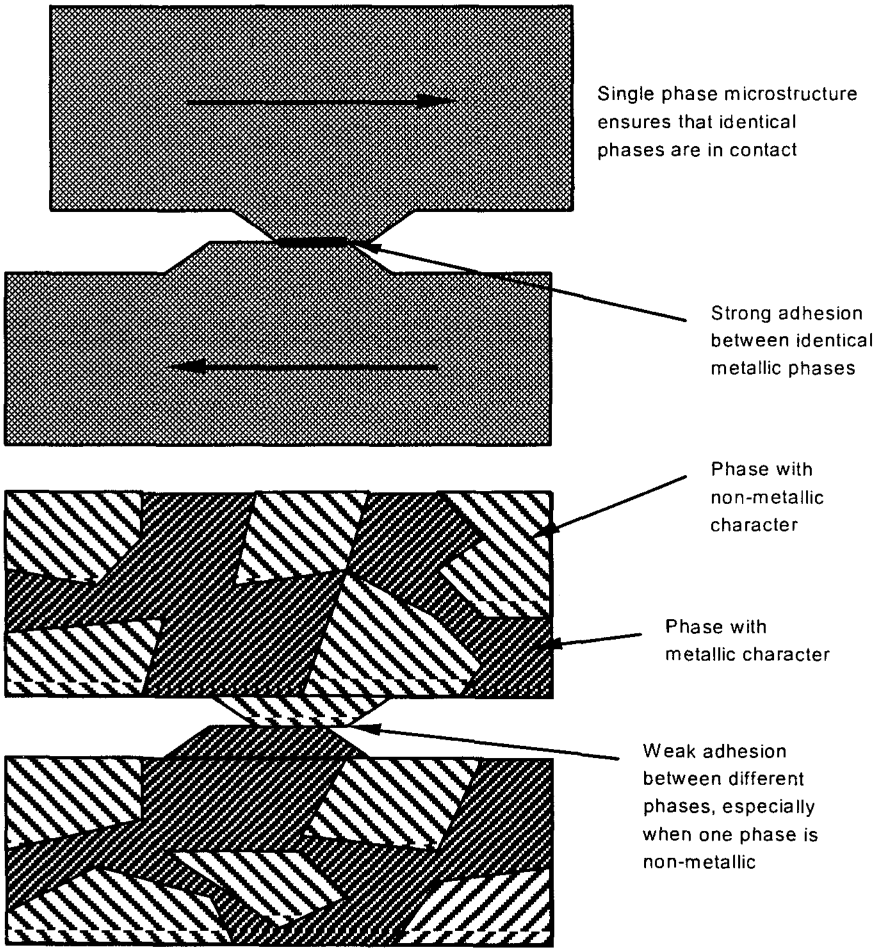


Figure 2.6 Microstructural effects on adhesive wear.

Adhesive wear is usually unacceptable during sustained operation of machinery and is associated with mechanical breakdown. The prevention of adhesive wear is often the primary purpose of the wear resistant coatings to be described.

2.2.2 *Abrasive and erosive wear*

The most economically significant form of wear is caused by abrasive and erosive wear of mineral processing machinery. Extremely rapid wear often occurs at the interface between the metal processing equipment and the rock, which contains the mineral to be extracted. It was originally thought that the rock particles or sand resembled miniature-cutting tools in their action on the worn metal surface. It is now known that this view is simplistic and that abrasive wear involves processes such as fatigue, grain detachment and brittle fracture. Erosive wear is caused by the impact of solid and liquid particles on a surface. Liquid particles can quickly erode even hard metals if they driven against a surface with sufficient velocity. Wear of aircraft fuselage when flying through clouds is a common example of this process.

Abrasive wear

During abrasive wear, apart from some gradual chipping of the abrasive particle and possible adhesive transfer of abraded material to the particle, the abrasive particle remains largely intact for the duration of one passage of the abrasive particle over the worn surface. The particle is hard enough to plastically deform the worn surface and with sufficient contact load will plough a groove in the worn surface. If the particle is sufficiently sharp and suitably orientated to the worn surface, cutting as well as groove formation will occur. In most cases however, the particle is too blunt for cutting and instead a wedge of material builds up in front of the abrasive particle or grit by plastic deformation. This wedge of material eventually detaches by crack propagation to form a wear particle. When the material is particularly brittle, cracks propagate beneath the wear track of the grit to cause rapid wear. Raising hardness at the expense of toughness is therefore largely counterproductive in suppressing abrasive wear. Materials with weak inter-crystalline bonding are liable to wear by removal of entire crystal grains, alumina is susceptible to this form of wear. Repeated passage of many grits over a worn surface produces overlapping wear tracks. At the edge of each wear track there is usually a lip of displaced material, repeated deformation by overlapping wear tracks will cause this lip material to detach by a fatigue process. The mechanisms of abrasive wear are illustrated schematically in Figure 2.7.

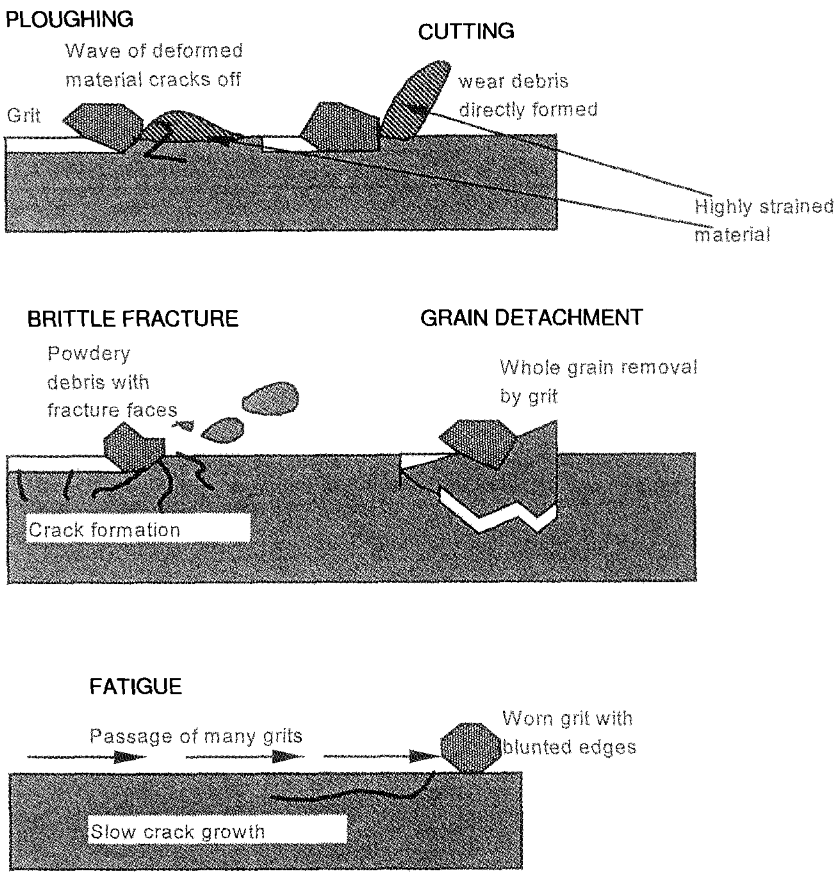


Figure 2.7. Mechanisms of abrasive wear.

Abrasive wear has the virtue of being possibly the easiest of the wear mechanisms to suppress completely if adequate materials are available. Rapid abrasive wear only occurs when the ratio of particle hardness to material hardness is greater than 1.2. When the ratio of particle hardness to material hardness is less than 1.2, abrasive wear declines rapidly but does not cease entirely until the material hardness is greater than the particle hardness by a similar factor of 1.2. It is observed that when the particles have equal hardness to the material, some scratching and wear of the material occurs although damage to the particles by crushing and plastic deformation is extensive. The effect of hardness on abrasive wear is illustrated schematically in Figure 2.8.

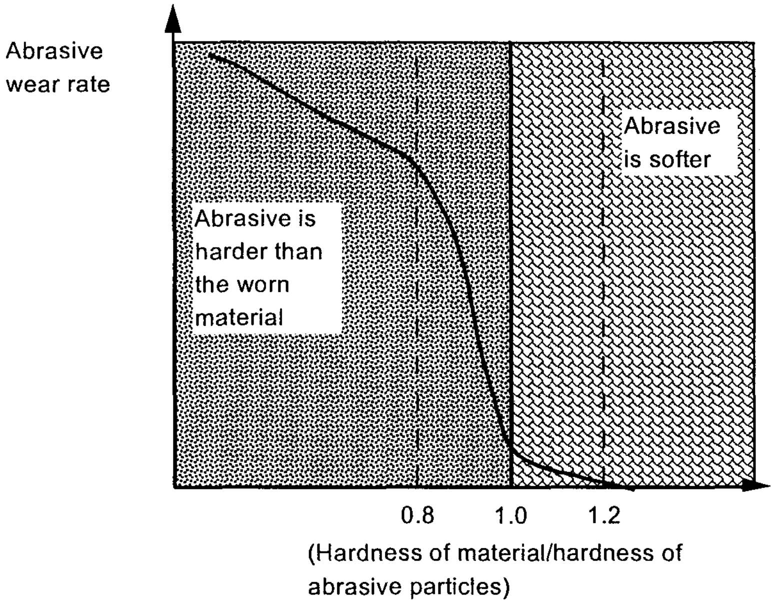


Figure 2.8 Effect of hardness of abrasive grit on abrasive wear rate.

A major practical difficulty to the elimination of abrasive wear from most practical engineering situations is that the most common abrasive is silica with a Vickers hardness (VHN) of 1100 approximately. To prevent abrasive wear by silica, the minimum hardness of the worn material should be 1100/1.2 VHN = 917 VHN. This is unfortunately just a little harder than martensite at 700 VHN so that the abrasive wear of steel implements persists with the absence of cheap hard-coatings. Most hard coatings with a hardness of 1000 VHN approximately are based on tungsten or other expensive materials, which cause the coatings to be comparably expensive.

Where the hardness of the particle or 'grit' is much larger than the hardness of the worn material, a linear proportionality between material hardness and wear rate is observed. For example, when diamond grits, which are the hardest known abrasive, are tested against various metals, it is found that if one metal is twice as hard as another, the former metal will wear half as fast. This rule applies so long as only one class of materials is considered. If plastics are compared with metals, it is found that both materials conform to proportionality between hardness and wear but the constants of proportionality are different for polymers and metals. Another important exception to the rule is that when metals are hardened by alloying or heat treatment, the gains in wear resistance are less than predicted by an exact proportionality between wear and hardness. These characteristics of abrasive wear by hard abrasives are illustrated schematically in Figure 2.9.

Detailed experimentation reveals that abrasive wear is not truly proportional to hardness measured by a hardness tester, which is hardness at very low strains. Abrasive wear is controlled by hardness at extreme strains similar to those that occur in the chip of material, which is extruded in front of an abrasive particle. When hardness at extreme strains is measured, it is found that the hardness of alloys and heat-treated metals becomes very similar to the plain metal. A graph of wear rate versus high-strain hardness shows wear-hardness proportionality for all metals irrespective of heat treatment and alloying.

Resistance to abrasive wear is not a materials property in the same way as tensile strength, since the wear rate is strongly dependent on grit characteristics as well as material strength parameters. Grit *abrasivity*, which is an index of propensity to cause abrasive wear, is strongly dependant on grit shape, hardness and toughness. Grit shape dictates the sharpness of edges on the exterior surface of the grit. Rounded blunt edged grits cause less wear than the same grit material with many sharp edges. Sharp edges on grits facilitate the more rapid cutting mode of abrasive wear and so accelerate wear. Freshly crushed rock which produces grits with surfaces composed of sharp edged fracture faces is likely to cause far more wear on processing equipment than well worn river sand where any sharp edges would have been worn away by repeated contact with other grits. The hardness of the grit dictates the maximum hardness of materials that are prone to severe abrasion as discussed above. Toughness of grits controls the formation of new sharp edged fracture faces on the grit surface. During abrasive wear, grits progressively reduce in size by fracture. Size reduction in grits is particularly common during high stress abrasive wear. There appears a critical toughness which results in maximum wear rate since this toughness represents the

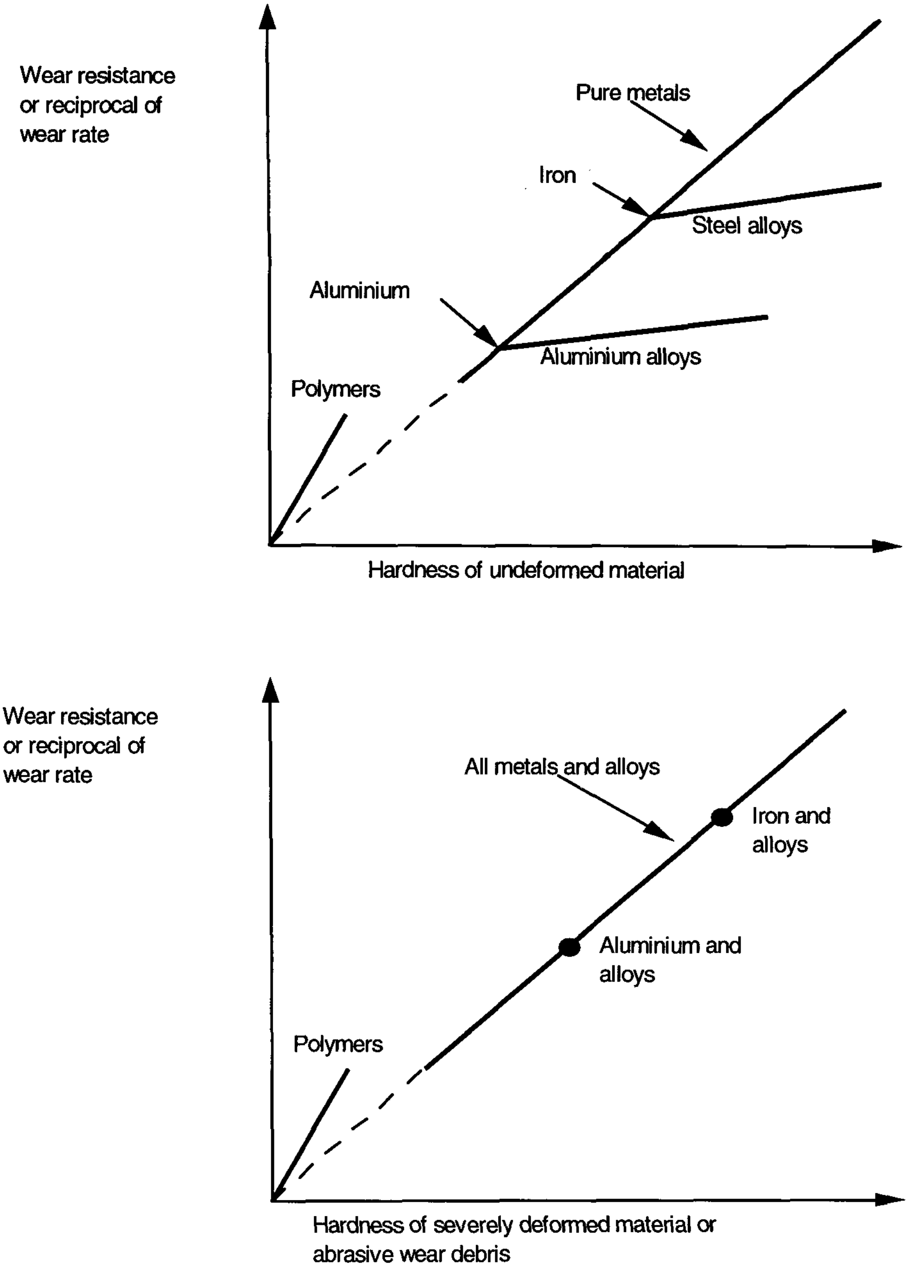


Figure 2.9. The relationship between wear rates with hard abrasives and the hardness of the worn material.

optimum balance between formation of new sharp edges and the survival of the grit. Very brittle grits rapidly degrade to a fine powder while very tough grits change to a rounded profile by gradual wear of the grit. The role of grit shape and grit toughness in controlling abrasive wear is shown schematically in Figure 2.10

Figure 2.10 Role of grit shape and toughness of grit material in abrasive wear.

Erosive wear

The mechanisms of erosive wear are not constant but are controlled by the angle of impingement of a particle, its speed, its size and the phase of material that constitutes the particles. Erosive wear mechanisms are different when a direct or normal impact is substituted by a glancing impact, the speed of impact controls the energy dissipation level and the temperature rise at impact while erosive wear by high speed droplets of liquid differ from erosion by solid particles. The angle of impingement and the other listed parameters are illustrated schematically in Figure 2.11.

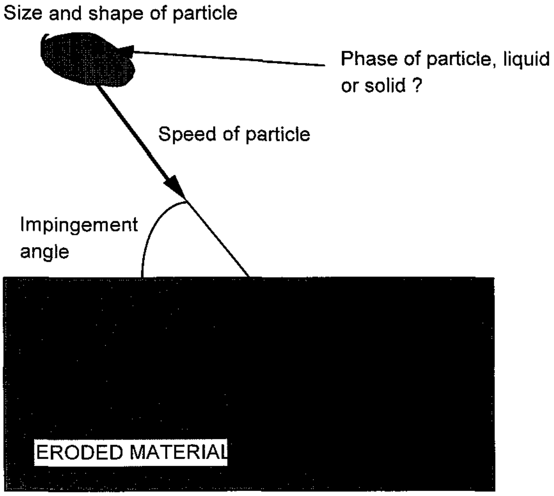


Figure 2.11. Physical and material parameters controlling erosive wear

Erosive wear can resemble abrasive wear when hard solid particles of microscopically visible size are the eroding agent and the angle of impingement is low and the impingement speed is of the order of 100 m/s. For all other conditions, wear mechanisms not resembling abrasive wear become dominant. For particles of microscopically visible size and an impingement speed of the order of 100 m/s, wear at high impingement angles proceeds by a combination of plastic deformation and fatigue or by cracking for brittle materials. For very large particles or objects moving at very high speeds, superplastic flow and melting in the impact zone becomes the controlling factor, while for very small particles or free atoms, wear proceeds by dislodgement of atoms from the crystal lattice. Superplastic flow is extremely rapid plastic deformation, which resembles the flow of a liquid. These latter two forms of erosive wear have few terrestrial examples but are highly significant to space technology. When a satellite is located in what is called Low Earth Orbit (LEO), it is within the reach of atmospheric gases. The satellite has to maintain a high speed in order to remain in orbit, which results in continuous impingement by ions originating from the atmosphere. Impingement by oxygen ions is particularly harmful as it causes oxidation of the worn material as well as destruction of its crystalline lattice. Planetary erosion by meteorites is the prime example of erosive wear by high speed large size objects as meteorites can impact on a planet with enormously high speeds. Practical significance of this mode of wear is at present quite limited, but it is conceivable that in the future, meteorite impact may be engineered to expose a valuable subsurface deposit of minerals on an uninhabited planet. The discussed mechanisms of erosive wear are illustrated schematically in Figure 2.12

With small but not atomic particles at moderate speeds, a distinction is observed between what is called the *ductile mode* of erosive wear (at moderate particle speeds) and the *brittle mode* of erosive wear is reflected in the response of a worn material to the angle of impingement. The ductile mode of erosive wear is dependant on tangential movement of the impinging particle across the eroded surface for a wear particle to be produced by shearing or cutting. This tangential movement is favored by a very small angle of impingement as opposed to normal impact. If the angle of impingement is zero however, there will be no kinetic energy of impact between particle and worn surface to initiate indentation by the particle and chip formation. A maximum in the ductile mode of erosive wear is usually found close to an angle of 30°, which represents a compromise between the requirements of tangential particle movement and impact energy. For the brittle mode of erosive wear, the maximum in erosive wear occurs around 90° where the kinetic energy of impact is at a maximum. The distinction in wear rates as a function of impingement angle for the brittle and ductile modes of erosive wear is shown in schematically in Figure 2.13.

Erosive wear by liquid particles is mostly limited to the brittle mode and the rate of wear is far more sensitive to particle velocity than when solid particles are involved. For liquids, erosive wear is observed to be proportional to approximately the 5th power of impact velocity while for solids wear is proportional to the square of impact velocity approximately.

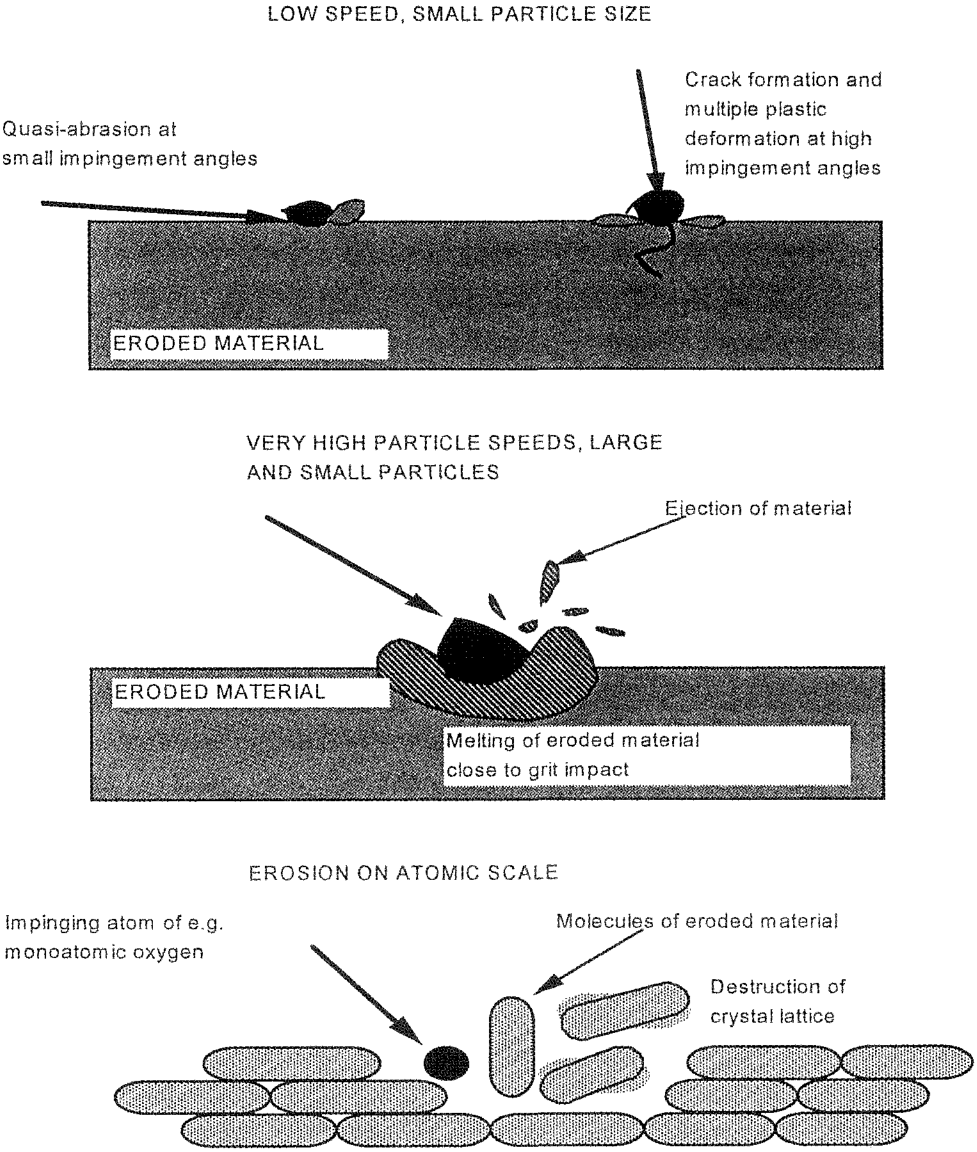


Figure 2.12 . Mechanisms of erosive wear for varying particle speeds, sizes and impingement angles.

Materials with high toughness are usually the most effective at resisting erosive wear as hard materials are often too brittle to be wear resistant. The size of particle size is critical to wear resistance as small particles, e.g. 8-μm average size tends to erode by ductile shearing of the worn material. This characteristic is due to the fact

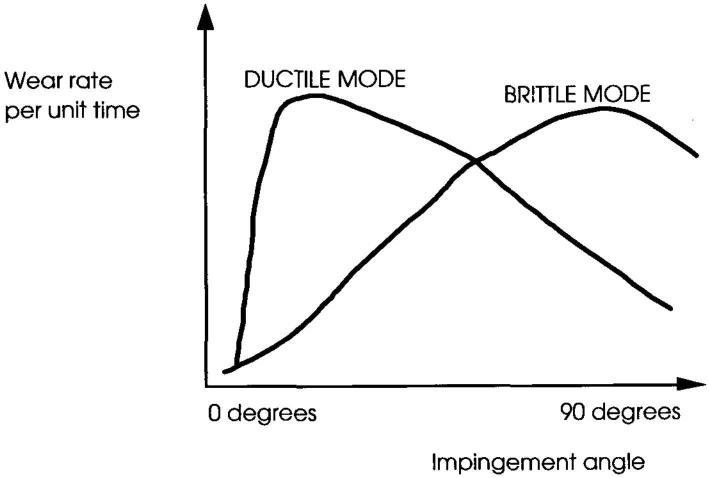


Figure 2.13. Contrast in dependence on impingement angle for brittle and ductile modes of erosive wear.

that brittleness depends on the presence of at least one defect in the deformed volume. Since the deformed volume decreases in proportion to particle size, erosive wear at small particle sizes is more likely to be ductile than erosion at large particle sizes where the apparent brittleness of the worn material is higher. High hardness materials are therefore more likely to be effective in resisting erosive wear by small particles whereas tough materials are suitable for larger particles of e.g. 50-μm size and larger. Rubber is durable against erosive wear for particles below a critical kinetic energy. When the particle kinetic energy is low, the rubber is able to deform and rebound during particle impact. Beyond the critical kinetic energy, particles are able to drill holes into the rubber and rapid wear ensues.

Another type of wear that is similar but not identical to erosive wear is cavitational wear. Cavitation is the result of negative pressure developed on the downstream faces of bodies immersed in a flow of liquid. Within the region of negative pressure, bubbles of dissolved gases and vapor successively form and collapse. When these bubbles collapse on the downstream face of the immersed body, wear is caused by the hammering effect of repetitive bubble collapse on a hard surface. A slow form of wear involving the formation of holes by localized crack development then results. It is found that high resistance to mechanical fatigue correlates with resistance to cavitational wear. Cavitation wear can be destructive to propellers and water turbines, but is useful in medicine. Kidney stones can be removed without surgery by using ultrasonic sound waves to induce cavitation around kidney stones in a patient's body. The kidney stones disintegrate and the debris is later expelled by urination. The mechanism of cavitation wear is shown schematically in Figure 2.14.

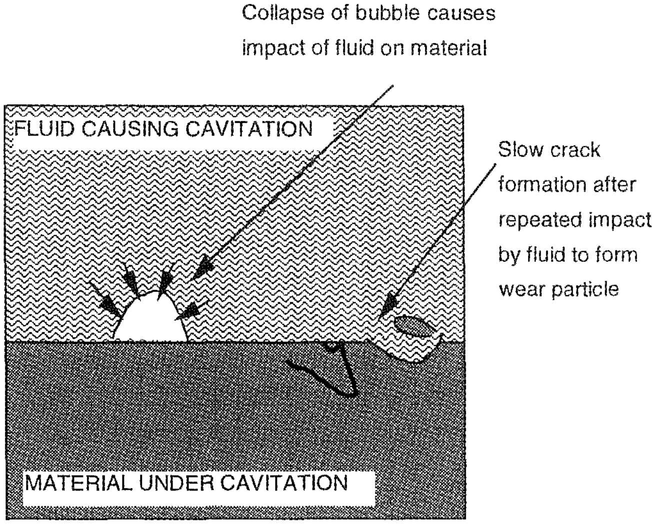


Figure 2.14 Mechanism of cavitation wear.

2.2.3 *Wear induced by mechanical fatigue of the worn surface*

Wear is generally observed to be a gradual process involving repetitive passage of one solid body over another. The number of cycles of relative movement between two wearing bodies can easily reach one million or more. Common examples of such cyclic relative movement are found in rolling and sliding bearings. Even when adhesive wear is suppressed by effective lubrication or surface coatings, or where sliding is substituted by rolling, fatigue processes will occur on the contacting surfaces. When two bodies make sliding contact without strong adhesion, a surface wave is developed between the asperities. The wave moves across the sliding surface in a manner analogous to a rack in a carpet that is driven across the carpet. The wave is generated by large strains in the surface layers of the worn material. When repeated sliding contact and wave formation occurs, these strains lead to fatigue wear. Wave generation at the surface of a wearing solid is illustrated schematically in Figure 2.15.

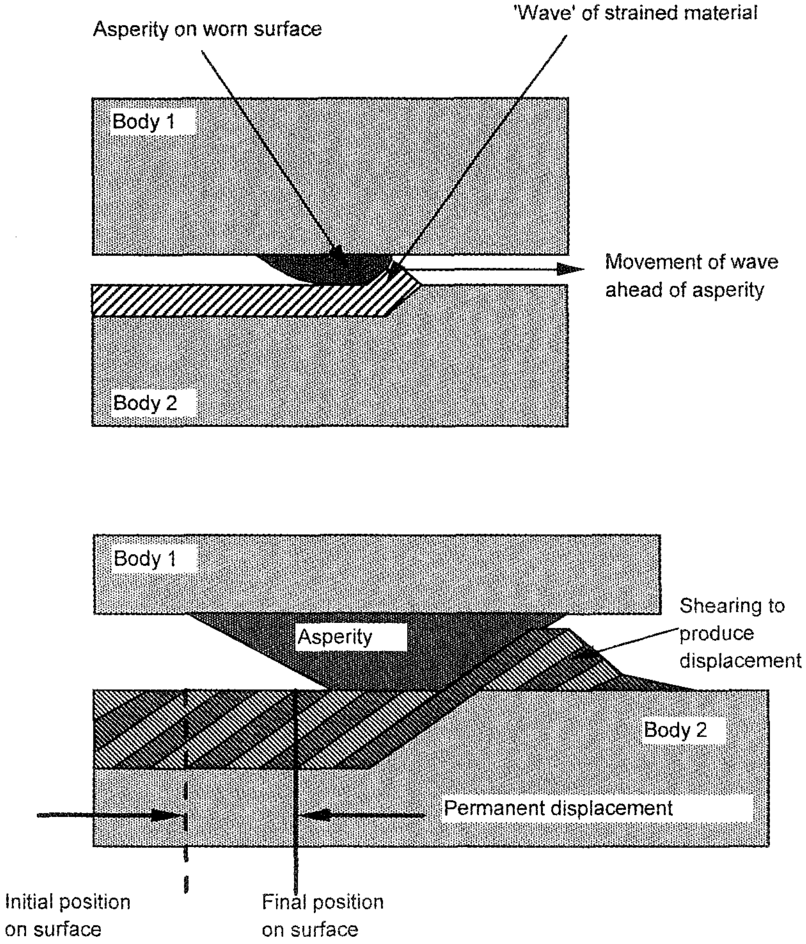


Figure 2.15. Wave formation at the worn surface of a material.

A fundamental characteristic of contact between solids that is predicted by the theory of Hertz and later workers is that a maximum in shear stress occurs below the contact interface unless the friction coefficient is high. At the contact interface, stress is high but it is mostly hydrostatic in nature unless the friction coefficient is high enough to impose a maximum in shear stress at the surface. When a body is subjected to repetitive rolling or sliding contact, material at a depth of typically a few micrometers below the contact interface is subjected to cyclic shear stresses. Crack propagation may then result to form a wear particle. This mechanism of wear is basically a form of subsurface crack propagation to produce lamellar wear particles and is known as delamination. It is possible for the subsurface crack to reveal few signs of wear on the surface prior to release of a wear particle. Sliding as opposed to rolling often causes the formation of severely strained layers directly beneath the worn surface. When there is sufficient frictional force, surface material is stretched tangentially to the surface in the direction of sliding. These strained layers are vulnerable to crack formation and consequent release of wear particles. The cracks usually propagate from the surface along fissures generated by the extreme surface strain and a more lumpy type of particle is usually generated than with the subsurface or delamination mechanism. The subsurface and surface mechanisms of crack propagation to cause wear are illustrated in Figure 2.16.

Oxidation of the worn surface prevents re-adhesion between adjacent folds of the highly strained surface layers thus facilitating the propagation of cracks along the interfaces between the folds.

Critical material properties relating to the prevention of delamination wear appear to be the lack of defects to initiate crack formation. The precursors of cracks, dislocations are trapped by inclusions and when sufficient number of dislocations is collected, a crack is initiated at the inclusion. Material purity, e.g. lack of sulfur in steels, is therefore critical to the prevention of delamination wear. For surface originated fatigue wear, resistance to crack growth appears critical, as even very hard materials when

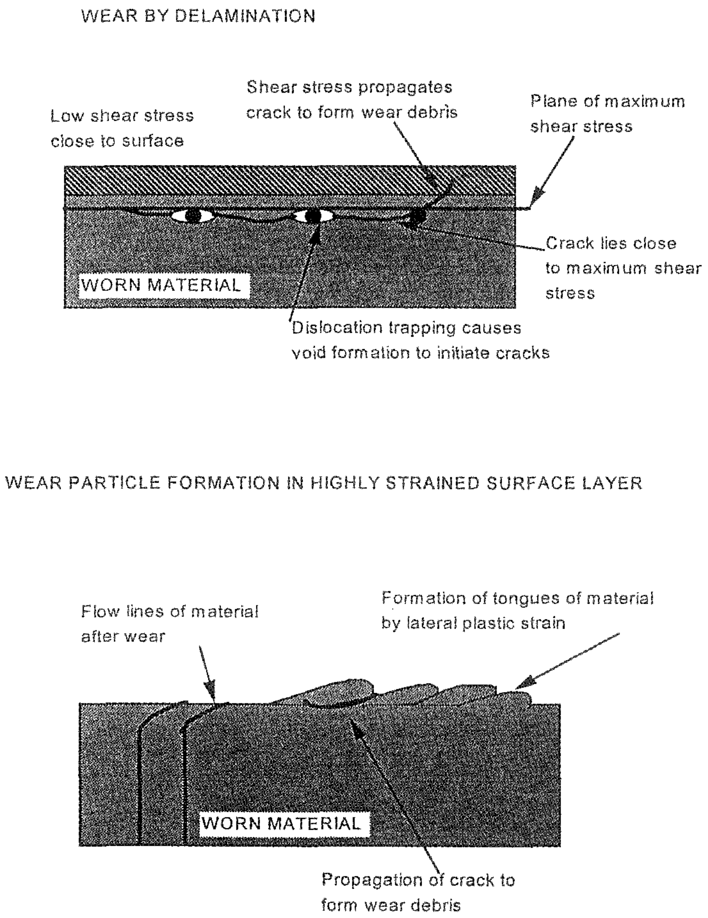


Figure 2.16. Mechanisms of wear particle formation by delamination and crack formation from surface strained layers.

slid against each other will form highly strained surface layers. High fatigue resistance is therefore beneficial to wear resistance in this instance.

2.2.4 *Melting wear, fretting wear and diffusive wear*

Melting wear and fretting wear are forms of wear that prevail at what may be considered the margins of engineering application. With a few important exceptions such as the moment when a shell is fired from a gun barrel, most practical sliding contacts do not slide at sufficient speed for melting wear of metals to occur. Fretting wear is generally associated with parasitic movement within a nominally static interface such as interference fit between a shaft and bush. Melting wear and fretting wear are however significant wear mechanisms that cannot be overlooked in any discussion of wear. Diffusive wear is usually found at the interface between a cutting tool and chip of cut material where it causes eventual tool failure. Diffusive wear has been likened to the dissolution of sugar in hot water but requires considerable mechanical energy for its activation.

Melting wear

When two bodies slide against each other at high speed, the frictional temperature rise at the sliding interface can reach the melting point of the more fusible of the two materials in contact. Frictional heating is characterized by extreme localization of heat to a depth of only a few micrometers below the worn surfaces. It is therefore possible for a layer of molten material to form on the worn surface with only negligible heating of the material below.

Melting prevents severe adhesive wear from occurring because although inter-asperity adhesion is strong, the fracture strength of the asperity contact is greatly reduced. The molten layers are however easily expelled from the wearing contact to promote wear. Friction is therefore reduced by the melting of asperity contacts but wear is usually more rapid than wear without surface melting. The mechanism of melting wear is illustrated schematically in Figure 2.17.

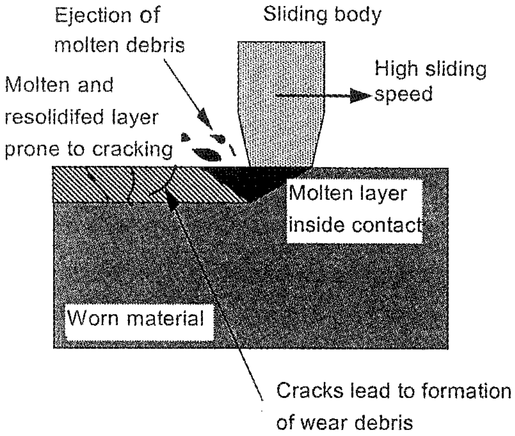
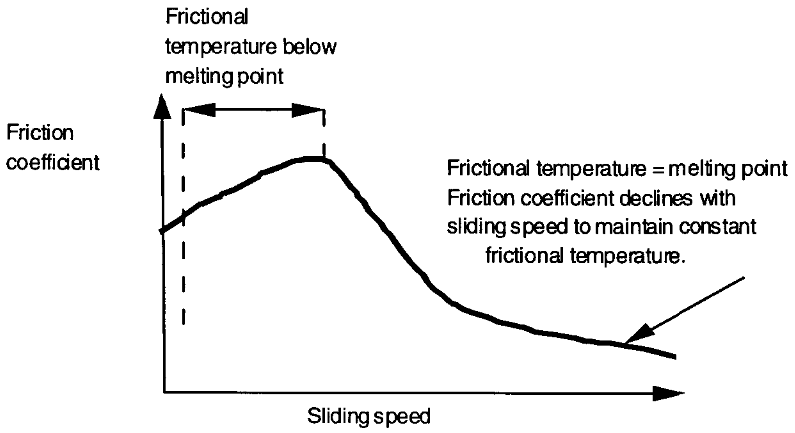


Figure 2.17. Mechanism of melting wear.

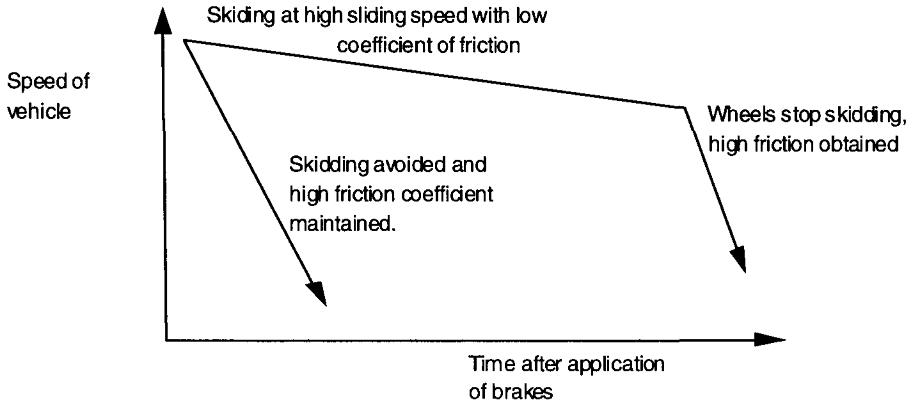
Melting wear has two characteristics of general importance; thermal control of friction and a strong interaction with atmospheric oxygen. Thermal control of friction is a condition where a limiting frictional surface temperature rise dictates an inverse relationship between sliding speed and the coefficient of friction. This phenomenon occurs because as stated above, it is observed that the maximum temperature in a sliding contact never exceeds the lower melting point of the two materials in contact. The high latent heat of fusion of most materials is probably the reason for this characteristic. At high sliding speeds, the frictional temperature rise is proportional to the product of the friction coefficient and the square root of sliding velocity.

For a constant temperature rise, it follows that the friction coefficient will be inversely proportional to the square root of velocity. This relationship between friction and velocity is seen at moderate sliding speeds of the order of 10 m/s for polymers, which have low melting points, and at higher speeds of 100 m/s or more for metals. An important practical application of this phenomenon is high speed skidding of road tires. Such skidding will not only cause very rapid wear of the tire but also ensure an almost total loss of frictional grip to the road compared to low speed skidding. The relationship between sliding speed of rubber and its coefficient of friction against a hard surface is illustrated graphically in Figure 2.18. A graphical illustration of the effect of skidding by rubber tires on the braking distance and wear of the tires is also shown in Figure 2.18.

The melting of surface material by frictional heat may cause surface oxidation if sliding of a reactive material occurs under atmospheric conditions. For steels, a hard surface layer known as 'white layer' because of its non-etching characteristic is observed to form. This



SPEED DEPENDANCE OF RUBBER FRICTION



EFFECT OF SLIDING (SKIDDING) SPEED ON BRAKING BY A VEHICLE

Figure 2.18 Decline in friction with sliding speed of rubber caused by melting wear and its effect on skidding by rubber tyres.

white layer is believed to be composed of rapidly quenched steel and iron oxide, which have been blended into a finely divided mixture by sliding contact. The formation and structure of white layers is illustrated schematically in Figure 2.19.

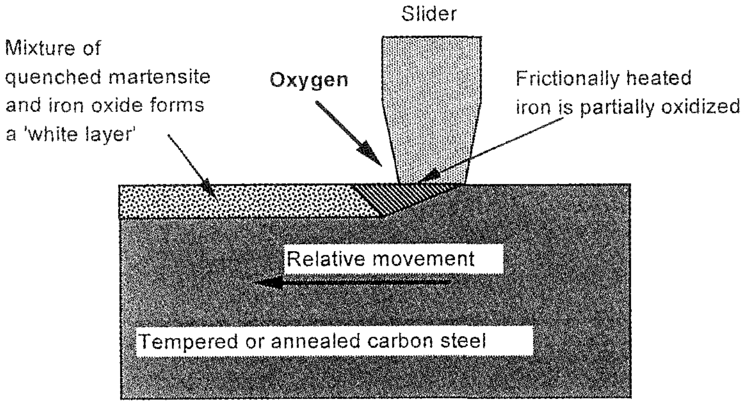


Figure 2.19. Structure and formation of white layers by melting wear.

The formation of white layers is believed to be associated with increased wear resistance of the steel and possibility of deliberately forming white layers has been invoked on more than one occasion.

Fretting wear

Accumulated observations over a long period of time have revealed that in practice it is very difficult if not impossible to entirely prevent sliding motion in any contact that is subject to varying load. While it is not difficult to stop gross movement from occurring, short reciprocating movements with amplitude of a few micrometers appear to be impossible to prevent without the use of an adhesive in the contact interface. The severity of wear is not diminished by the short amplitude of sliding involved and fretting wear is involved in the premature fatigue failure of shafts at the edges of interference fits with bushes and of wire ropes at the interfaces between individual wires of the rope. These sites of fretting wear are illustrated schematically in Figure 2.20.

In wire ropes, the wires have to move by several micrometers between each other to allow the rope to flex. Wear and wire damage occurs mostly during contact with a sheave where the combination of flexing and contact forces between sheave and rope

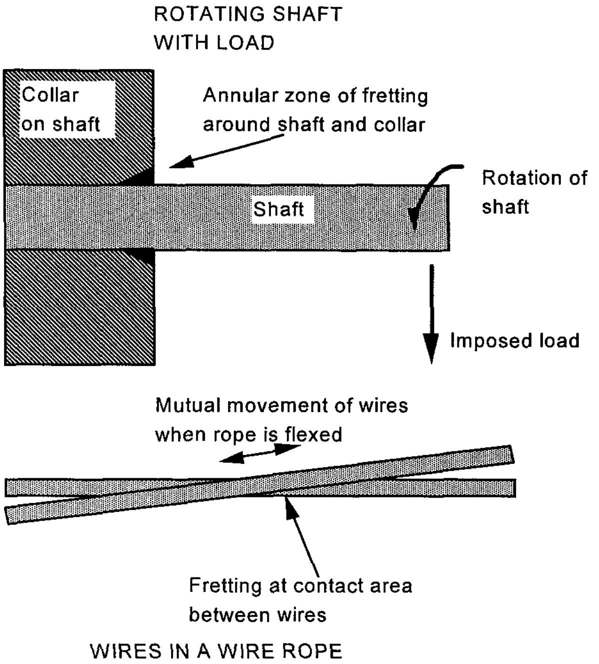
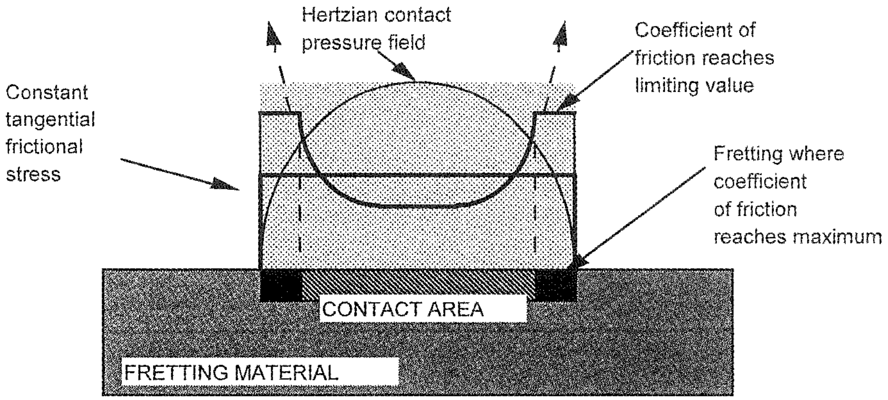


Figure 2.20. Location of fretting wear in some common engineering components.

results in heavily loaded sliding between individual wires. A basic analysis of the reason for micro sliding in a nominally stationary contact is attributed to Mindlin and in summarized form is as follows.

Any contact between rough surfaces can be modeled as a series of discrete Hertzian contacts. A fundamental characteristic of a Hertzian contact is that the contact pressure declines in a continuous manner from a central maximum to zero at the margins of the contact. If an ideal contact where no sliding under frictional force is considered, when a frictional force is applied to the Hertzian contact, a uniform tangential stress is established over the Hertzian contact area. The local friction coefficient at any point in the contact area is equal to the tangential stress divided by the local contact pressure. The friction coefficient therefore rises from a finite value at the center of the contact to an infinite value at the margin of the contact. This result is clearly impossible and Mindlin predicted that an annular region of slip would surround a central region in the contact where the local friction coefficient did not exceed a critical value. K.L. Johnson and co-workers later confirmed Mindlin's analysis experimentally. Mindlin's analysis is illustrated in Figure 2.21 graphically for the simplest case where a circular Hertzian contact is subjected to a frictional torque.



MECHANICS OF FRETTING IN A CONTACT

SHAPE OF WEAR SCAR DURING FRETTING

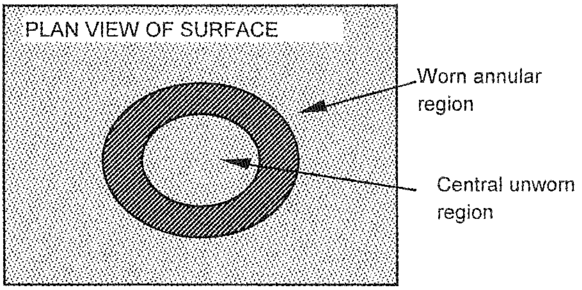


Figure 2.21. Origins of microsliding in a Hertzian contact that sustains frictional force.

Such localized micro sliding, which is typically of the order of one micrometer, occurs in interference fits between e.g. shaft and bush or bolt and washer. When the frictional force reaches a sufficiently high value, the central non-sliding region in the Hertzian contact vanishes and uniform sliding begins. Fretting wear at amplitudes of tens of micrometers then occurs as is found in the interface between roping wires. While the damage and wear in a small number of cycles is quite negligible, many engineering components reach one million cycles of reciprocating micro sliding during their service lifetime which leads to significant damage and wear by fretting. Fretting is found to accelerate mechanical fatigue by the formation of surface cracks around wear scars. These cracks then propagate to cause fracture.

Fretting wear involves basic wear processes such as adhesive wear and fatigue based wear but is different from wear as normally seen in unrestricted sliding. The small amplitude of sliding compared to the dimensions of the sliding contact causes wear debris to become trapped within the sliding interface. A layer of wear debris progressively forms on the worn surface and wear becomes a function of both the formation of new wear debris and the later release of wear debris from the trapped layer. For reasons that are still not fully understood, the presence of trapped debris layer affects friction coefficients and the relative wear resistance of materials to fretting. The effect of small amplitude of sliding on wear and the formation of a wear debris layer in fretting is illustrated schematically in Figure 2.22.

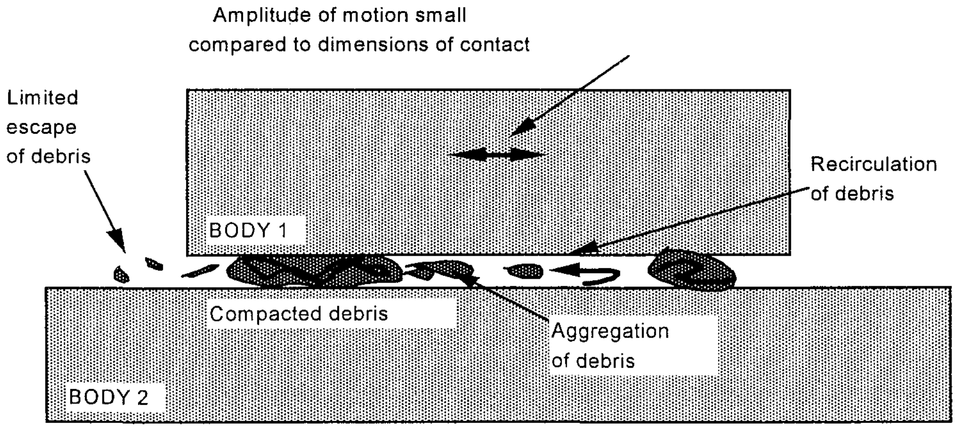


Figure 2.22. Formation of wear debris layer by entrapment in restricted amplitude sliding.

Materials that offer usefully high wear resistance in unrestricted sliding may fail prematurely in fretting. Wear-resistant coatings contain many examples of this characteristic. If adhesion between asperity contacts occurs then cracking may occur adjacent to the asperity contact, which remains intact during micro sliding. Such cracking is the cause of accelerated mechanical fatigue at even very low amplitudes of fretting movement. Fretting remains one of the more poorly controlled types of wear that is implicated in the unexpected failure of many critical machine components.

Diffusive wear

Diffusive wear occurs at high contact stresses and temperatures where two materials are in intimate contact and diffusion processes are rapid as compared to low temperatures. If the contact stress is sufficiently high, then the true contact area becomes comparable or equal to the nominal contact area since the asperities on at least one of the surfaces become completely flattened by high load. If a very hard material is in contact with a softer material, then only the asperities on the softer material will be flattened and made to conform to the opposing surface. Whatever is the hardness, the contact model of isolated asperity contacts are randomly distributed over a contacting interface and is no longer appropriate and near perfect contact between the opposing surfaces can be assumed.

These conditions of near perfect contact at high load are usually found on the surface of a cutting tool immediately adjacent to the cutting edge where contact stresses reach the yield stress of the softer material and temperatures can reach 1000°C. If one of the contacting bodies is far larger than the other or else supplied as a continuous flow of material, then it is possible for atoms from the smaller body to be continuously dissolved into the flow of opposing material. For cutting tools, strengthening elements in the metal alloy that constitutes the tool are lost by dissolution into the cut metal. The weakened alloy that remains is then rapidly worn away to expose fresh alloy. As well as direct dissolution, chemical reaction is also possible between the contacting materials. Examples of diffusion wear are: tungsten carbide-cobalt composite tools cutting steel or titanium at high speeds, silicon nitride cutting tools for steel. Aluminum oxide and aluminum oxide composites are suitable for cutting steel even though aluminum oxide can react with iron oxide (present on all steel surfaces) to form iron aluminate. The mechanism of diffusive wear is illustrated schematically in Figure 2.23

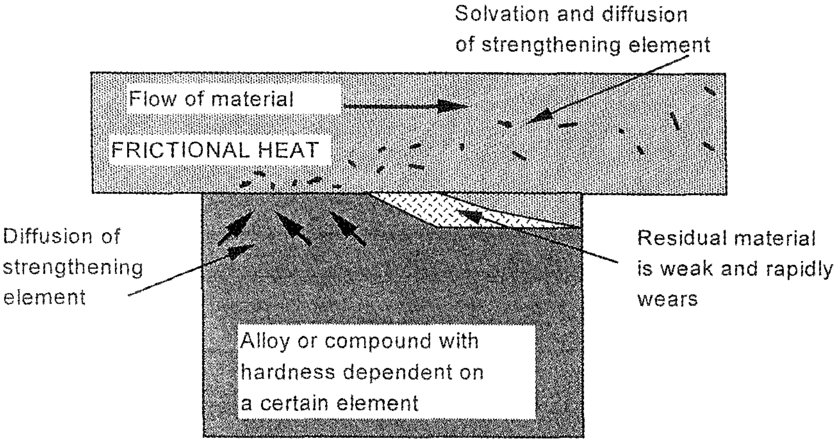


Figure 2.23 Schematic illustration of the mechanism of diffusive wear.

Many cutting tools depend on carbides for their hardness and it is found when cutting steel that carbides tend to react with the steel to form complex carbides in the steel swarf. According to the model of wear developed by Loladze, the transfer of carbide hardening phase to the steel swarf causes rapid wear of the softer metal matrix in the metal-carbide composite. Elemental carbon is also reactive and prone to diffusive wear. It is found that diamond tools display severe wear despite their extreme hardness when cutting steel or iron. This is because carbon is soluble in iron and diffusive wear is therefore rapid. Diamond displays little wear when cutting non-ferrous metals such as zinc and aluminum because of the low solubility of carbon in these metals.

2.2.5 *Analytical models of wear*

The complexity of wear with its multitude of causes defies interpretation by an exact analytical model. An approximate model was developed by Archard several decades previously and has remained a popular tool for interpreting the severity of wear. The basic principal of the Archard model is that wear is proportional to the true area of contact between asperities of the opposing surfaces and the sliding distance. This principle is based on the idea that the size of wear debris must be proportional to the area of asperity contacts. In the simplest form of the Archard model, a single circular asperity contact is imagined to generate a hemispherical wear particle with the same radius as the radius of the contact area. The total volume of wear debris is equal to product of the sliding distance and projected area of the hemisphere when the sliding distance is much greater than the radius of the hemisphere. The projected area of the hemisphere is equal to half the true contact area; this observation forms the basis of the Archard model. This model of wear is illustrated schematically in Figure 2.24.

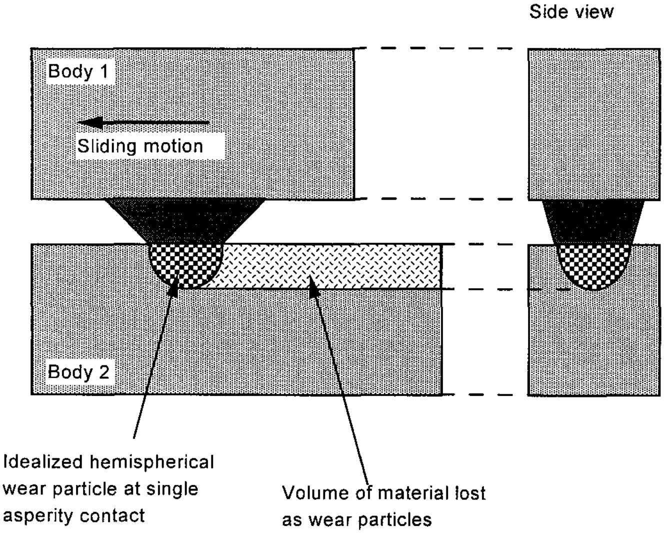


Figure 2. 24 The Archard model of wear.

The true area of contact is independent of the nominal contact area and is approximately equal to the contact load divided by the flow stress of the material. The flow stress is typically 1/3 of the measured hardness. A constant known as the Archard constant is then introduced, it is typically referred to as 'K'. The equation of wear then becomes.

$$V = 3*W*D*K/H \qquad \{2.1\}$$

Where V is the wear volume [m^3], W is the contact load [N], D is the sliding distance [m], K is the Archard constant [dimensionless] and H is the hardness [Pa]. An expression for K can then be derived which is:

$$K = V^*H/(3^*W^*D) \qquad \{2.2\}$$

A critical advantage of the Archard model over other models of wear is that values of 'K' can be determined experimentally. It is found by experiment that typically $K<0.001$ which lends credibility to the fatigue models of wear discussed above where a large number of asperity contacts precede detachment of a wear particle. In severe abrasive wear by hard sharp abrasives, the value of the Archard constant might reach 0.01 whereas in mild wear or wear under lubricated conditions, values of the constant as low as 1.0E-6 can be reached. For adhesive wear, $K = 0.001$ is typical. The Archard constant is extremely useful for crudely estimating a scale of wear to within a factor of 10 ~ 100 by using literature values of the constant and applying them to a particular situation. This method of calculation can become grossly inaccurate when the wrong wear mechanism is assumed, e.g., adhesive wear instead of fatigue wear.

The reader may note an apparent inconsistency between Archard's emphasis on material hardness and the general observation that hardness has only a weak influence on wear excluding abrasive wear. This observation is true for small variations in hardness while Archard's model deals with very large variations in hardness between e.g. silicon carbide and a polymer. Archard's model is also only approximately correct and cannot be used to predict small variations in wear as a function of hardness.

2.2.6 Wear resistant materials

The different mechanisms of wear are each inhibited by different properties and characteristics of materials. To obtain a high level of wear resistance, the specific wear mechanism should be considered.

Abrasive, erosive and cavitation wear

For resistance to abrasive and erosive wear, hardness without excessive brittleness is critical. For erosive wear, brittleness should be more carefully controlled than for abrasive wear in most instances. Hard metals, e.g. martensitic steels have moderate abrasive wear resistance. Higher levels of abrasive wear resistance can be obtained from ceramics such as alumina, zirconia toughened alumina or silicon carbide. If the abrasion or erosion is associated with impact sufficient to cause spalling of ceramics, it will be necessary to accept a higher rate of abrasive/erosive wear and select hard metals.

For cavitation wear, fatigue and corrosion resistance are the controlling properties since the wear process is gradual and occurs in a liquid, which is usually water. Stainless steels, titanium alloys and cobalt alloys e.g. stellite are the most resistant. Ceramics and composite materials where a hard phase is embedded in a soft phase are unsuitable because cavitation selectively attacks weaker phases and microstructural flaws that are commonly found in ceramics.

Adhesive wear

Choosing a material with a complex microstructure that is either non-metallic in character or contains non-metallic phases best controls adhesive wear. A high hardness is also desirable in at least one of the sliding surfaces to suppress adhesive contact. Grey cast irons are more resistant to adhesive wear than steel because of the interstitial graphite, which is non-metallic and acts as a solid lubricant. Ductility is undesirable for adhesive wear resistance because it promotes tensile adhesive contact between asperities. Ceramics are resistant to adhesive wear for this reason. Low chemical activity in metals is also desirable to suppress electron transfer between contacting surfaces. Silver and gold are useful in suppressive adhesive wear because of chemical inertness despite their comparatively high ductility. Aluminum and titanium show very poor resistance to adhesive wear because of high chemical activity and ductility. Proneness to adhesive wear combined with low hardness restricts aluminum to a very low wear resistance in almost all instances.

Fatigue-based wear

The critical feature for resistance to fatigue-based wear is a clean microstructure with the absolute minimum of brittle incoherent inclusions and bubbles or voids. A high quality microstructure in a common material, e.g. steel can generate as much wear resistance as selecting a nominally more advanced or precious material. Increases in hardness are not usually useful in controlling fatigue-based wear because of possible embrittlement and acceleration of crack growth. A tough yet hard material such as a high strength cobalt alloy is resistant to fatigue-based wear.

Minor wear modes

There is a considerable variety of wear modes, the most significant forms of wear, fretting, melting wear and diffusive wear each have radically different wear resistant materials. Fretting wear appears to be suppressed by a combination of toughness and hardness, rather than hardness or toughness acting alone. While it is difficult to identify materials, which show universally high fretting resistance, it is easier to recognize those materials, which are prone to fretting. These include aluminum, titanium and zirconium alloys where these metals are the bulk component. Steels have moderate to good fretting wear resistance yet cannot be said to be entirely immune to damage by fretting. Oxide ceramics such as alumina can display much better resistance to fretting than steels.

Materials with high resistance to melting wear are those with high melting points, e.g. refractory metals and ceramics since melting wear only becomes significant when frictional temperatures reach the melting point of a material.

Diffusive wear of e.g. cutting tools is best suppressed by avoiding those combinations of cutting tool and work piece materials, which are known to generate severe diffusive wear problems.

2.3 Fatigue, fracture and creep

Ductile and brittle fracture, Fatigue and Creep are the primary causes of structural failure of materials. Ductile or brittle fracture usually occurs immediately after a high load is applied to a structure. Fatigue is a mechanism of slow crack growth, which depends on exposure to oscillating or variable stresses. Fatigue occurs at lower loads and stresses than immediate ductile or brittle fracture. Creep involves the continuous extension or stretching of a material until it finally fractures. Creep is commonly associated with high temperatures but many materials suffer from creep at low temperatures. The relationship between fracture, fatigue and creep is summarized schematically in Figure 2.25.

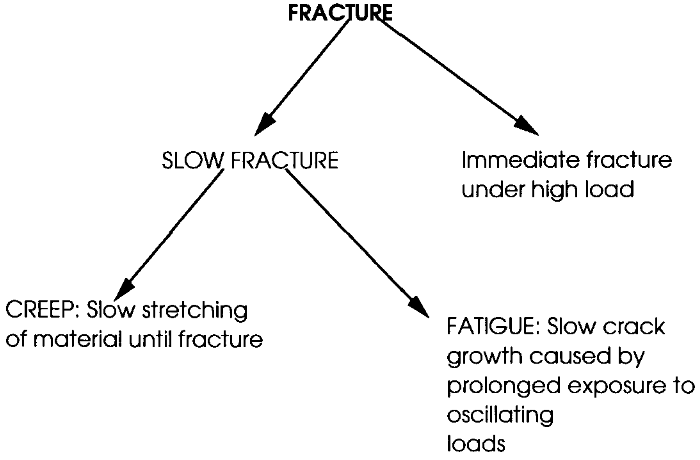


Figure 2.25 Basic classification of fracture modes.

Ductile and brittle fracture occurs on exposure to stress levels higher than the ultimate tensile strength of the materials. Since the resultant failure is more or less of immediate imminence upon application of stress, they are not generally classified as forms of materials degradation. However, the ductile or brittle nature of the materials has a significant effect on the fatigue and creep resistance of the materials which is dealt in the following sections.

Ductile and brittle fracture occurs either immediately or fairly rapidly after a mechanical load is applied and is not usefully classified as forms of materials degradation. As described in Chapter 1, materials degradation refers to a progressive loss of performance over a significant period of time. A reader seeking more information on brittle and ductile fracture is referred to standard materials science texts.

Although fatigue is just one form of progressive mechanical failure of structural materials, there are several sub-categories of fatigue corresponding to different loading conditions. The most important types of purely mechanical fatigue are described in this chapter, while fatigue mechanisms involving synergy with corrosion are discussed in Chapter 5.

Creep involves bulk deformation of mechanically stressed material but is subject to influence by surface effects. The basic causes of creep are summarized below, for more detailed information the reader is referred to standard materials science texts.

2.3.1 Mechanisms of fatigue and creep

Fatigue induced fracture

Fatigue induced fracture is a phenomenon involving mechanical failure after a large number of cycles of oscillating stress. The amplitude of the oscillating stress is always less than the stress required to cause immediate ductile or brittle fracture. There is an inverse proportionality between the number of cycles of oscillating stress at failure and the amplitude of oscillating stress. Failure of the material is caused by progressive crack growth that continues until the remaining cross section of a component reaches a stress sufficient to cause rapid brittle or ductile fracture. Most mechanical components in service can be assumed to be subject to fatigue because of the ubiquity of cracks in most components. While the fatigue crack is small, the component can still be used but it is vitally important to replace the component before the crack becomes too large. Crack growth can be predicted by the theory of Fracture Mechanics.

Almost all known materials are subject to fatigue, which means that unless the material is subjected to a truly constant stress, eventual failure will occur. Constant stress is not the norm for most engineering structures, stress typically varies with level of imposed duty and environmental factors. The classic examples of fatigue sites are rotating shafts of turbines and other high-speed engines where it is immediately apparent that stresses will oscillate at the same frequency as shaft rotation. Less easily recognized examples are ships hulls and oil-rig structures where the continuous battering of waves generates variable stress. Weighbridges where vehicles continuously pass over a weighing-plate are also subjected to fatigue problems. In general, it should be assumed that stress is variable unless demonstrated to be constant.

Steel and iron offer the favorable distinction from other metals of having an endurance limit to fatigue. The endurance limit is the maximum stress level where fatigue remains too slow to measure. If the stresses in a steel structure remain below the endurance limit, fatigue is considered unlikely to occur. The endurance limit for steel is only observed for situations where pure fatigue without aggravation by wear or corrosion occurs. If a wear scar or a corrosion pit is formed and then functions as a crack initiation site, the steel will display fatigue failure at any stress level. This topic is discussed further in Chapter 5. The inverse proportionality between amplitude of stress and cycles to failure and the endurance limit are illustrated schematically in Figure 2.26 via a graph of fatigue lifetime versus amplitude of oscillating stress.

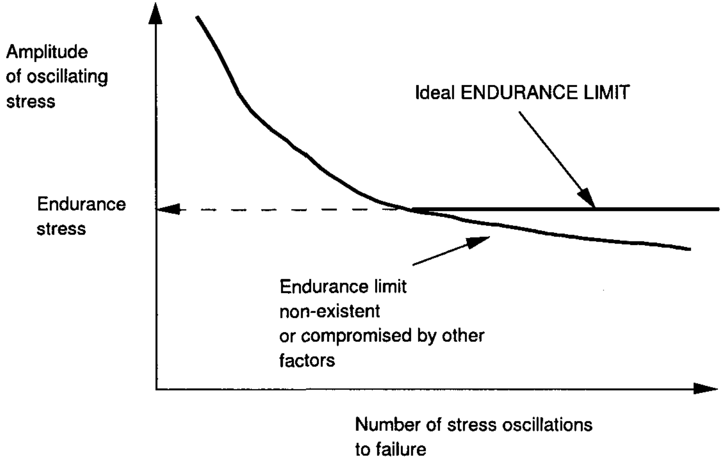


Figure 2.26 The endurance limit of metal fatigue

Many factors have a controlling effect on Fatigue failure of materials, which are listed below

1. Size
2. Loading effects
3. Surface finish
4. Surface treatments
5. Temperature and
6. Environment.

Fatigue failure of the materials is largely dependent on the interaction of stress with the critical flaw size. A large component will have a smaller stress gradient as compared to a smaller one in case of stress concentration occurring due to a critical flaw thus exposing a larger volume of material to high stress. This leaves us with an increased possibility of initiating a fatigue crack in large sized components.

Loading has considerable effect on the endurance limits of the material. The endurance limit varies with the type of loading the structural material is subjected to. Under normal conditions the limits are specified for axial loading conditions where there is no stress gradient present in the material. In the case of rotating bending or torsional loads due to the presence of stress gradients the endurance limits are on the lower side and varies from 0.6 to 0.9 and 0.3 to 0.45 times the endurance limits in the axial direction. Moreover the mean stress values in fatigue loading have a profound effect on the cycles to failure.

The surface finish has an immediate effect on the fatigue crack nucleation as the surface irregularities act as a source of stress concentration. The irregularities act as an aggravating factor in addition to the geometry effects. This is one reason why mirror-polished surfaces are preferred in members, which are subjected to cyclic or oscillating stresses.

Fatigue cracks always initiate in the free surface of the materials. Thus any surface treatment can have a significant effect on the fatigue crack initiation. The fatigue crack initiation is quicker in ductile materials than brittle materials upon application of cyclic loads. This is due to the fact that dislocation motion is an important factor in fatigue crack initiation. However, the crack propagation is faster in brittle materials than ductile materials as the crack initiated act as a source of stress concentration to open up the crack. It is thus preferable to have a brittle surface with a ductile core to delay the fatigue crack initiation and propagation stages respectively. So generally the members subjected to cyclic loading are given prior surface treatment such as shot peening to induce compressive residual stress on the surface to delay the crack initiation process in the material. Thermal processes such as carburising or nitriding involve diffusion at higher temperatures, which also help to improving the fatigue strength by inducing a compressive residual stress. Metal plating operations, which induce tensile residual stresses, are often detrimental to the fatigue resistance of the material.

At higher environmental temperatures creep predominates over fracture. But at slightly elevated temperatures the endurance limit disappears in steel due to the increased mobility of dislocations. Elevated temperatures also anneal the surface removing the beneficial effect of compressive residual stress produced by surface treatments. The endurance limit increases at lower temperatures, but a significant reduction in fracture toughness is experienced in most materials at low temperatures. Moreover the ductile to brittle transition takes place below the critical temperatures resulting in increased flaw size dependence.

The operating environment influences fatigue phenomena and the endurance limit is often altered due to the corrosion pits produced. The alternating stress intensity/ mean stress values drops to a much lower value when corrosion coincides with fatigue of materials.

Most of the theory of fatigue relates to metals rather than polymers and ceramics. These materials show similar fatigue characteristics to metals but with some significant differences.

Mechanical fatigue

The mechanism of fatigue involves 3 distinctly different stages: Crack Nucleation, Crack Propagation and Fracture. Nucleation is the stage most affected by the surface characteristics and features of a material while crack propagation is strongly affected by environmental agents, e.g. air as discussed above.

Cracks are usually nucleated at the surface of a component because the surface contains more severe stress concentrations than the subsurface unless there is a severe pre-existing flaw in the material. Surface roughness, corrosion pits, wear scars can all generate significant stress concentrations at the surface. Even a fine surface finish such as produced by grinding will generate a measurable increase in stress close to the surface. A half-exposed inclusion or void generates a higher stress concentration at the surface than the same inclusion or void when located subsurface. As explained above the fatigue crack initiation normally takes place on the surface of the material and is delayed in brittle surfaces than ductile surfaces while the crack propagation is slower in ductile surfaces. Thus a composite material with brittle surface layers and a ductile core is often desired to have higher fatigue strength. The only redeeming factor to the above is the production of stressed layer on the surface of the material with mirror finish surface. The stressed layer is produced by various treatments either mechanically using cold rolling, shot peening, etc. or by thermally assisted diffusion such as carburizing or nitriding process generating a compressive residual stress. Tensile residual stress produced due to processes such as chrome plating or nickel plating is detrimental to fatigue strength as explained above. The other option is to produce a mirror finish surface to minimize the effect of stress concentration due to the surface irregularities. The only redeeming feature is that plastic deformation caused by most forms of surface finishing can create a compressive residual stress.

During the nucleation stage, dislocations in a metal move under the influence of stress and eventually congregate at some favoured location on or in the metal. In most cases, the site is a stress concentration located at the surface of the component. Dislocations require only a very low stress to become mobile; it is the generation of dislocations via Frank-Reed sources for bulk plastic deformation that requires a high stress. When the dislocations reach the surface they contribute to the formation of a crack nucleus. The crack nucleus consists of a series of extruded metal lamella that reach above the surface by a few micrometers. Each metal lamella is separated by a 'fatigue slip band', which is where the dislocations have concentrated. The structure of the crack nucleus is illustrated below schematically in Figure 2.27.

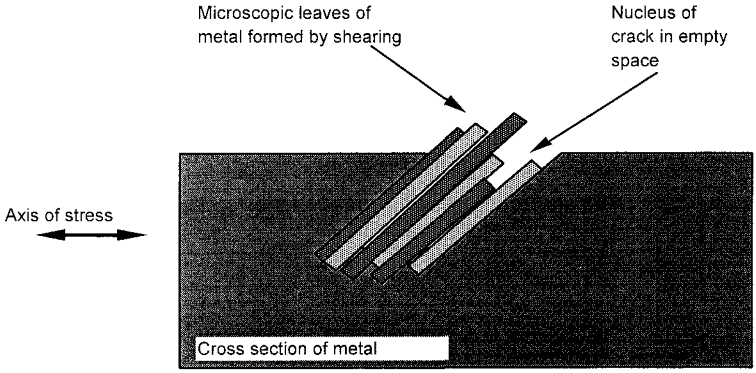


Figure 2.27 Mechanism of fatigue crack nucleation in a metal.

Several crack nucleation structures form on a component at the initial stages of fatigue, the most advanced crack nucleation structure is likely to be the one that later becomes a crack. Despite the small size of the crack nucleation structure, a large portion of the fatigue life is expended by its formation.

In the Propagation stage, a distinct crack or micro-crack forms from the void. The initial direction of growth of the micro-crack, which is controlled by shear stress, is at 45° to the direction of the applied stress. Once the micro-crack has enlarged, the tensile stress concentration around the apex of the crack becomes the controlling factor.

Crack growth is controlled by the cyclic nature of oscillating stress and the path of the crack is never entirely straight. The crack grows most during the period in each cycle of oscillating stress where the tensile stress is at a maximum. When the tensile stress is at a minimum or is converted to compressive stress, crack growth is slow or can even reverse. It is found that fatigue is slower in a vacuum than in air for chemically active metals e.g. nickel if there is a period in the stress cycle, which is compressive. In a vacuum, crack growth can be reversed during the compressive portion of the stress cycle because the lack of oxide films permits near perfect re-adhesion of the opposing surfaces of a crack. Striations or 'beach marks' are formed on the crack surface as the crack front performs an oscillation about the preferred direction of crack growth. The fracture face of a component after fatigue induced failure is covered with the 'beach marks' apart from a small area of ductile or brittle failure where the component finally breaks. While each striation corresponds to crack extension over one cycle of oscillating stress, the number of striations is generally less than the number of stress cycles to failure. This means that there are inactive or latent stress cycles, which are not associated with a specific 'beach mark'

Contact fatigue

Contact fatigue occurs on or close to the surface of any rigid body that sustains repeated rolling contact with only a limited amount of sliding in the contact. The prime cause of contact fatigue is the high stresses generated by Hertzian contact between any pair of curved solid bodies. For example, stresses of 1 GPa or more are commonly found in contacts between steel spheres. Few materials can resist surface damage from repeated exposure to such large stresses. Repeated rolling contact generates a fatigue process of oscillating stress for any material located in the rolling path. Examples of contact fatigue are found in gear teeth, rolling bearings, steel rails and wheels. The prime mode of damage in contact fatigue is pitting of the contacting bodies. Where loads are higher and the size of pits is much larger, the pits are called spalls and the process is termed 'spalling'. Pits and spalls are however closely related and associated where the stresses in rolling contact are high but below the plastic flow stress of the contacting material. Examples of rolling contact fatigue are illustrated schematically in Figure 2.28.

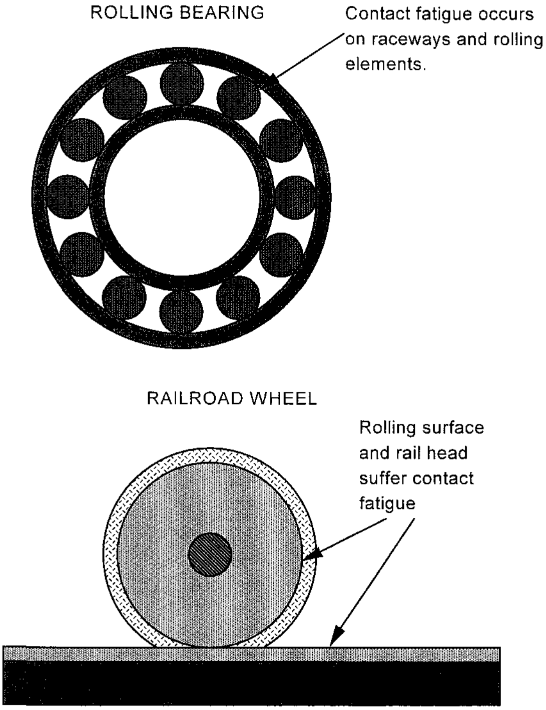


Figure 2.28. Common examples of rolling contact fatigue.

As is the case with other forms of fatigue, the three stages of initiation, propagation and fracture are identified. In the initiation stage a small crack originates from either a subsurface inclusion or at the surface from wearing contact between the two rolling bodies. Initiation is strongly correlated with shear stress and most crack initiation sites from subsurface inclusions are located close to the subsurface maximum of shear stress for a Hertzian contact. Thus, unlike mechanical fatigue where surface initiation of cracks predominates, contact fatigue is often initiated at a small depth below the surface. The precise location of the crack nucleation site is usually close to the Hertzian shear stress maximum except for when a major stress concentration occurs at a different depth. Inclusions and voids can become favored sites for crack nucleation. If the rolling of the contacting bodies is combined with some sliding and wearing contact, then surface scratches may initiate cracks. Where crack initiation occurs close to the surface to form a shallow pit, this is classified as the Surface Mode of contact fatigue. When crack initiation leads to the formation of a deeper pit with the origin of the pit evidently at some distance below the surface, the Subsurface Mode of contact fatigue has been said to occur.

The initiation stage often lasts for a considerable amount of time and does not impair mechanical function of the rolling contact. The next stage is crack propagation where more rapid growth of the crack occurs leading to the formation of what is in fundamental terms a wear particle. The rolling contact remains functional during this stage but its lifetime is already limited. The final stage is fracture where the crack is large enough to allow detachment of the wear particle. Once the wear particle is detached, the rolling contact vibrates since one of the contacting bodies no longer has an accurately circular profile and if rolling speed is high, considerable noise is generated. The rolling contact is usually (but not always) considered to be unusable at this stage and replacement of the rolling parts is required. The mechanism and three stages of rolling contact fatigue are illustrated schematically in Figure 2.29.

Contact fatigue occurs in both lubricated and unlubricated situations, examples of lubricated contact fatigue are found in gears and rolling bearings. Unlubricated contact fatigue can be seen in railway wheels and rails. The presence of oil in lubricated contact fatigue does modify the mechanisms involved. The oil can contain additives; particularly sulfur and chlorine additives that can accelerate contact fatigue by a form of Corrosion Fatigue (see Chapter 5, section 5.3) while all oils can accentuate contact fatigue by a process known as the Hydraulic Crack Propagation Mechanism. Any oil present inside a rolling contact is subjected to very high pressures of 1 GPa or more. This high pressure enables the oil to separate the contacting surfaces, suppress the surface mode of contact fatigue and thereby prolong the working life of the rolling materials. The same high pressure does however accelerate crack growth by pressurizing the space between the crack surfaces. The apex of the crack sustains an increase in tensile stress as a result of this pressurization and the crack propagation rate is increased. Hydraulic crack propagation is most significant where some degree of sliding as well as rolling is present. It is usually found that the slower rolling body shows more signs of hydraulic crack propagation than the faster rolling body. The mechanism of hydraulic crack propagation is illustrated schematically in Figure 2.30.

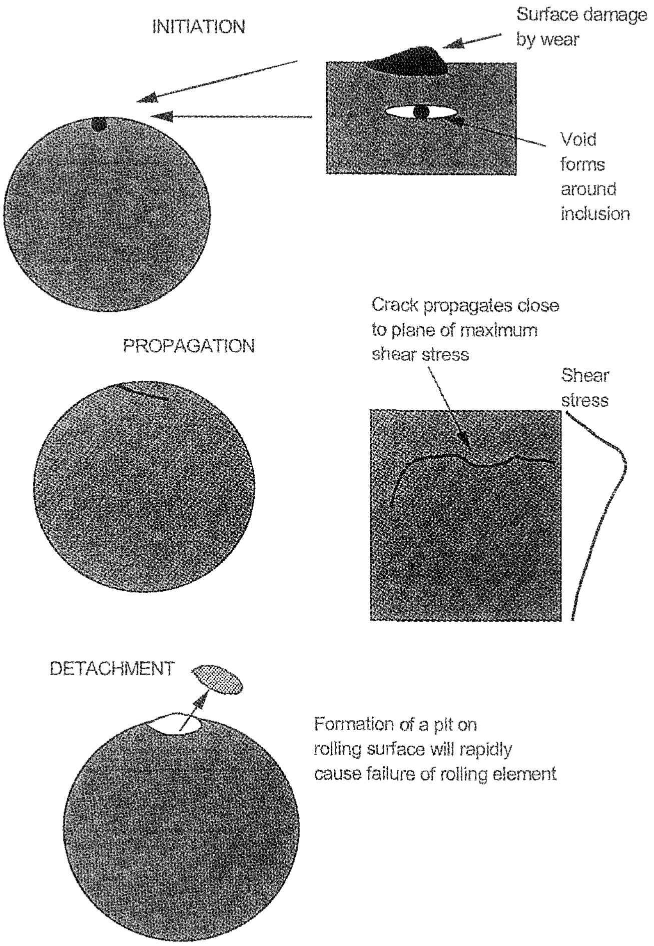


Figure 2.29 Mechanism of rolling contact fatigue.

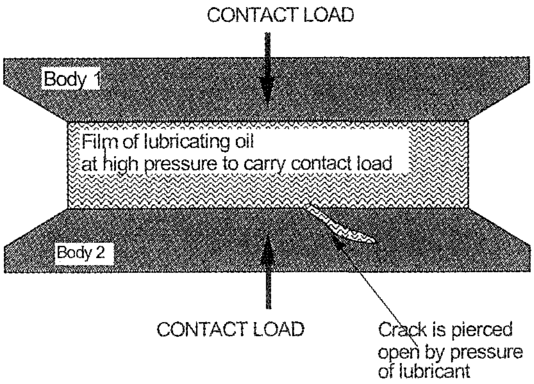


Figure 2.30 Mechanism of hydraulic crack propagation in lubricated rolling contact fatigue.

Dry unlubricated rolling contact fatigue always occurs simultaneously with wear of the rolling surfaces. For steel rails and wheels, a mild form of oxidative wear (see Chapter 5, section 5.2.1) occurs which produces iron oxide granules as wear particle. When the contact stress reaches the plastic limit of either of the rolling bodies, a process known as corrugation occurs on the softer rolling body. Corrugation is the distortion of the original flat or curved surface of the rolling body to a waved profile. It is frequently found on steel rails with the resulting nuisance of noise and vibration. The exact mechanism of corrugation is still unknown but clearly involves plastic flow just below the rolling surface.

High frequency rolling contact does not only generate surface defects in metals. Heavily used steel rails are prone to a mode of fracture originating in the interior of the rail, which is known as *tache ovale*. This is a large oval shaped defect that develops close to the center of the rail cross-section. The tache ovale is usually aligned with longitudinal axis of the rail and can be much longer (i.e. axially along the rail) than its height. When the tache ovale reaches sufficient size after prolonged usage of the rail, there is a sudden fracture of the rail as it splits open. The concealment of tache ovale inside the rail with no external signs of damage and the sudden and catastrophic failure of the rail represent a severe problem for the management of railways. Many railways practice regular monitoring with eddy current detectors or other devices designed to reveal tache ovale and other internal defects.

Thermal fatigue

Stresses leading to cyclic crack growth are not only generated by the deliberate application of mechanical loads but also are caused by rapid changes in temperature. If a solid body is made to rapidly cool or become hot, the variation in temperature between the exterior and interior of the body will cause thermal expansion stresses. When the temperature rises or declines of several hundred degrees Celsius in a few seconds are involved, thermal expansion stresses become severe and a phenomenon known as thermal fatigue results. Instances of thermal fatigue are found in gas turbine blades of e.g. aero-engines, which are exposed to very rapid changes in temperature during start up. Cylinder blocks and heads of internal combustion engines also suffer from thermal fatigue.

Thermal fatigue reaches final fracture more quickly than mechanical fatigue at ambient temperatures. Fracture typically occurs after 1,000 to 10,000 cycles but the frequency of the stress cycle is lower than of a rotating shaft so that the lifetime is still usefully long.

Fatigue of non-metallic materials

Polymers and in some instances, ceramics, display similar characteristics of fatigue, i.e. slow crack growth under alternating stress but the micro-structural mechanisms of fatigue are very different from metals. The mechanisms of fatigue in polymers are also fundamentally different from that of ceramics.

Crazing and shear band formation in polymers

Crack growth under fatigue loading in polymers is largely controlled by the formation of either a crazed region ahead of the crack or a series of shear bands. Ductile crystalline polymers are more likely to display shear bands, while brittle glassy polymers tend to form crazed regions. Shear bands are planar regions where shear strain is concentrated and are aligned along the plane of maximum shear stress. Crazing is characterized by the formation of a diffuse 'crazed' region ahead of the crack front with fibrous tendrils that extend across the gap between the opposing surfaces of the crack. Unlike metals here, there may not be a distinct crack apex. Formation of the fibrils and crazed region are associated with chain scission and extreme alignment of molecules under mechanical stress. Shear bands and crazing around a fatigue crack are illustrated schematically in Figure 2.31.

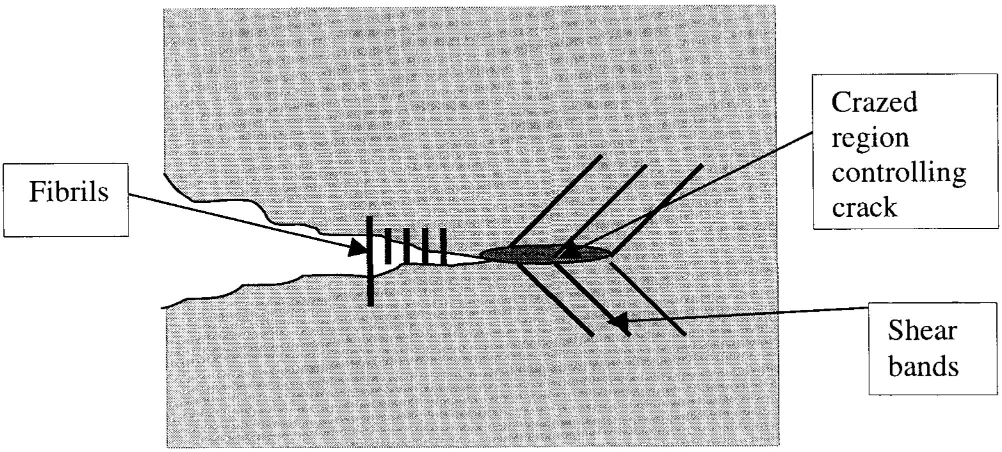


Figure 2.31 Fatigue mechanisms of polymers.

Crazing is often located at the surface of the polymer where it generates an initiation site for a crack. Consequently, fatigue in polymers is usually surface originated, as is the case for metals.

Crazing and shear band formation can cause crack growth to proceed by a series of steps rather than in a continuous manner. This is considered to be a benign phenomenon because of the slower rate of crack growth that results.

Crazing can be induced by chemical reagents, leading to a phenomenon known as Environmental Crazing. Polymers when stressed and is in contact with chemical reagents undergo a slow brittle failure termed as environmental stress cracking and crazing (ESC). ESC is the premature initiation of cracking and embrittlement of a plastic due to the simultaneous action of stress and strain and contact with specific fluids. This definition provides the key to the problem of predictability. Failure is due to a combination of influences, which would not cause the same problems if encountered individually. A phenomenon known as static fatigue can also occur where a steady stress causes delayed fracture by the formation of a crack from a region of degraded polymer.

High frequencies of fatigue loading may cause significant heating of the strained polymer. Since the mechanical strength of polymers are extremely sensitive to temperature, thermal softening may occur with high frequency fatigue loading.

Ceramics

At low temperatures, the mechanical properties of ceramics are dominated by their inherent brittleness, which can lead to sudden fracture. Brittle fracture is determined by a critical level of stress intensity, with little effect from oscillating stress amplitude. Growth of long cracks at room temperature is accelerated when, a static load is substituted by an oscillating load. Long cracks are cracks, which are larger than crystalline grains. At higher temperatures, e.g. 1000 degrees Celsius, the ductility of ceramics, such as aluminium oxide increases significantly and thermal fatigue or slow thermal fracture becomes an important mode of materials degradation. In this instance, the thermal contraction stresses cause numerous microcracks to form in the ceramic before it finally fractures.

Creep

Creep, like fatigue, is a phenomenon displayed by most known materials. It is widely considered to be an exclusively 'high temperature' mode of failure but low temperature creep is common and significant to engineering materials such as polymers. Creep is characterized by a slow growth or contraction in the length of a component under applied stress. Compressive stress causes a component to progressively contract and conversely tensile stress causes gradual extension. A parameter that is fundamental to the measurement of creep is strain rate, which has the scientific dimensions of reciprocal seconds but is more often quoted as % strain per year. At constant true stress, creep can be categorized into three periods. An initial period where strain rate is comparatively high but then declines to a near constant value, a middle period where the strain rate remains at this constant value and a final period where strain rate increases at an accelerating rate until final failure occurs. These three stages are commonly known as primary, secondary and tertiary creep respectively. The progress of creep through the primary, secondary and tertiary phases is shown schematically in Figure 2.32.

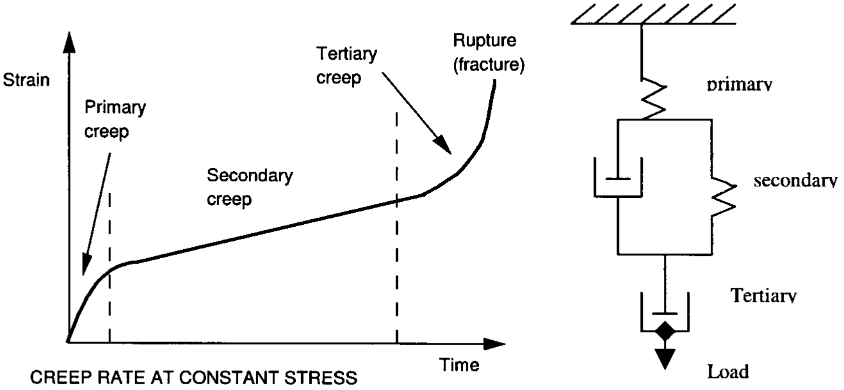


Figure 2.32 Development of creep at constant stress level and a model of different stages of creep.

A model depicting the behavior of materials at different stages creep is shown in the figure 2.32. During the primary stage the creep curve can be modeled as a spring, which shows increasing resistance to deformation with stress. The strain rate during secondary creep is not precisely constant but can be modeled as constant for the purpose of failure prediction. It is during this stage of creep where it can be equated to a spring and a damper in parallel and the resistance and the deformation increases uniformly with applied load. Prevention of failure by creep usually involves ensuring that a material remains within primary and secondary creep within the prescribed period of service.

The propensity of material to creep is not controlled by temperature in an absolute sense but instead is controlled by the ratio of service temperature to the melting point of the material. It is generally observed that observable creep occurs only when the prevailing temperature is greater than 40% of the melting point of the material. For this reason, ice as found in glacier creeps at climatic temperatures since these are very close to its melting point. Another typical practical example could be creep of ice cream during transport due to self-weight. Among the metals, lead and indium also creep at room temperatures because room temperature which is close to 300°K (23°C) is half of the melting point of lead at 601°K (328°C) and greater than half the melting point of indium at 430°K (157°C). For steel and metals with higher melting points, the creep only becomes significant at temperatures exceeding several hundred degrees Celsius. Ceramics with higher melting points than steels, e.g. alumina, are creep resistant to at least 1000 °C while polymers are sensitive to creep at room temperature.

There are several mechanisms of creep which all contribute to the overall creep rate additively, i.e. the creep rate is the sum of creep rates predicted for each mechanism. The most significant creep mechanisms are listed below.

(i) Dislocation glide where dislocations move along slip planes and pass over barriers by thermal activation. Typical barriers are precipitates, solute atoms and other dislocations. This mechanism of dislocation occurs at high stress, which are approximately 1% of the shear modulus of the material sustaining creep.

(ii) Dislocation creep, which involves the movement of dislocations in the same manner as dislocation glide but is controlled by the balance between strain hardening and thermal recovery. Thermal recovery is the dissipation of strain by temperature dependent dislocation movement. This mechanism is significant at stresses ranging from 0.01% to 1% of the shear modulus.

(iii) At high temperatures and low stresses, diffusion of vacancies and interstitial atoms becomes a significant mechanism of creep. The effect of applied stress is to reduce the activation energy for diffusion along the direction of stress. Vacancies and interstitial atoms then diffuse along the axis of stress to generate creep. This mechanism is dominant at low stresses, which are approximately 0.01% of the shear modulus.

(iv) Another high temperature mechanism of creep is intergranular sliding. The strength of grain boundaries is more acutely temperature sensitive than strength of material within a grain. Thus at high temperatures, the grain boundaries become the weakest part of a metal. Intergranular sliding is a major cause of final fracture in creep. Integranular sliding is more probable at low strain rates than at high strain rates.

The Arrhennius law of thermal activation accurately models the temperature dependence of creep. In metals, the Arrhennius activation energy for creep is found to closely correspond to the activation energy of either lattice self-diffusion or grain boundary diffusion or dislocation pipe diffusion. Creep is also strongly influenced by stress where the strain rate can be reasonably modeled as proportional to stress^n where 'n' varies according to the mechanism of creep. Where creep is diffusion based: $n = 1$. For grain boundary sliding and intergranular creep, $n > 1$ where the actual value of 'n' may reach 10. At the end of tertiary creep, voids may form in the component to cause the final fracture.

Interaction of creep and wear

Creep can cause a form of 'quasi-wear', which involves migration of a contact area, or wear scar deeper into one of the contacting bodies but without any weight loss of either contacting body's. The classical model of wear always involves weight loss as wear particles. With quasi-wear, frictional heating and the shear and compression stresses of frictional contact cause material directly beneath the contact area to be displaced away from the contact area. A common example of this effect is found on polymer pins after a wear test. A tongue of displaced material is found at the downstream side of the pin. This tongue contains polymeric material that sustained rapid creep during the wear test until it was removed from the contact interface. Eventually the tongue of material detaches from the pin to form an unusually large wear particle. Creep induced pseudo-wear is illustrated schematically in Figure 2.33.

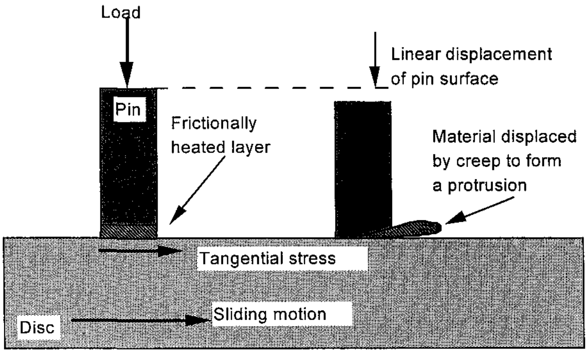


Figure 2.33 Pseudo-wear generated by creep and friction.

Creep of non-metallic materials

Both polymers and ceramics suffer from creep but the controlling physical mechanisms are different. The creep of ceramics displays some similarities to metals where strain is generated by the movement of dislocations. In polymers, strain is dependent on the relative sliding and uncoiling of macromolecules. A further distinction between creep of ceramics and polymers is that the brittleness of ceramics ensures that creep competes with brittle crack growth as the dominant fracture mechanism even at high temperatures. Creep can interact with brittle crack growth by causing closed cracks to open and in doing so, become blunt. Creep in ceramics therefore tends to suppress sudden brittle fracture.

2.3.2 *Fatigue and creep resistant materials*

Fatigue

In principle, fatigue resistance is enhanced by a very clean microstructure with the lack of inclusions and bubbles, which provide sites for crack nucleation. A general guide for enhanced fatigue resistance is to choose a high strength material in order to ensure that the applied stress is only a small fraction of the yield stress. Fatigue resistance is strongly influenced by careful design to avoid the generation of notches and sharp corners and manufacturing quality where surface finish is of high quality. Even a high strength material will suffer early fatigue failure if design and manufacturing standards are not maintained. Many high strength materials also have the disadvantage of being brittle which compromises fatigue resistance as well as entailing the risk of sudden fracture. Steels are effective and cheap materials for fatigue resistance.

Creep resistant materials are usually evaluated in terms of the temperature threshold of rapid creep. On this basis, the most creep resistant materials are the refractory ceramics such as alumina and silicon carbide, which suffer from negligible creep up to temperatures of 1000°C. The limitation of ceramics is brittleness, which often necessitates the use of metals such as nickel and cobalt. Molybdenum is also creep resistant because of its high melting point but as shown in Chapter 3, suffers severely from high temperature oxidation and so cannot be used where oxygen is present. To enhance the creep resistance of a particular metal alloy, the component is often manufactured from a single metal crystal. The lack of grain boundaries in a single crystal suppresses inter-grain boundary creep.

Almost all known organic polymers are subject to creep, the most creep resistant are the thermosetting polymers which have extensive cross-linking between adjacent carbon chains. A high glass transition temperature is also essential for creep resistance.

2.4 Summary

The main mechanical causes of degradation involving wear and fracture are discussed in this chapter. Both fracture and wear comprise of many separate mechanisms, each of which has distinct characteristics. Some forms of wear can be modeled as micro-fracture at the surface but other forms of wear e.g. melting wear cannot be so classified. Frictional contact generates local heating and this is the prime reason why wear can be fundamentally different from fracture. The predominance of compressive stresses in a wearing contact also favors plastic deformation over fracture. Plastic deformation can reach very high levels in a wearing contact and this is one of the reasons why wear rates are usually comparatively low (because material flows over a surface rather than fracturing and detaching).

Fracture is affected by the nature of stress, i.e. steady or variable stress and the prevailing temperature. Variable or oscillating stress causes slow crack growth, which is termed fatigue. Elevated temperatures lead to continuous stretching of a material until it fractures, this phenomenon is called creep. It is very difficult to devise a material, which is permanently resistant to fracture. Eventually fatigue or creep will cause failure of the material, even if a thousand years or more is required.

RECOMMENDED READING.

G.W. Stachowiak and A.W. Batchelor, Engineering Tribology, 2nd edition, Butterworth-Heinemann, 2001, Woburn, Massachusetts, USA.

I.M. Hutchings, Tribology, friction and wear of engineering materials, Edward Arnold, 1992, United Kingdom.

Fatigue under thermal and mechanical loading, Proc. Symposium, Petten, The Netherlands, 22-24 May 1995, edit. J. Bressers, L. Remy, M. Steen and J.L. Valles, publ. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

M. Klesnil and P. Lukas, Fatigue of metallic materials, publ. Elsevier, Amsterdam, 1992.

J. Cadek, Creep in metallic materials, publ. Elsevier, Amsterdam, 1988.

H-H. Kausch, Polymer fracture, Springer Verlag, Berlin-Heidelberg, 1987.

S. Suresh, Fatigue of materials 2nd edition, Cambridge University Press, 1998, United Kingdom.

CHAPTER 3

CHEMICAL CAUSES OF MATERIALS DEGRADATION

3.1 Introduction

Perhaps the most common causes of materials degradation are chemically mediated since they can proceed without the expenditure of mechanical work. It is an unfortunate reality that technology currently relies on metals in a reduced or unoxidized state while under terrestrial conditions; the oxidized state represents chemical equilibrium. Chemically mediated materials degradation largely involves a reversion of metals to chemical equilibrium. If at some time in the future metal oxides, e.g. alumina becomes widely used as engineering materials, the severity of chemically mediated materials degradation will greatly reduced though not eliminated. Alumina and other metal oxides are still vulnerable to hydration and other forms of chemical attack.

There are almost as a great a variety of types of chemically mediated materials degradation as there are chemical reactions. The most significant types of chemical reactions and associated materials degradation are illustrated schematically in Figure 3.1

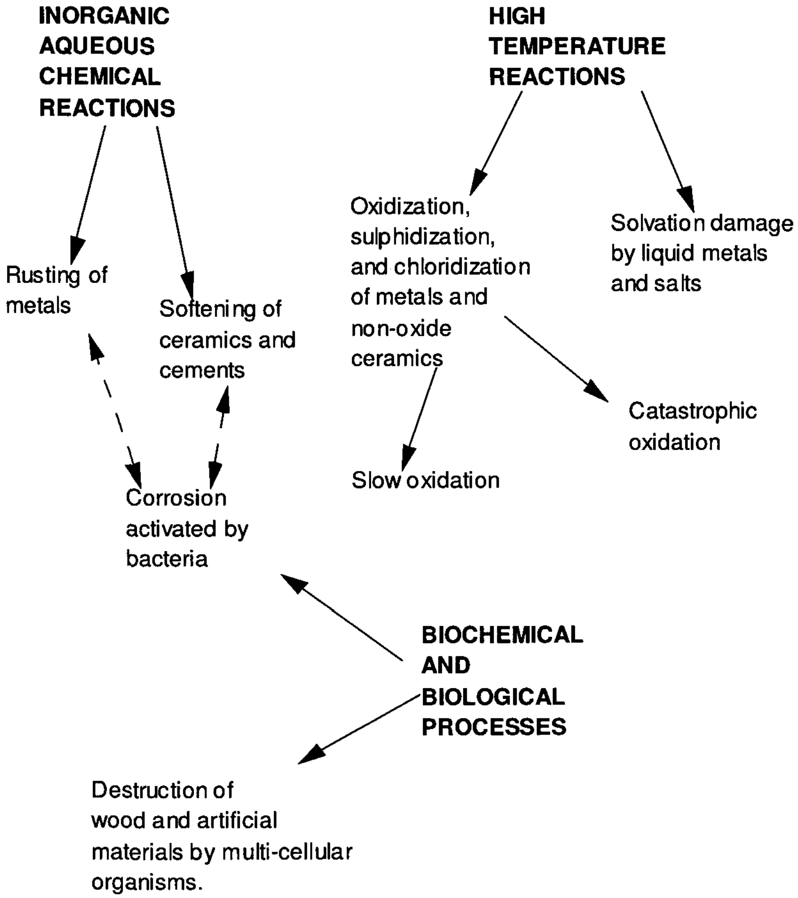


Figure 3.1 Relationship between the various forms of chemical degradation of materials.

Chemically mediated degradation can be divided into 3 basic classifications: aqueous, high temperature and biochemical/ biological. Aqueous degradation involves damage by water, salt water or acids to metals, which either rust or dissolve. Water and acids can also damage non-metals e.g. strong acids rapidly attack concrete. High temperature corrosion is endemic to any metal used at high temperature and is in fundamental terms a form of burning although the protective effect of oxide layers usually prevents rapid destruction of metal from occurring. Molten metals and salts can be very destructive solvents of metals and some non-metals, thus posing problems for moulds and furnace crucibles. Biological organisms can treat solid materials as a food source or else, the material is degraded by corrosive excretions from organisms. Life-forms ranging from single-cell bacteria to mammals (rats) are the culprits. Chemical degradation at very low temperatures is similar to that at ambient temperatures except that the rate of degradation will be less and may even be negligibly small. All of these modes of materials degradation shown in Figure 3.1 will be discussed in this chapter.

In terms of economic costs, aqueous oxidation of metals at ambient or moderately elevated temperatures and dry oxidation of metals at high temperatures are the most significant modes of chemically mediated materials degradation. Polymers and ceramics are also prone to damage by water and solvents, these and other related mechanisms of materials degradation are discussed below.

3.2 Corrosion of metals in aqueous media

The metal of greatest economic and technical value, iron, is prone to corrosion in aqueous media or *aqueous corrosion*. This phenomenon is commonly known as 'rusting' and if unchecked will lead to complete destruction of a steel component. Rusting and other types of aqueous corrosion are an undesirable form of the same *electrochemistry* that is used to store electric energy in batteries. During rusting, the electrical energy produced is dissipated as resistive heating (which is usually too slow to be noticed) while the iron or steel is converted to iron hydroxide, which is too weak to be useful as a structural material. Other metals such as zinc and magnesium suffers a similar process of electrochemical reaction or *electrochemical corrosion* when wetted by water. For metals, aqueous corrosion is almost always equivalent to electrochemical corrosion but for other materials e.g. ceramics this equivalence does not apply.

3.2.1 Electrochemistry and aqueous corrosion

The key to understanding and controlling aqueous corrosion is to analyze the processes involved in electrochemical reactions. Knowledge of basic electrochemistry is therefore a prerequisite for a study of aqueous corrosion.

The electrochemical cell as a mechanism for corrosion

A convenient starting point for discussion of electrochemistry is events occurring when two bars of dissimilar metals are placed in a weak acid solution together and connected by a metal wire. In the context of electrochemistry, the bars (or any other physical form) of the two dissimilar metals are known as *electrodes*. One of the metals could be copper and the other zinc while only a small amount of acid need to be added to the water. The acid is an example of an *electrolyte* since when dissolved in water, the acid raises the level of dissociation of neutral water into positively and negatively charged ions. Other soluble substances such as salt but not sugars are electrolytes. An electrolyte is so called because it facilitates conduction of electricity by the water. The entire assembly of electrodes, electrolyte and connecting wires is an example of an *electrochemical cell*. Recognition of hidden electrochemical cells inherent in engineering structures is an essential skill for corrosion prevention. So long as there is no wire connection between the copper and zinc, neither metal shows any visible sign of reaction although some very slow reaction is almost certain to occur. As soon as the wire connection is established, gas bubbles will be emitted from the copper electrode while the zinc electrode will slowly start to show changes in surface topography as reaction progresses. If a galvanometer is connected between the copper and zinc instead of only a wire, electrical current will be detected. These events are illustrated schematically in Figure 3.2.

The rate of reaction of the more chemically active metal which in this case is zinc, is increased by sharing some of the stages in chemical reaction with a less active metal, which is in this example is copper. This rapidity of reaction is a critical feature of electrochemistry. Chemical reactions in the form of corrosion are almost inevitable but where electrochemical

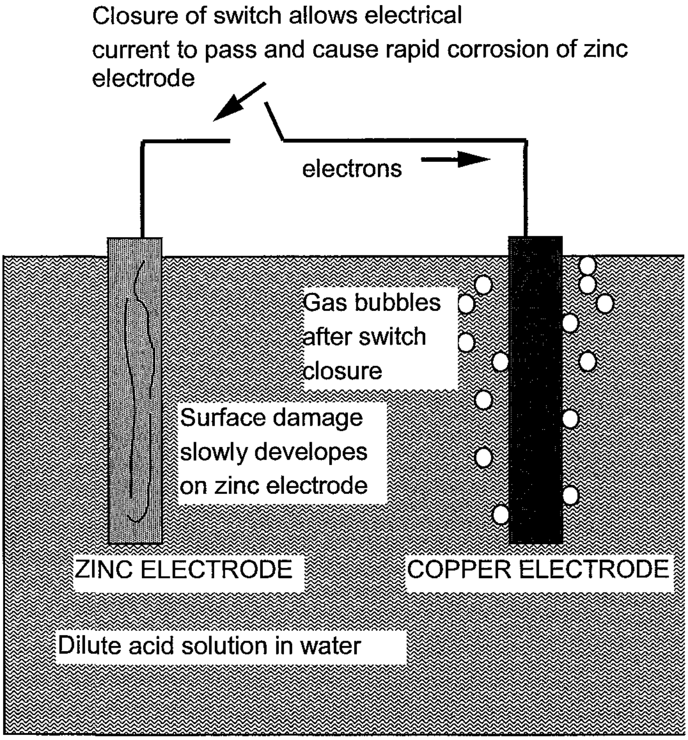


Figure 3.2 Electrochemical interactions between dissimilar metals immersed in an aqueous solution of electrolyte.

factors are favorable, significant corrosion can occur in days but instead of years. If electrochemical reactions can be prevented then rapid aqueous corrosion is much less probable. For example, aluminum and most of its alloys are protected by a strong but extremely thin film of aluminum oxide which blocks the electrical circuit needed for electrochemical corrosion to occur. This is the reason why boats made of aluminum can be expected to last for many years.

Electrochemistry involves two chemical reactions proceeding at different sites but in coordination as observable accumulations of electrons either in the electrodes, electrolyte or conducting wires are not found. At the zinc electrode the following reaction occurs:

$$Zn = Zn^{++} + 2\,e^- \qquad \{3.1\}$$

This reaction means that metallic zinc atoms are converted to zinc ions, which are then sequestered by water molecules to form a *coordination complex* in the water. This coordination complex is a formalized arrangement of water molecules around an ion or alien molecule and forms a basic part of the process of solubilization.

The electrons released by the formation of zinc ions do not remain in the zinc electrode but flow to the copper electrode. At the copper electrode, electrons are consumed in the following reaction to produce hydrogen gas bubbles and water.

$$2e^- + 2\,H_3O^+ = H_2 + 2\,H_2O \qquad \{3.2\}$$

The copper does not suffer any chemical changes other to be a site for this reaction to occur. The zinc electrode, which donates electrons and loses metal to the electrolyte is said to be *anodic* while the copper electrode, which acquires electrons, is said to be *cathodic*. In electrolytes composed of acidic solutions, hydrogen evolution at the *cathode* is the most probable reaction as shown above.

A rate-controlling factor in this electrochemical corrosion is apart from temperature, often the concentration of ions e.g. H_3O+ in the electrolyte. The ionic forms of water, H_3O+ and OH- are illustrated schematically in Figure 3.3.

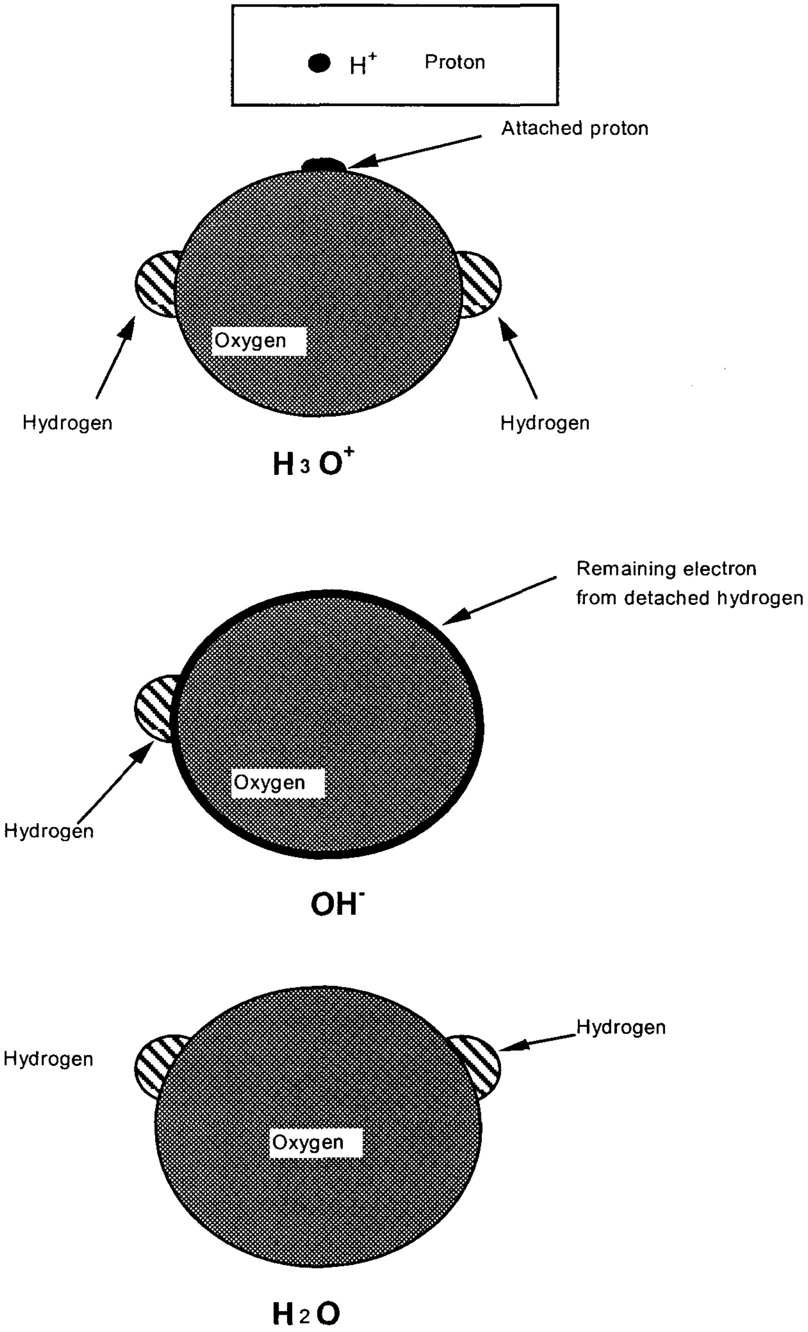


Figure 3.3 Molecular configuration of water in neutral and ionised form.

The hydrogen atoms and hydrogen ions are considerably smaller than the oxygen atom and form only small extensions to the central oxygen atom when bonded to form a water molecule. A water molecule can either gain a hydrogen ion to form H_3O^+ or lose a hydrogen ion to form OH^-. The hydrogen ion itself, H^+ is just a proton which is minute compared to atoms and ions with electron shells. Being so small, the proton would generate such an intense electric field that it would be immediately bonded to an adjacent molecule. In some discussions of corrosion reactions, H+ is written as a reactant not H_3O^+ because the water molecule H_2O appears to be entirely passive in the chemical reaction. It is then possible to omit H_2O, which simplifies the writing of the chemical equation. This abbreviation of the reaction process overlooks the kinetic aspects of the reaction. As will be discussed later, the speed of corrosion reactions is often controlled by diffusion of H_3O^+ to the reaction site. H_3O^+ is much larger than H^+ and so can only diffuse at a much lower rate through water.

Distilled and tap water have rather small concentrations of H_3O^+ and OH^- ions, but sea water which contains about 5% by weight of sodium chloride has a much larger concentration of ions. An abundant supply of ions to facilitate electrochemical reactions is the basic reason why seawater is so *corrosive*. The corrosive nature of sea water is the reason why shipping companies are prepared to spend a small fortune to extract their ships from the sea and send small armies of workers to paint the entire exterior of each ship owned.

The significance of surface structure in electrochemistry

A point often overlooked in discussions of corrosion is the process whereby metal atoms are dislodged from their crystal lattice to be ionized and later solubilized by water. Gulbransen observed that corrosion, as we know it depends on minute imperfections on a metal surface. A typical metal surface is not perfectly planar but in fact consists of terraces of atoms. The terraces are irregular in shape and there are also numerous missing single atoms or single atoms known as *adatoms* that stand isolated above a terrace. Ionization of metal atoms always occurs preferentially at adatoms in particular or along terraces since this where atoms are bonded to the least number of adjacent atoms compared to subsurface atoms. This concept of localized corrosion is illustrated schematically in Figure 3.4

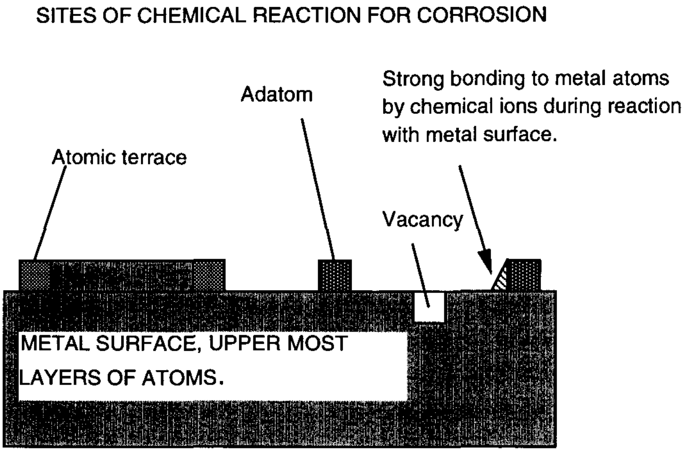


Figure 3.4 The role of imperfections in atomic layering of surfaces in promoting corrosion.

A reasonable objective for future technical development may be to produce a perfectly planar surface, which is therefore corrosion resistant without recourse to rare alloying elements such as chromium. A typical metal surface is not perfectly planar but in fact consists of terraces of atoms. This model of the surface is known as the Terrace-Ledge-Kink (TLK) model.

Electrochemical potentials

Innumerable experimental measurements have lead to the observation that there is a systematic ranking between the various metals of the periodic table in terms of their electrochemistry. Some inactive metals such as copper release bubbles of hydrogen i.e. are cathodic, when connected to almost other metals whereas zinc is only cathodic when connected to a much smaller number of metals. Iron is moderately active so that it is anodic to copper but cathodic to zinc. Metals are usually ranked in terms of their *standard electrochemical potential.* The standard electrochemical potential is the voltage measured between an electrode of the metal to be tested and a nominal *hydrogen electrode, which is in fact,* a platinum electrode immersed in a one molar solution of hydrochloric acid with hydrogen gas flushed through. The voltage is measured at the minimum possible current to prevent voltage loss caused by accumulations of ions around each electrode. Each electrode is surrounded by a pool of electrolyte containing either metal ions of the same metal as used in the electrode, e.g. zinc ions for a zinc electrode or hydronium ions (H_3O+) for the hydrogen electrode. The two different solutions of electrolyte are maintained in electrical contact but physically separate by a permeable membrane. The hydrogen electrode and a connected test electrode are illustrated schematically in Figure 3.5.

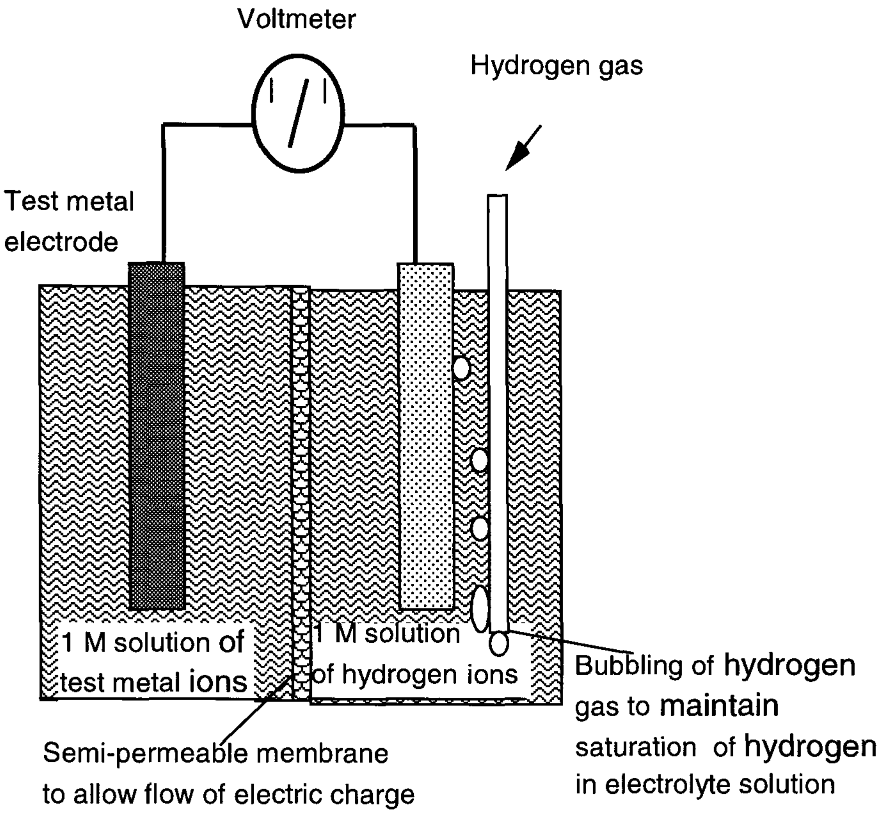


Figure 3.5 Use of the standard hydrogen electrode to measure the standard electrochemical potential of a metal.

The electrochemical potential represents the difference in chemical energy release between when a metal is ionized and solubilized in water and when hydrogen gas is ionized and absorbed by water to form H_3O^+. Where two metals differ in electrochemical potential or energy of ionization, the metal with the lesser electrochemical potential (i.e. the more negative) is more likely to be corroded. A list of some common electrochemical potentials is shown below in Figure 3.6.

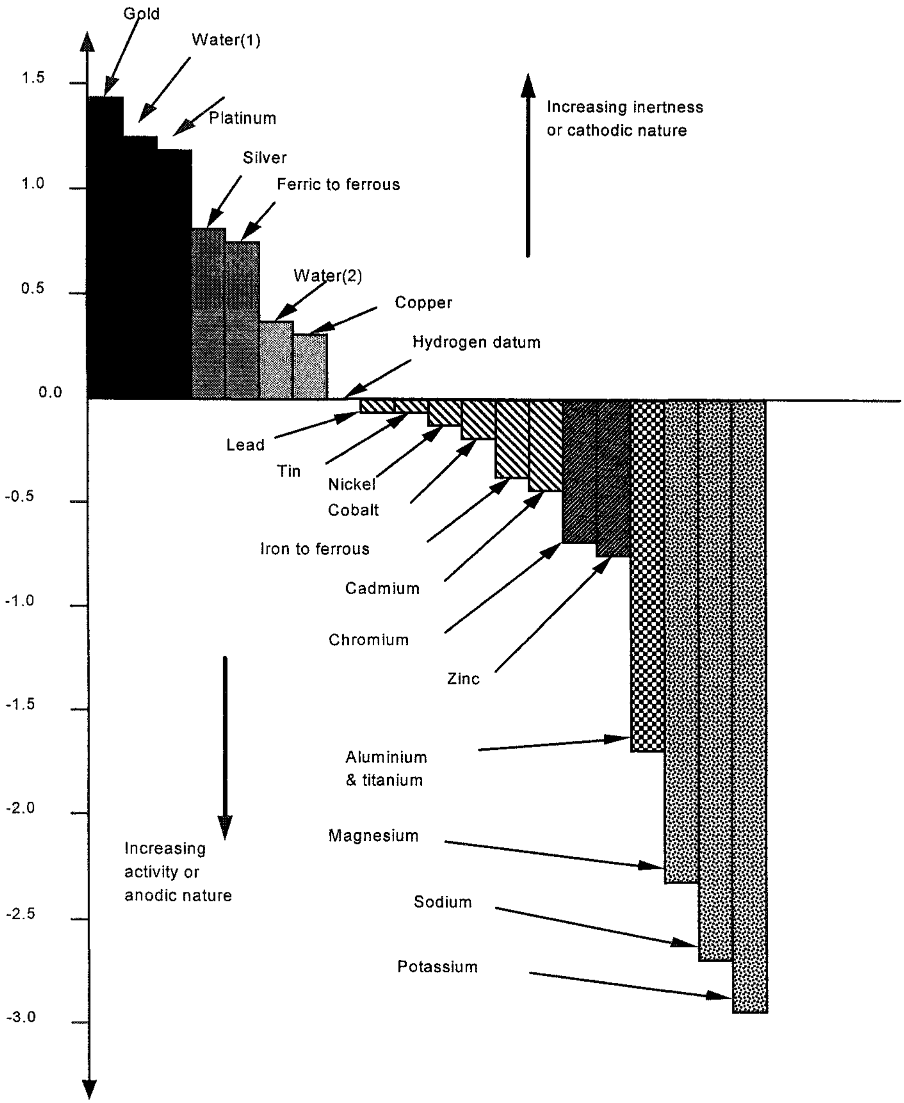


Figure 3.6 Electrochemical potentials of common elements.

The standard electrochemical potential between e.g. iron and zinc in an electrochemical cell can be found by computing the difference in electrochemical potentials relative to the reference electrode. It is possible to refer directly from data based on the hydrogen electrode to data involving a different metal electrode because the only influence that reactions at one electrode have on the other connected electrode is a supply of electric current. This view of the independence of separate electrode reactions is based on the assumption that there is no mixing at all between the corresponding solutions of electrolyte. If such mixing occurs then the electrochemical potential will be affected. To calculate the relative electrochemical potential between iron and zinc, it is necessary to find the electrochemical potentials of the respective electrode reactions, which are as follows.

$Fe^{++} + 2\,e^{-} = Fe$ Electrochemical potential = -0.440 V {3.3}

$Zn = Zn^{++} + 2e^{-}$ Electrochemical potential = 0.763 V {3.4}

In this instance it has been assumed that ferrous ions will be de-ionized and deposited on the iron electrode because iron has a lesser negative electrochemical potential than zinc. The value of the final calculated electrochemical potential will reveal if this assumption is correct. It should be noted that the electrochemical potential of the zinc electrode reaction is the reverse of the table value because the reverse reaction, ionization of zinc is considered here. The two chemical equations and their corresponding electrochemical potentials are then added separately to form two equations representing the overall reaction and the overall release of electrochemical energy. The combined reaction of the two electrodes and the corresponding electrochemical voltage is shown below.

$$Zn + Fe^{++} = Zn^{++} + Fe \qquad \{3.5\}$$

Electrochemical potential = -0.440 V+ 0.763 V = 0.323 V {3.6}

The electrons, which appear on both sides of the combined equation, are eliminated as in reality they do not accumulate (unlike the zinc ions and iron) but are circulated between the two electrodes. A positive value of calculated electrochemical voltage means that the reaction is energetically favorable while a negative voltage means that it is unfavorable. In this example, it is apparent that iron does not usually corrode in the presence of zinc and this result is used in many protective coatings for irons and steels.

Limitations of electrochemical potentials and the galvanic series

Electrochemical potentials are usually only observable in carefully controlled chemical experiments where pure chemicals and clean metals are studied. In practical situations the relationship between metal and propensity to corrode may differ from what is directly predicted by the electrochemical series, i.e. the larger the electronegative potential the more corrodible the metal is. Another series called the *galvanic series* has been devised which gives an approximate allowance for complicating factors such as oxide films which envelope a metal to alter its effective chemical characteristics. Aluminum, as discussed above, is a particularly good example of this phenomenon. The galvanic series is illustrated schematically in Figure 3.7.

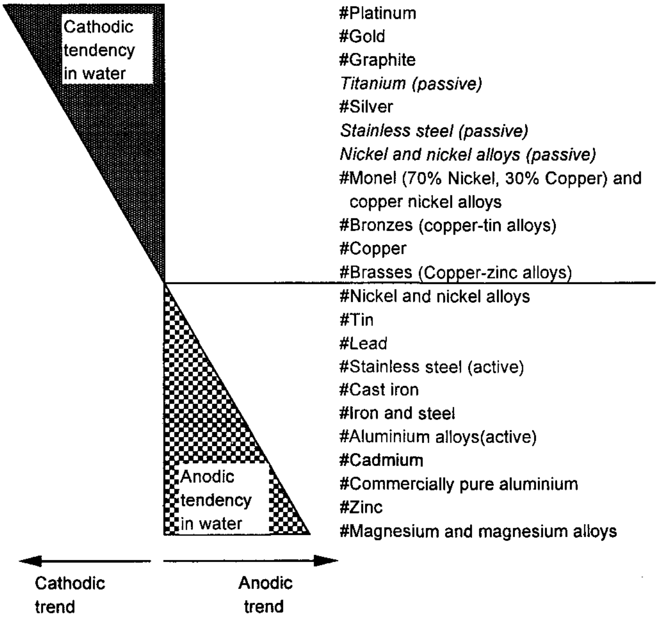


Figure 3.7 The galvanic series of metals.

It can be seen that the galvanic series provides a qualitative indication of corrosion tendency unlike the electrochemical series, which offers quantitative information.

Application of electrochemical potentials and the Nernst equation

Electrochemical potentials are not absolute constants but are dependent on temperature and concentration of the electrolyte solutions. The electrochemical potentials listed in Figure 3.6 are based on measurements at 25°C cell temperature and 1.0 Molar concentration of electrolytes. A fundamental equation known as the Nernst law describes the effect of electrolyte concentrations and temperature on the electrochemical potential. A simplified form of the Nernst law is shown in {3.7}.

$$E = E'' - R.T/(n.F).\ \ln(\text{ratio of concentrations of electrolytes}) \quad \{3.7\}$$

Where E is the actual electrochemical potential, E" is the standard electrochemical potential calculated from tabulated values, R is the universal gas constant (8.31 J/mol. K), T is the absolute temperature of the cell, n is the number of electrons involved and F is the *Faraday constant* which has a value of 96,500 C/mole. The Faraday constant represents the amount of electric charge in coulombs to deposit from solution or solubilize one mole of monovalent electrolyte. In the case of iron and zinc, which are both bivalent 2 * 96,500, C would be required to deposit or ionize one mole of each. In this example, since the number of electrons involved in {3.3} and {3.4} is 2, then n = 2. The ratio of concentrations of electrolytes is by convention has the numerator equal to the concentration of electrolyte that gains ions while the denominator is the concentration of the other electrolyte that loses ions should the reaction proceed. In the example given above this means the concentration of ferrous ions in solution is assigned to the denominator while the concentration of zinc ions is assigned to the numerator. If the concentration of ferrous ions is 10 M (*molar*, i.e. 10 moles per litre) and the concentration of zinc ions is 1 M then the ratio of concentrations is 1/10 = 0.1. At a temperature of 27°C or 300°K, where ferrous ions are present in 10 M concentration and zinc ions are present in 1 M concentration, the electrochemical potential of a zinc and iron cell is calculated in the following manner.

$$E = 0.323\ \text{V} - (8.31\ \text{J/(mol. K)} * 300\ \text{K})/(2 * 96{,}500\ \text{C/mol})*\ln(0.1)$$

$$= 0.323\ \text{V} + 0.0297\ \text{V} = 0.3527\ \text{V}$$

It should be noted that since the dimensional quantity Joules/Coulomb is the definition of electric voltage, the above calculation is dimensionally consistent.

The effect of temperature on electrochemical potential can be seen to depend on temperature. Where there is a negative logarithm of concentration ratio, which means in the current example that the concentration of ferrous ions is higher than the concentration of zinc ions, then increased temperature raises the electrochemical potential of the reaction. Conversely if the logarithm of the concentration ratio is positive then increased temperature will reduce the electrochemical potential. What are the practical consequences of this result? To answer this question, suppose that an iron bolt protrudes from a zinc casing, which is exposed to a concentrated solution of zinc sulfate with a smaller concentration of sulfuric acid. If the temperature is sufficiently high, e.g. if the zinc casing forms a high pressure autoclave where the boiling point of water is elevated beyond 100°C, then it is possible that the electrochemical potential described in {3.7} will become negative. This model of events is based on the premise that the electrochemical potential of {3.5} does not increase. If the electrochemical potential of {3.7} becomes negative then iron will become anodic to zinc and so corrode. For reasons described in section 3.1.2, the small size of the iron bolt compared to the zinc casing will cause far more intense corrosion of the iron bolt than the casing. The possibility of extreme concentration differences reversing electrochemical potentials is illustrated schematically in Figure 3.8

It is found in practice that zinc coatings do not effectively protect iron or steel structures when they are immersed in hot water. The cause of this loss of corrosion protection is a decline in electrochemical potential of the zinc coating with respect to the iron substrate with increase in temperature.

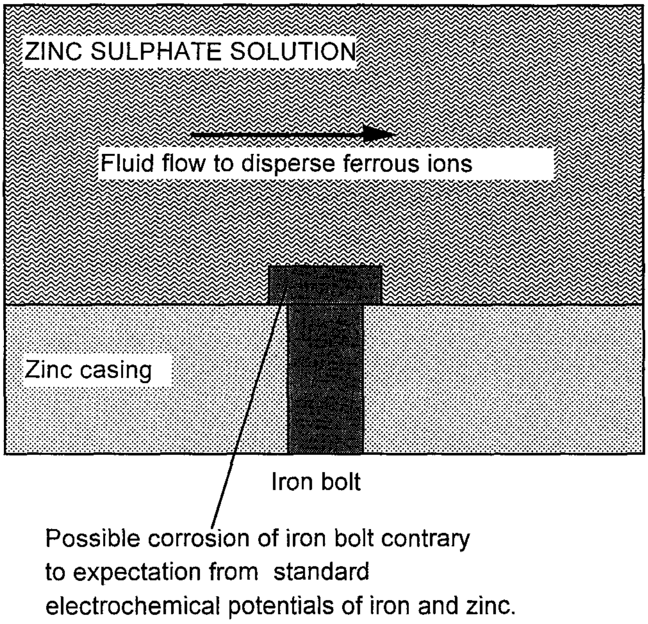


Figure 3.8 Generation of abnormal electrochemical potentials by large concentration differences in electrolytes and associated corrosion problems.

Rate of electrochemical corrosion

Equations of electrochemical potential only predict whether a corrosion reaction can proceed and cannot be used to directly predict the rate of corrosion. Corrosion rates are of critical importance to estimating how long a specific metal will last in a corrosive fluid. A typical question is whether a steel vessel will withstand acid solution for hours, days or months? If the answer is a few months or longer then steel can be used otherwise it will be necessary to specify a more resistant but also more expensive material. The rate of aqueous corrosion is usually answered by calculating *the corrosion current,* which is the current that passes between anode and cathode of the electrochemical cell causing the corrosion. It is convenient to measure corrosion in terms of corrosion current instead of mass of corroded material since corrosion current can be directly calculated from electrochemical theory. Corrosion current also provides a common basis for comparison of corrosion of different metals.

The corrosion current of an electrochemical cell is determined from the Tafel equation, which predicts a linear relationship between electrochemical potential and the logarithm of corrosion current. The Tafel model is a superb simplification of some very complex processes that occur on electrode surfaces and provides a very useful means of estimating corrosion currents. The Tafel model is most effectively illustrated by considering a graph of measured current passing through an electrode versus an imposed voltage. Instead of letting the electrochemical cell generate its own characteristic voltage between its electrodes, the voltage between one electrode and its adjacent solution of electrolyte is controlled externally. As may be instinctively realized, if the voltage is lowered, i.e. a larger negative voltage is generated, the current emanating from the electrode increases. This is because a negative voltage means that electrons are being forced into the electrode. However, when a positive voltage is imposed, the current also increases, as now the electrode is receptive to electrons from the solution of electrolyte. It is also found that the plot of the logarithm of the modulus of current versus voltage has a linear portion and a non-linear portion when the imposed voltage is small. The linear portion represents an exponential relationship between imposed voltage and resulting electrode current, the non-linear portion is caused by the precipitate decline of a logarithm to negative infinity when the current approaches zero. At extremely small currents, the current decomposes into individual electron movements where the concept of continuous electric current is inapplicable so that 'negative infinity' cannot actually be reached. This level of current is hard to observe, of limited practical importance and so is not considered further. When the voltage is progressively lowered from positive values to negative values, the electrode current passes through zero before reversing in polarity. The non-linear portion of the relationship between voltage and logarithm of current is usually overlooked in applications of the Tafel model because unless the electrode is very large, very small currents are involved. Inspection of actual voltage-corrosion current measurements plotted on a Tafel diagram reveals significant non-linearity between voltage and logarithm of current close to a reversal in current polarity.

For large-scale electrodes, which typically consist of entire metal structures, non-linearity becomes detectable at measurable currents and therefore should be analyzed accordingly. Extrapolation of the linear portion of the voltage-current relationship for both positive and negative imposed voltages defines a characteristic current at the intersection of the two linear extrapolations. The linearization of the voltage-current relationship and isolation of a characteristic current, which form the basis of the Tafel model, are shown graphically in Figure 3.9 for an idealised metal electrode.

IDEALISED GRAPH OF ELECTRODE CURRENT VERSUS IMPOSED VOLTAGE

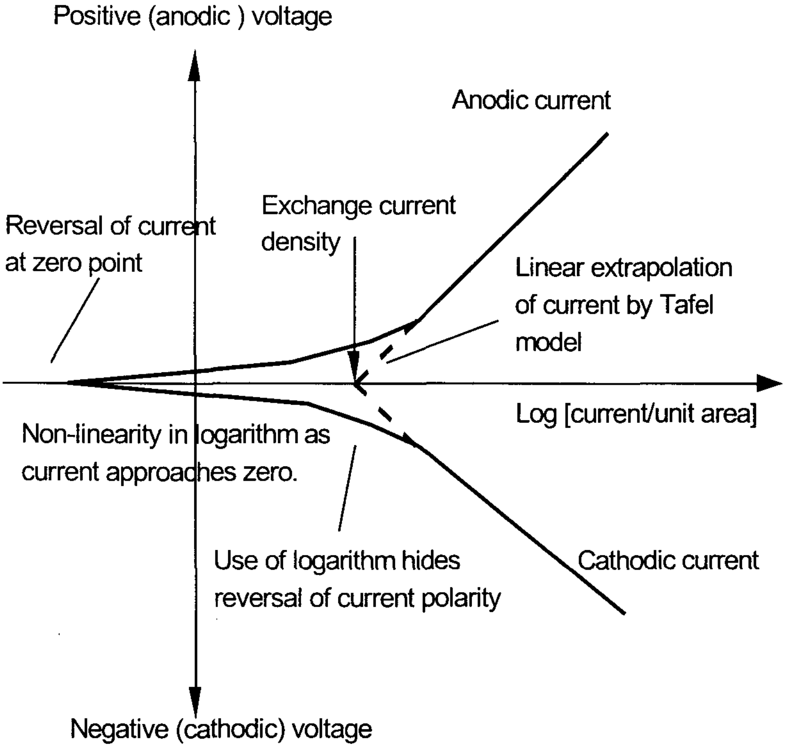


Figure 3.9 Tafel model of relationship between current and electrochemical potential for a zinc electrode.

The Tafel extrapolation of current versus external voltage reveals a minimum non-zero value of electrode current. As may be readily appreciated this current does not flow from the electrode, instead it is recycled within the surface of the electrode where ionization of metal ions are exactly balanced by deposition of metal ions from solution. When an electrode is immersed in a solution of electrolyte, the electrode does not remain inert but continuously changes ions from solution with atoms of electrode material or else continuously dissociates water to form hydrogen while simultaneously synthesizing water. This is an exchange process and the minimum current is known as an *exchange current*. The effect of an externally imposed voltage is to suppress one direction of the current flow while raising the other in order to create a net current flow to or from the electrode. The phenomenon of change in electrode potential (voltage) with current is known as *polarization*. The effect of imposed voltage is therefore not to initiate some previously dormant chemical reactions but merely to bias the rates of the various reactions occurring so that a change occurs in the overall state of the electrode, i.e. it is corroded or it gains material. The concept of exchange current is illustrated graphically in Figure 3.10. Exchange current dictates the rate of corrosion and chemical activity as much as electrochemical potential. Zinc is more rapidly corroded in an acid solution than iron not just because of the difference in electrochemical potential but also because zinc has a higher *exchange current density* per *unit area*. However, electrochemical potential and exchange current density are independent parameters, which do not correlate with each other. For example, the metal with one of the highest exchange current densities, platinum, is well known as a noble metal.

So long as the electrode is isolated electrically, the two opposite components of exchange currents are equal with no net current from the electrode. As soon as the electrode is connected either to another electrode or power source or even to the earth (which is effectively an electrode), the electrode will sustain an externally applied voltage. The difference between the external voltage and the electrode potential calculated according to {3.7} is often referred to as an *over voltage* and its effect is to alter the ratio of exchange current components from its undisturbed value of unity. The over voltage is caused by the activation energy of the rate limiting process on the electrode. There are several reactions occurring at an electrode, e.g. adsorption of ions, electron transfer, supply of non-ionic molecules. The slowest reaction controls the rate of electrochemical reaction and its rate is controlled by activation energy of chemical reaction. Part of the electrochemical potential is lost at the electrode surface to satisfy this activation energy. The full or standard electrochemical potential of an electrochemical system is only

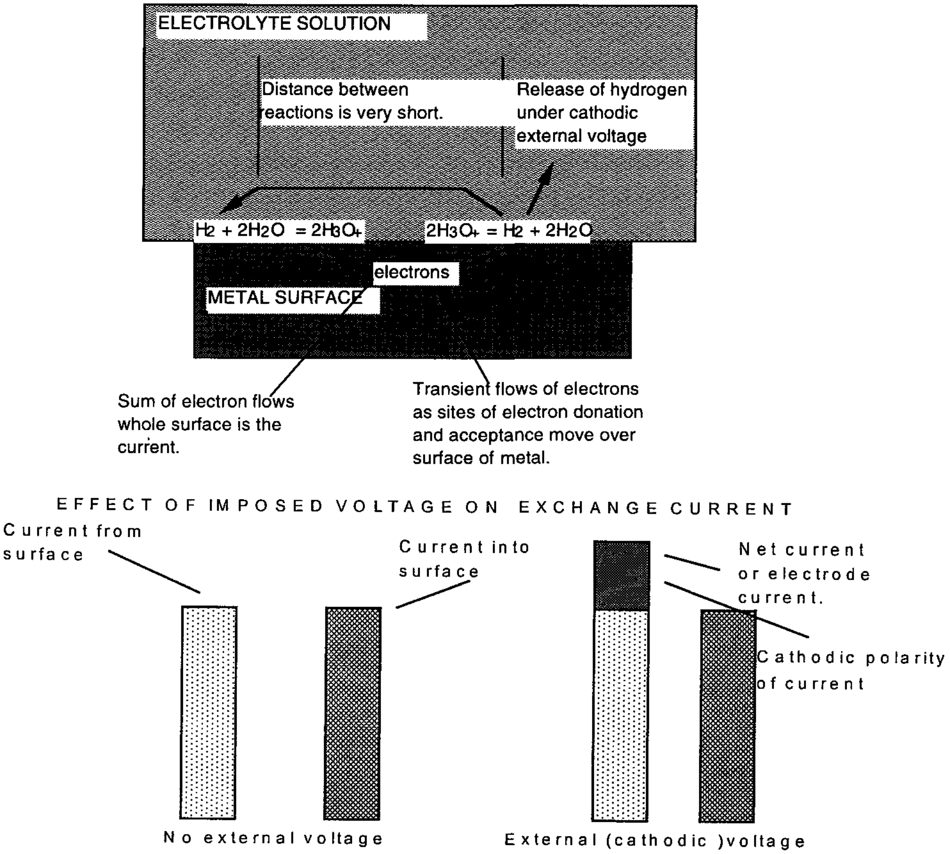


Figure 3.10 Exchange current between an electrode and electrolyte solution and the mechanism of current flow from an electrode.

realized at an infinitesimally small electric current. This loss of electrochemical potential is known as *activation polarization*.

The net current from an electrode as a function of exchange current and over voltage is usually determined from the following approximation {8}, which is accurate for the levels of over voltage usually encountered in corrosion problems. At very small over voltages, the component of exchange current flowing in opposite direction to the voltage becomes comparable to component that flows with the voltage. In this case a different relationship is applicable for which the reader is directed to more specialized texts.

$$E^* = E_{ex} - E'' = \mathrm{K} \log (i/i_{ex}) \qquad \{3.8\}$$

Where E^* is the over voltage [volts], E_{ex} is the externally applied voltage [volts], E'' is the electrochemical potential[volts] calculated from {3.7}, K is known as the Tafel constant [volts], i is the current density passing through the electrode [A/m^2] and i_{ex} is the exchange current [A/m^2].

The Tafel constant is the gradient of voltage versus logarithm of current density shown in Figure 3.6. The Tafel constant varies over a narrow range for most metals of interest and is typically 0.1 V in value. Exchange currents vary over a much wider range, from 100 A/m^2 for platinum to 0.001 A/m^2 for iron and 10 A/m^2 for platinum at ambient temperatures. Exchange currents are a function of temperature and in general rise with temperature.

The next problem to consider before a meaningful corrosion current can be obtained is to analyze how two different electrodes with different exchange currents and electrochemical potentials are coupled to produce a functioning electrochemical cell. The deciding factor is that unless there is considerable resistance in the electrical connection between the two electrodes, both electrodes attain the same voltage or with only a negligible difference in voltage to allow current flow. Most of the electrical resistance of an electrochemical cell is located in the electrolyte solution and on the surface of the electrodes. The electrode currents are also strictly identical as electrons are not stored in the connecting wire or diverted elsewhere. These two conditions of commonality in electrode current and voltage imply a pair of simultaneous equations involving the respective exchange currents and Tafel constants of either electrode. In practice however, a graphical solution is often preferred as an easier and quicker method, which allows for any deviation from strict linearity in the relationship between electrode voltage and logarithm of current. Electrode current or corrosion current is determined by plotting the relationship between over voltage and logarithm of current for both electrodes as a graph of absolute voltage (not over voltage) versus logarithm of current. Corrosion current is found from the intersection of the two plots. This graphical method of solution is shown schematically in Figure 3.11 for zinc and iron electrodes separately placed in dilute acid.

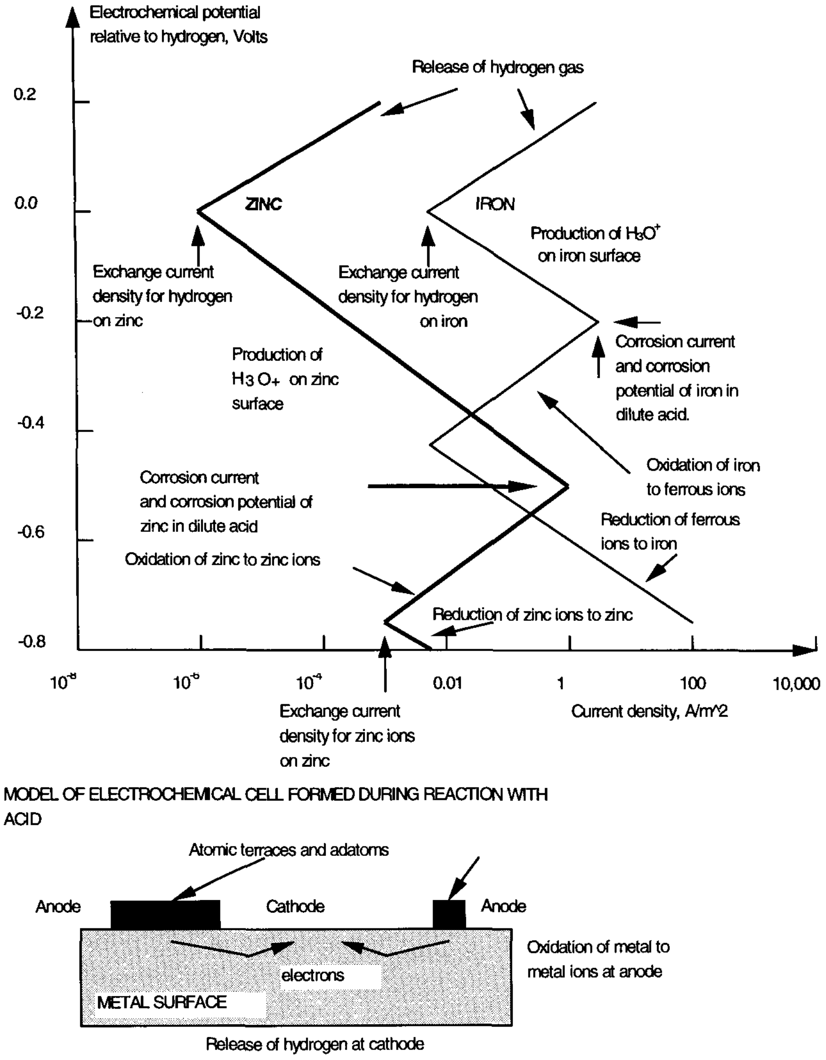


Figure 3.11 Analysis by Tafel model of corrosion of zinc and iron electrodes in acid.

The microscopic electrochemical cells, which are modeled to form on a metal surface, are also illustrated schematically in Fig. 3.11. A real metal surface is not smooth but is covered by numerous terraces and adatoms as shown in Figure 3.4. Adatoms and atoms forming the terraces are more weakly bonded than other atoms to the metal and so will be preferentially ionized. The sites of ionization become the anodes of an electrochemical cell while other areas of the metal surface become cathodic to facilitate hydrogen gas evolution. It is convenient to invoke a model where a nominal hydrogen electrode and the real metal electrode coexist on the same metal surface. The hydrogen electrode is identical to the hydrogen electrode illustrated in Fig. 3.8, except that the corroding metal substitutes the platinum, e.g. zinc so that the corresponding exchange current and Tafel constant are specific to the corroding metal. The corrosion rate is then found from the conditions of equality of electrode currents and voltage. This means that: (i) the current required to ionize or *oxidize* the metal must equal the current released by de-ionization or *reduction* of hydronium ions to produce hydrogen (ii) activation polarization of the metal electrode and nominal hydrogen electrodes must be sufficient to equalize the electrode potentials. The values of current density and electrochemical potential satisfying both these conditions are known as the *corrosion current density* and *corrosion potential* respectively. Both these parameters are fundamental to corrosion engineering. A comparison of zinc and iron is shown in Figure 3.7 to demonstrate why zinc has a lower corrosion rate in dilute acid than iron even though it has a larger negative electrochemical potential. The reason is that the exchange current of hydrogen reduction/ oxidation is much lower on zinc than on iron. Overall corrosion rates, which are of critical interest to the engineer, are often controlled by factors not directly related to chemical activity or position in the galvanic series. By careful manipulation of these factors it may be possible to reduce the corrosion of even very reactive metals.

The mass of corroded metal or the depth of corrosion per unit time is found from the corrosion current by the following calculations. The Faraday constant is 96,500 Coulombs per mole per electron lost, most metals of interest lose 2 electrons per atom upon ionization, e.g. Cu++ or Zn++, the molar weight of zinc is 65 grams. To corrode one mole or 65 grams of zinc requires 2 * 96,500 Coulombs of electricity, which is approximately 1 gram per 2969 Coulombs. 2969 Coulombs is equivalent to 0.82 Amp-hours.

To estimate the depth of material corroded, consider the above example of zinc, 1mm depth of zinc over 1 square meter is equal to 0.001 cubic meters and would have a mass of 7.14 kg. This amount of zinc corroded would be equivalent to 7140 g * 2969 C/g = 21198660 C = 5888.5 Ampere-hours. In other words, 1 Ampere per square meter of corrosion current is equivalent to 0.169 µm per hour depth of zinc lost as corrosion product or 1.48 mm per year. The latter quantity is more useful as a measure of long-term corrosion rates. For the purposes of rough estimation of corrosion of almost any metal, it is safe to assume that 1 Ampere per square meter is equivalent to 1 mm per year of corroded metal depth.

Corrosion prediction for varying environments

The use of the Tafel diagram can often be quite restricted, as each solution for corrosion current is limited to one specific condition. Very often in engineering problems it is required to know which environments cause corrosion for a specific metal. *Pourbaix diagrams* were developed by a researcher of that name for this purpose. According to the Pourbaix model, a corrosive environment can be usefully classified in terms of just two parameters, the *corrosion potential* of the material and the acidity of the fluid that the material is immersed in. For example, the material can be stainless steel and the immersing fluid can be sulfuric acid. The corrosive response of the material, i.e. inert or reacts can be mapped on a diagram where one axis represents corrosion potential and the other axis represents *pH* which is a measure of acidity. The corrosion potential is an engineering term used to describe the electrochemical potential of a specific material in a specific environment. It can be applied to alloys and composite materials in complex fluids and not only pure metals in idealized fluids, as is usually the case for electrochemical potentials. The pH is the negative log of the concentration of the hydronium ion H_3O^+ where p denotes negative decimal logarithm. Neutral water contains approximately 0.1 ppm or 10-7 parts of hydronium ions hence its pH is 7.2. Acidification of water by e.g. adding concentrated sulfuric acid generates a higher concentration of hydronium ions, which causes the pH to decline. Mapping of the corrosive response of a material is based on dividing the graph into regions that denote the equilibrium reaction product. The Pourbaix diagram for iron in water at room temperature is shown in Figure 3.12.

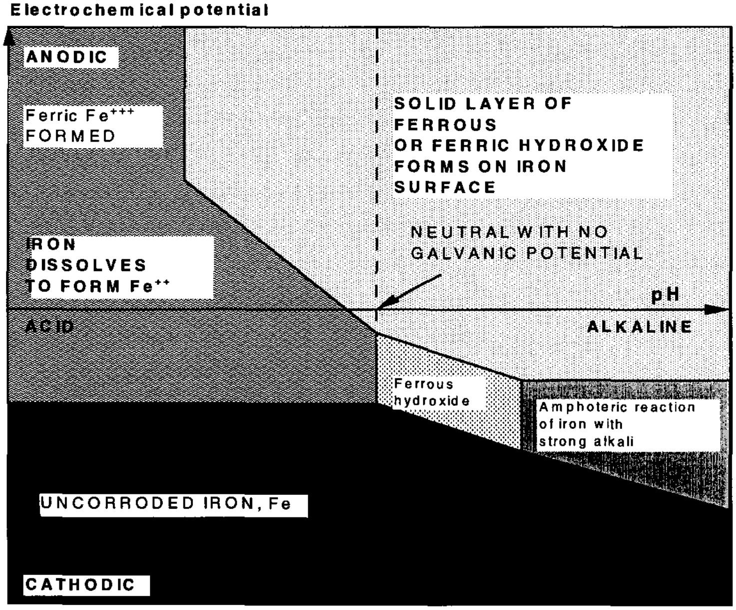


Figure 3.12 Example of Pourbaix diagram: corrosive response of iron to acidified and alkalized water. Adapted from *Fontana and Greene.*

The Pourbaix diagram shows graphically that iron is only stable in water under an imposed cathodic voltage, which can be achieved either by an external electricity supply or connecting the iron to another more anodic metal such as zinc. Corrosion of iron in water is simply the process of establishing chemical equilibrium between iron and water. At low pH (high acidity), dissolution of iron to ferrous or ferric ions occurs. At higher pH, hydroxides are formed which are form an insoluble deposit on the iron surface. At extremely high pH, which corresponds to very strong alkalinity, an amphoteric reaction between iron and the alkali may occur where, the iron acts as a weak acid. This reaction is of only specialized interest in corrosion studies. The 'safe' region for survival of iron is generally considered to be less than the theoretically safe region denoted on the Pourbaix diagram. For combinations of electrochemical potential and acidity that lie close to the transition between un-reacted iron and a reaction product, it is prudent to assume that some corrosion will occur.

In terms of corrosion resistance, the ideal case is inertness; the second best case is passivation by the formation of a solid reaction product. The worst case for corrosion is the production of soluble ions of metal. Temperature is also a major environmental variable but unfortunately cannot be directly considered in a Pourbaix diagram. It is necessary to use a series of Pourbaix diagrams for different temperatures to cover a specified temperature range. It should also be noted that the Pourbaix diagram does not provide any information about the rate of chemical reactions and should never be used even to attempt an 'estimate' of corrosion rates.

The Pourbaix diagram is dependent on the thermodynamic principle discussed previously that corrosion does not usually occur unless the net energy release as predicted by the Nernst equation {3.7} is positive. It is possible for some reactions to occur even when the Nernst equation indicates a negative result because of large entropy changes involved which is an application of the Third Law of Thermodynamics. Such reactions are termed *endothermic* and are not commonly found in corrosion problems.

3.2.2 *Electrochemical corrosion of machinery and structures*

A basic reason why corrosion causes so much damage to machinery and structures is that components of differing materials and shape are placed in close contact with each other. Another reason is that pure metals are rarely used to manufacture components; instead alloys and composite materials are usually selected. These characteristics of mechanical construction and manufacture ensure that a machine or structure contains many electrochemical cells even though in most cases no electrochemical cell was deliberately designed into the system. A simple example of the ease with which electrochemical cells are generated is to consider a steel shaft contacting a bronze bearing when immersed in seawater. Where no electrochemical cell is present, a metal object may last almost indefinitely. An iron column sited in India is observed to have not significantly rusted over many hundreds of years. It is believed that the purity of the iron used contributes to its corrosion resistance by ensuring that electrochemical cells are not formed between the iron and any other metals or carbon that are commonly present in iron. Alloying introduces electrochemical cells on a microscopic scale between grains of different metallic phases.

Another factor that heightens corrosion damage is the increasing hostility of the environments to which manufactured items are subjected. There is a continuing demand for ever higher temperatures when exposed to increasing concentrations of acids and alkalis. Even when there is no deliberate environmental stress, environmental pollution brings increasing amounts of sulfur compounds and other corrosive agents into contact with manufactured materials. These pollutants combine with atmospheric oxygen and water to generate more potent forms of electrochemical corrosion, which can rapidly destroy equipment and structures.

Crevice corrosion and pitting are the most common manifestations of electrochemical corrosion. Pitting and crevice corrosion are examples of localized corrosion, which is associated with early fracture of components before corrosion has damaged the bulk of the component or structure.

Rusting

Rusting is the aqueous or atmospheric corrosion of iron, zinc, copper, aluminum and other metals to either cover the metal component in a layer of corrosion products or else in more severe cases cause the metal component to be damaged by extensive conversion of metal to corrosion products. In most cases, the corrosion products are oxides and hydroxides of the rusted metal but other compounds such as carbonates and sulfates are also possible. Atmospheric oxygen, water, and atmospheric pollutants such as sulfur compounds and to a lesser extent atmospheric carbon dioxide activate rusting. The rate of rusting and corrosion product formed is closely dependent on environmental factors so that accurate prediction of rusting *a priori* is very difficult. The temperature, frequency of rainfall and sunshine, orientation of the rusted surface relative to gravity, concentration of sulfur dioxide, salt crystals and dust particles all influence rusting rates. In general, iron and steels suffer the most severe rust damage of all the metals in common use. The basic reason for that is the product of rusting, iron hydroxide is not impermeable and does not protect the underlying surface to the extent that aluminum oxide protects aluminum.

Under typically humid conditions, a film of water about 1 μm thick will cover any exposed metal surface. Proximity to the sea will expose metals to airborne salt crystals, which lower the minimum humidity for water accumulation on the metal surface. The salt crystals on the metal surface will dissolve in the water film to create saline solution. Exposure to polluted atmospheres adds sulfates and nitrates to the aqueous solution. Iron and steel surfaces, particularly those, which lie horizontally, are therefore covered in a layer of corrosive solution of electrolytes unless the surface is continuously irradiated by the sun or some other source of heat. Light rain increases the amount of water available for corrosion while heavy rain-washes off the electrolytes and delays rusting. Corrosion often begins around a crystal of salt or some other hygroscopic material, which is in the process of absorbing water. A minute electrochemical cell is then formed close to this deposit. A basic characteristic of electrochemical corrosion is that the formation of an electrochemical cell is often a slow process. Once formed, the electrochemical cell usually grows in size and strength to cause an ever-accelerating corrosion rate. After a period of initial corrosion, a layer of iron hydroxide covers a small pit in the metal, which is filled with an aqueous solution of sulfate ions, ferrous ions and hydronium ions. This aqueous solution can be become highly concentrated compared to surface water away from the rust pit and will generate an electrochemical potential that is large enough to permit rapid corrosion of iron. The adjoining metal becomes cathodic to complete the electrochemical cell. The surface of the rusting iron or steel is covered with an inner layer of amorphous ferrous hydroxide ($FeOOH$) and magnetite (Fe_3O_4), which has sufficient electrical conduction to convey current to the surrounding metal. A schematic diagram of the rusting process is shown in Figure 3.13. The chemical reactions of rusting appear

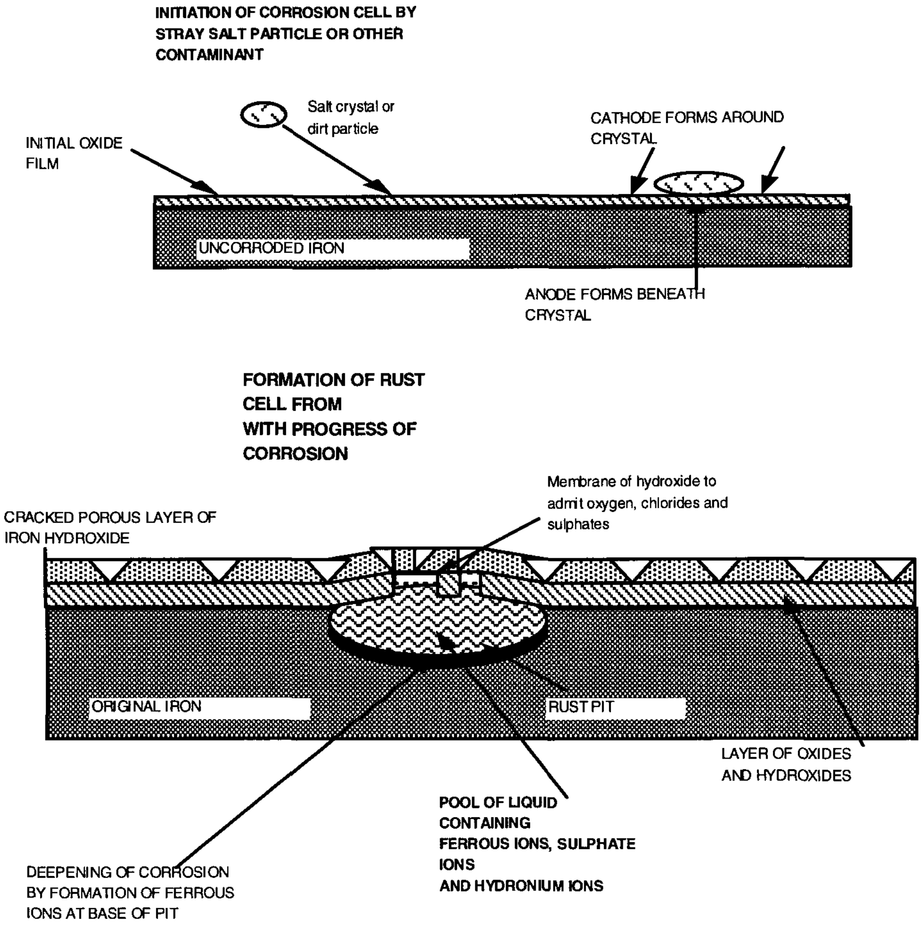


Figure 3.13 Mechanism of rusting of iron in open air.

to proceed in two stages where initially metallic iron is converted to ferrous ions followed by the formation of visible products in the form of ferrous hydroxide. This two-stage process is described by the following chemical reactions.

Dissolution of iron inside the anodic pool of liquid beneath the rust layer:

$$Fe = Fe^{2+} + 2e^-$$

Formation of hydroxide ions on the surrounding metal which functions as the cathode:

$$O_2 + 2H_2O + 4e^- = 4OH^-$$

Combination of hydroxide ions and ferrous ions inside the layer of rust to form more rust:

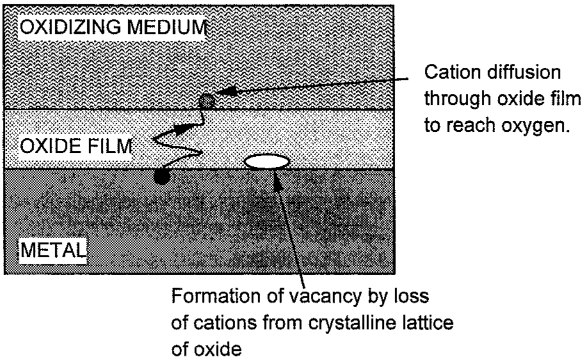
$$4Fe^{2+} + 10H_2O + O_2 = 4\,FeOOH + 4H_3O^+$$

The production of hydronium ions and the flux of electrical current between the anode and cathode attract more sulfate ions so that the fluid at the anode becomes an increasingly concentrated solution of electrolyte. The rise in electrolyte concentrate contributes to the progressive acceleration in the rate of rusting. Protection of irons and steels against rusting mostly involves alloying to substitute the porous hydroxide layer with an impermeable protective layer of oxide. Washing of the surface to remove sulfates and chlorides is very difficult on a large scale. Wire brushing and dry blasting with sand particles to clean a steel surface have the disadvantage of causing plastic deformation of the steel surface which results in traces of sulfates and chlorides being trapped beneath folds of metal [Mattson].

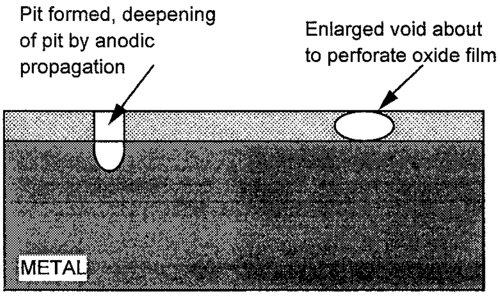
Localized corrosion

Corrosion being an uncontrolled process does not usually proceed in a uniform manner over a wide area. Instead, either specific areas of the surface are corroded more severely than other parts. This non-uniform characteristic of corrosion is reinforced by the nature of electrochemical reactions where for every anodic site of corrosion there must be a cathodic site of non-corrosion. It is extremely common for the surface to spontaneously divide into cathodes and anodes as illustrated in Figure 3.12 for rusting. Pitting is also associated with this segregation of cathodes and anodes. Most surfaces of engineering components are complex to the extent that they contain a mixture of exterior surfaces, which are visible and hidden interior surfaces. The interior surfaces usually enclose spaces that are known as crevices. Common examples of crevices are the space between a nut and bolt or between overlaid sheets of metal. For reasons discussed below, crevices usually become anodic to form hidden sites of corrosion. Much corrosion occurs beyond the range of easy inspection, which further hinders its control.

Pitting represents an intermediate stage between unrestrained corrosion of a metal and its protection by a passivating film [Fontana and Greene]. Pitting is caused by localized failure of a passivating surface film, which exposes small areas of metal surface to unrestrained corrosion. Once a pit is established on the surface, the pit becomes anodic to surrounding metal surface, which functions as a cathode. Rapid corrosion can then occur inside the pit to deepen and widen the pit so that extensive subsurface corrosion of a metal is possible even when the surface opening of the pit is comparatively small. Formation of deep pits by corrosion is a problem in boilers where the pits are found inside the tubes conveying water to be heated. Perforation of the tubes results when the pits reach sufficient depth and as can be readily imagined, the boiler then ceases to function. The initial process of pit formation in an unbroken passivating film is known as *nucleation* and subsequent pit growth is known as *propagation*. The causes of nucleation are still not established despite extensive investigation. One model of nucleation, which has received some quantitative experimental confirmation, invokes the mechanics of ion diffusion in surface film growth. Oxide or other corrosion product film growth is a consequence of metal cation diffusion from the interface between oxide and metal to the outer surface of the oxide. This metal cation diffusion causes the accumulation of crystalline vacancies at the interface between metal and oxide. The accumulation of vacancies causes void formation and when the voids reach sufficient size, localized collapse of the oxide film will occur. A small pit then rapidly forms at the exposed metal. The process of nucleation is comparatively slow so that a metal surface will remain apparently undamaged by immersion in a corrosive medium for a length of time that is known as the *induction period*. According to the ionic diffusion model of nucleation, the induction period is the time required for voids in the oxide film to reach a critical size. The mechanisms of pit nucleation and propagation are illustrated schematically in Figure 3.14.



FORMATION OF VOID IN OXIDE FILM



ENLARGEMENT OF VACANCY TO PERFORATE OXIDE FILM AND INITIATE A PIT

Figure 3.14 Mechanisms of pit nucleation and propagation.

The occurrence of pitting is determined by electrochemical potential of a surface and there is a minimum potential known as the *pitting potential* below which pitting does not usually occur for a particular corrosive medium. There are also other electrochemical potentials, e.g. *nucleation potential* which describes the minimum potential to cause pit nucleation, but a full description is beyond the scope of this book. The metallurgy of a metal has a strong influence on pitting both through the development of passivating films that are resistant to nucleation and through the provision of pitting nucleation sites in the metal structure. Manganese sulfide inclusions in steels are notorious as sites for pitting since manganese sulfide is unstable in chloride solutions. Mixed Manganese-iron-sulfide inclusions are considered to be the most active sites of pitting nucleation. Chromium steels require careful heat treatment to prevent *sensitisation* to pitting. Sensitization is caused by chromium segregation at high temperatures, which results in the formation of chromium-depleted areas of metal surface. Stainless steels are more prone to pitting than plain steel while alloying of steels with molybdenum usually inhibits pitting. Titanium is also very resistant to pitting but is usually too expensive to use.

Seawater and other corrosive mediums containing dissolved chloride ions are particularly active as pitting agents. Seawater contains about 35g dissolved salts per 1 kg of seawater where the dissolved salts have the following composition by weight : 77.76% sodium chloride, 10.88% magnesium chloride, 4.74% magnesium sulfate, 3.60% calcium sulfate, 2.46% potassium chloride , 0.22% magnesium bromide and 0.34% calcium carbonate. Approximately 91.1% of the dissolved salts are chlorides. The chloride ions in seawater are believed to form complexes on passivating films, which partially solubilize the passivating film during the initiation stage of pitting. Chloride ions also accelerate the propagation stage of pitting by a mechanism involving concentration of chloride ions in the pit. The progressive rise in chloride ion concentration with extent of pit formation is caused by migration of negatively charged chloride ions to the pit by diffusion and entrapment of positively charged metal cations by restricted diffusion from the pit. The combination of ferrous metal ions and chloride ions is acidic so that the pH of liquid inside the pit declines. The rate of anodic dissolution of metal ($Me = Me^{n+} + ne^-$) is accelerated by the decline in pH while the high concentration of metal chloride solution lowers oxygen solubility. Oxygen is therefore excluded from the pit and instead reacts to form hydroxyl ions at the margins of the pit, which are cathodic. The mechanism of chloride ion concentration is illustrated schematically in Figure 3.15.

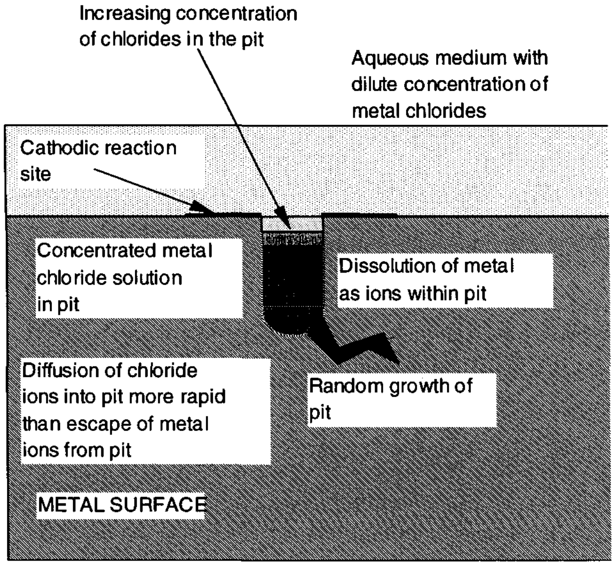


Figure 3.15 Mechanism of intensification of corrosion inside a pit where chloride ions are present in the corrosive medium.

While the sodium chloride found in seawater is a moderately severe pitting agent, hypochlorites (bleaches) are a more severe pitting agent because of their stronger oxidizing powder, which hastens the anodic dissolution of metals. Sodium chloride is also a less severe pitting agent than chlorides of metals such as iron and copper. Cupric chloride and ferric chloride are acidic and both cupric and ferric ions can function as electron acceptors, e.g. $Fe^{3+} + e^{-} = Fe^{2+}$. Electron acceptance or reduction of cupric and ferrous ions favors the anodic dissolution of more electropositive metals such as zinc, aluminum and magnesium thus promoting pit formation. Ferric chloride is often used as a model-pitting agent in accelerated testing. Pitting is considered to be most intense in stagnant liquids, the effect of liquid flow may be to scour the pits and lower the concentration of chloride ions inside the pits. Pitting is not however prevented by fluid flow as evidenced by the boiler tube example described above. For reasons that are still not fully understood, fluoride and bromide ions are much less severe pitting agents than chloride ions.

Crevice corrosion resembles pitting that has reached the propagation stage since the crevice obviates the requirement for initial formation of a micro pit. Crevice corrosion can be severe as the crevice becomes anodic to the surrounding exposed metal structure. Corrosion can proceed unseen inside the closed space to cause failure of critical engineering components such as nuts and bolts. As discussed above, there are innumerable examples of crevices in structures and devices so that crevice corrosion is a widespread problem. Crevice corrosion is highly significant to the phenomena of *corrosion induced fracture and fatigue failure* as is discussed later. A practical measure to prevent crevice corrosion is to substitute riveted, bolted or screw joints by welding so far as is possible. The use of tough inert coatings on the surfaces that form crevices, e.g., on the threaded surface of a bolt is also a possible remedy.

Corrosion in a flowing electrolyte

There is a substantial difference between the corrosivity of stagnant liquid and flowing liquid. A classic example is the rapid corrosion of lead by flowing concentrated sulfuric acid but not stagnant sulfuric acid. Acceleration of corrosion by fluid flow is also known as *erosion-corrosion.* Flow induced acceleration or even initiation of corrosion generates substantial problems for engineering construction. The fundamental question is: How to protect from corrosion pump impellers, valve parts, pipes, stirrers, piston, ships hulls and other devices exposed to flowing liquids? There are a wide variety of examples of flow-induced corrosion but they all share some common basic features. The effect of liquid flow is to remove diffusion limitations on the supply of corrosive reagents to the metal and to expose the passivating film on a metal to mechanical stress. Most passivating films are mechanically weak and therefore easily detached from the underlying metal. A very common process is corrosion by repeated formation and removal of a passivating film. If the intensity of flow is raised higher then a passivating film never covers the metal surface because any passivating material is removed as soon as it forms. In this case, corrosion proceeds in an unrestricted and often very rapid manner. Such unrestricted corrosion is very common in high velocity flow where impingement of a liquid against a surface is involved and usually leads to rapid localized perforation of a metal component. Corrosion by cyclic passivating film formation and removal and unrestricted corrosion where passivation is prevented are illustrated schematically in Figure 3.16.

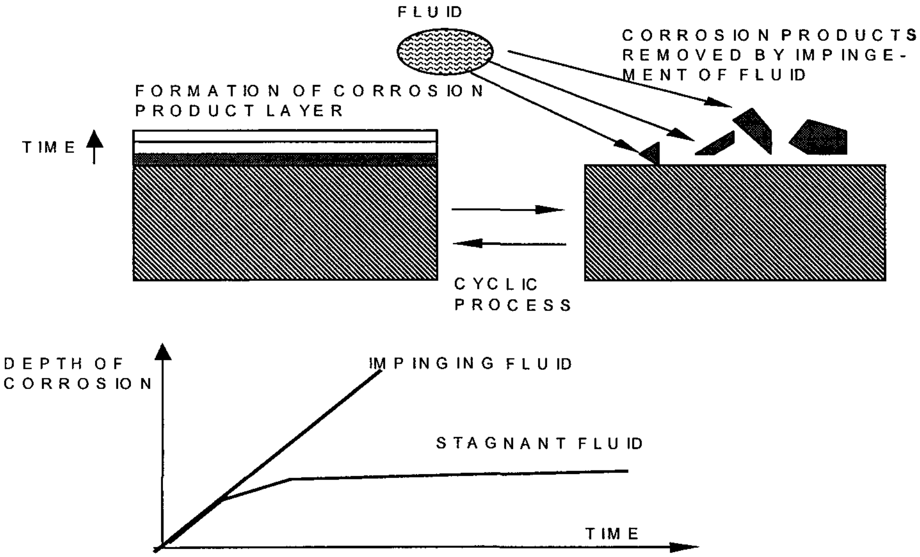


Figure 3.16 Effect of fluid flow on corrosion mechanisms.

Turbulence in a fluid flow accelerates corrosion compared to a laminar flow of similar speed, fluid density and viscosity. This is because turbulence introduces an eddy component of flow, which directly impinges against the passivating films.

Removal of passivating films by fluid flow warrants careful consideration since the overall effect is to convert a normally corrosion resistant material into a corrodible material. Corrosion resistance of metals such as stainless steel is the result of passivating oxide films. This oxide film can be removed by sufficiently high fluid velocity to cause rapid corrosion of the metal.

Even when the velocity of fluid flow is already more than sufficient to cause removal of passivating films, there is still an accelerating effect of flow on corrosion rate. The effect of moderate flow is to raise the diffusion limit on corrosion current. This diffusion limitation is caused by the formation of a boundary layer on the electrode surface where the concentration of the rate limiting reactant, e.g. hydronium ions declines significantly from the bulk value. Fluid flow disrupts this boundary layer and as a consequence raises the maximum current to an extent proportional to the velocity of fluid flow. Thinning of the boundary layer above a corroding surface by increased flow and the resulting enhanced diffusion are illustrated schematically in Figure 3.17.

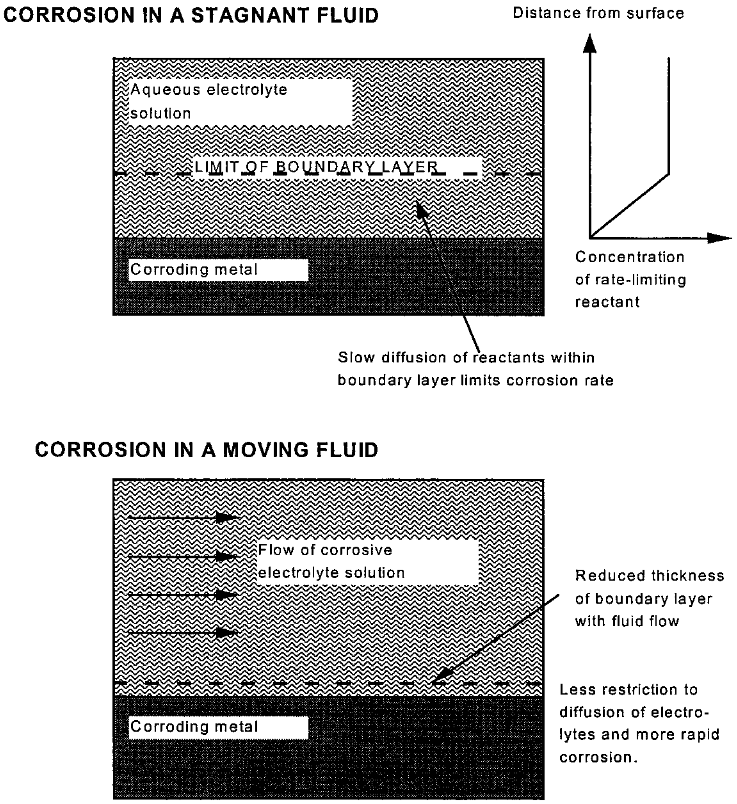


Figure 3.17. Thinning of a boundary layer and enhanced diffusion of corrosive agents during flow of a corrosive fluid.

According to the idealized conditions of the Tafel model, electrode current can rise indefinitely with imposed voltage. In reality however, the diffusion limitations to the transport of corrosive medium to the surface impose a maximum limit to the corrosion current (or current to the electrode). This limit is observed on the Tafel diagram as a sharp deviation of the current-voltage characteristic from the ideal straight line to a two-gradient characteristic where the gradient of voltage versus current at high currents is much larger than the corresponding gradient at lower currents. The effect of increased flow is an increase in the critical current where this change in voltage-current gradient occurs. This critical current is effectively the limiting corrosion current as very high voltages are extremely rare except where external voltage is applied. A schematic illustration of the effect of flow on the electrode current illustrated in terms of the Tafel diagram is shown in Figure 3.18.

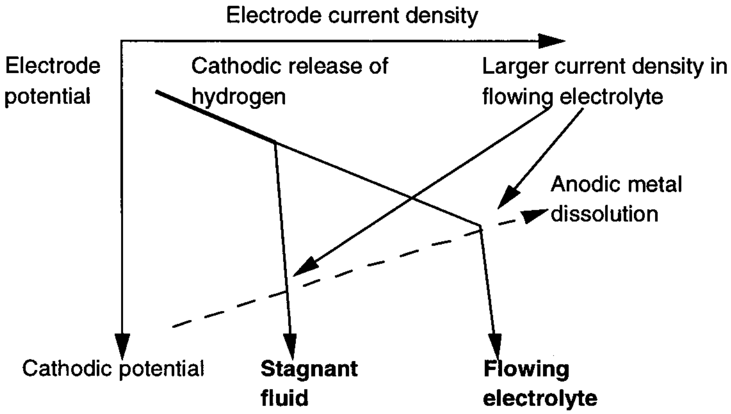


Figure 3.18. Effect of fluid flow on the limiting corrosion current..

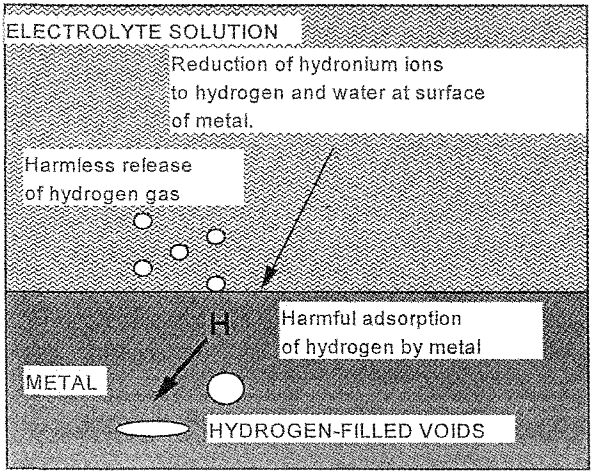
Cathodes are usually subject to more severe diffusion limitations than anodes due to limitations in hydronium ion transport. Under static non-flowing conditions, a small cathode will impose a more restrictive limitation on a larger anode than the opposite configuration, which is a small anode and large cathode. This is because in general, a larger cathode is required for the anode to corrode close to its diffusion limit. Where dissimilar metals are coupled together, intense corrosion of the anodic metal component will result if it is smaller than the cathodic metal. An example is the differing corrosion rates of iron bolts in copper water containers and zinc water containers. For the copper container, the iron bolt represents a small anode, which will corrode rapidly leaving the copper largely undamaged. The iron bolts in the zinc container become small cathodes, which do not significantly accelerate corrosion of the zinc container. Extremely high flow velocity can however cause significant corrosion even where the cathode is comparatively small by raising the cathodic diffusion limit to create an active electrochemical cell.

Control of the maximum corrosion current can have some important repercussions for passivation as is discussed later. Whatever the precise effect on corrosion, fluid flow rate is a basic parameter of any corrosion analysis.

Hydrogen embrittlement and hydrogen blistering

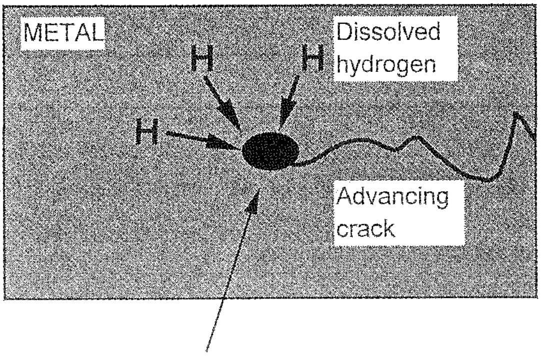
Hydrogen embrittlement is primarily found in steel components, which contain sites of cathodic reactions, and is a significant decline in toughness from the value at the time of manufacture with prolonged exposure to corrosion processes. Hydrogen blistering is an accumulation of gaseous hydrogen inside the metal that reaches sufficient pressure to stretch the metal and so form a bump or blister in the metal surface. The basic cause of hydrogen embrittlement and hydrogen blistering is that hydrogen gas bubble release from the surface of a cathode is not always spontaneous. Instead of being released to the corrosive medium as bubbles, hydrogen may penetrate the metal. Once inside the metal, hydrogen ions rapidly diffuse through the crystalline lattice as they are smaller than the metal atoms. The hydrogen ions continue to diffuse through the metal until they reach microscopic voids inside the metal. Steel contains many of these voids and as soon as the hydrogen ions reach a void they combine to form hydrogen molecules. The hydrogen gas attains a high pressure as hydrogen ions continue to reach the void but no gaseous hydrogen can leave the void. The pressure required to force hydrogen to ionize and re-dissolve in the metal is extremely high, of the order of 1 GPa. The void eventually expands to relieve the pressure and a blister is formed. Hydrogen blistering is illustrated schematically in Figure 3.19. The most severe problems of hydrogen blistering are found in oil drilling where steels are exposed to high concentrations of hydrogen sulfide in hot liquid petroleum. Treatment of steels by acid pickling followed by changes in temperature in service is also a cause of hydrogen blistering.

Hydrogen embrittlement is a complex and still incompletely understood process that involves the formation of metal hydrides. It is however known that hydrogen sulfide strongly inhibits the reaction between atomic hydrogen to molecular hydrogen. Atomic hydrogen therefore accumulates at the sites of cathodic electrochemical reactions before it later dissolves into the metal below. Plastic flow favors the absorption of atomic hydrogen and its transport within the metal since the dislocations produced by plastic deformation provide space for the atomic hydrogen to reside. It is believed that atomic hydrogen diffuses under the influence of mechanical stress to the vicinity of a crack. Accumulation of hydrogen around the crack creates a zone of brittle metal hydride ahead of the crack, which promotes crack extension. This mechanism of hydrogen embrittlement is illustrated schematically in Figure 3.20.



Two paths for hydrogen to escape from cathodic reaction site: Surface mode as gas, subsurface mode as dissolved hydrogen.

Figure 3.19 Mechanism of hydrogen blistering.



Plastic zone ahead of crack attracts dissolved hydrogen to create a brittle hydride region in front of the crack

Figure 3.20 Mechanism of hydrogen embrittlement.

Hydrogen embrittlement occurs at lower hydrogen pressures than hydrogen blistering and is therefore a more widespread problem. Iron and nickel are particularly prone to hydrogen embrittlement while copper and aluminum are far more resistant. Titanium and its alloys are prone to hydrogen embrittlement by the formation of brittle hydrides. Hydrogen embrittlement is favored by low stacking-fault energies, high negative binding energies between hydrogen and dislocations and high fracture strengths. Metals and alloys with low values of these parameters will probably be resistant to hydrogen embrittlement. Low temperatures also promote hydrogen embrittlement because the solubility of hydrogen in metals declines to minute values when the temperature is reduced. This low solubility promotes the formation of a separate brittle hydride phase in the metal structure.

Hydrogen embrittlement and blistering occurs in situations other than aqueous corrosion. Poor process control in metal refining where iron and other metals are allowed to contact water leads to the ingress of atomic hydrogen into the metal. Welding of steels under wet conditions can also cause hydrogen embrittlement.

3.2.3 *Corrosion inhibitors*

A corrosion inhibitor is a substance, solid or liquid, which when added in only moderate quantities to a corrosive fluid should significantly reduce the corrosivity of that fluid towards a particular metal. Given the elaborate nature of most surface engineering techniques, the simplicity of a corrosion inhibitor is a very useful attribute since the engineer is only required to blend a specified amount of proprietary chemical into the corroding system.

Corrosion inhibitors are usually classified according to the chemical mechanism of corrosion inhibition. Most of the inhibitors used currently are either organic or inorganic chemicals for retardation of corrosion progress under different environments. The inhibitors interact with the material surface in question forming a passive film or an adsorbed layer acting as a barrier film. Corrosion inhibitors are used in a wide range of applications such as the corrosion of steel surfaces by oxygenated oil in oil wells, corrosion by acidic mine waters or by hot boiler waters. Corrosive-abrasive wear can also be minimized, by adding inhibitors to the abrasive slurry. It is very important for an engineer to understand the corrosion problem before deciding on a particular inhibitor for application. Since a large amount of inhibitor may be required to inactivate the entire volume of corrosive fluid, the cost of the inhibitor should be predicted before committing to a specific inhibitor. When the optimum concentration of an appropriate inhibitor has been added to a corroding system, a very large reduction in corrosion exceeding a factor of 10, may be obtained.

The type of inhibitors can be classified into 1. Passivators ,2. Barrier inhibitors, 3. Poisons, 4 Scavengers, and 5. Neutralizers. Before discussing, inhibitors, the basic concepts of passivation and passivity will be discussed.

Passivity

Passivity is a phenomenon where a reactive metal shows an unexpected resistance to corrosion by a fluid, which is known to produce severe corrosion in other circumstances. Michael Faraday developed a classic demonstration of passivity in the 19th century. Faraday observed that when an iron specimen was immersed in concentrated nitric acid, very little corrosion could be detected by weight measurements. When water was progressively added to dilute the nitric acid while containing an unscratched and unshaken specimen, corrosion was not observed to occur. However, if the specimen was scratched while immersed in the diluted nitric acid or else the jar containing the acid and specimen was shaken vigorously so that the specimen hit the side of the jar, then rapid corrosion of the iron specimen resulted. When the iron specimen was directly immersed in dilute nitric acid, severe corrosion occurred. This sequence of events is illustrated schematically in Figure 3. 21.

Highly concentrated nitric acid is not a corrosive medium for iron because it causes the formation of a passivating layer. This passivating layer can subsequently protect iron against the far more corrosive dilute nitric acid. The passivating layer is mechanically weak so that scratching and even impact will causes its detachment from the underlying surface. This results in rapid corrosion of the unprotected metal by dilute nitric acid. Immediate exposure of iron to dilute nitric acid causes intense corrosion without the intervention of a passivating film. It is also evident that while a passivating film can provide an extremely useful degree of corrosion protection, it also requires careful manipulation to be effective and reliable.

Passivation can also be induced on a surface by applying an anodic over voltage to the test electrode. Where the ideal Tafel diagram shows an indefinite rise of anodic current with positive over voltage, experimental measurements show that anodic current reaches a maximum and then

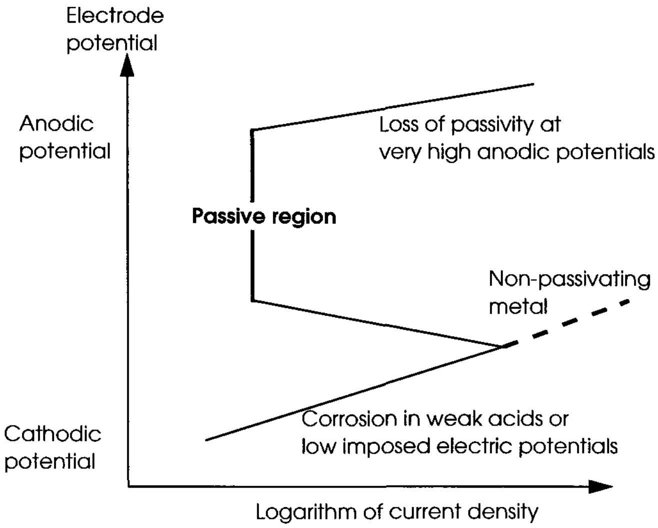


Figure 3.22 Display of anodic passivation on the Tafel diagram.

The main effect of increasing temperature and reduced pH (increased hydronium ion concentration) is to raise the maximum value of anodic current at small over voltages and lower the larger value of over voltage required to destroy passivity. The passivation phenomenon therefore becomes far more distinct for a hot concentrated acid than for a colder dilute acid.

Electrochemical corrosion always proceeds as an interaction between a cathode and an anode so that depending on the electrochemical potential and exchange current density of the partner electrode, a passivating material may either be protected or else corrode severely. The partner electrode can either be a galvanic couple or else a nominal hydrogen electrode on the same metal surface. To ensure reliable passivation, it is necessary to ensure that the exchange current - over voltage characteristic of the partner electrode only coincides with the passivated region of the anodic current characteristic and does not also coincide with the non-passivated region. If there is a coincidence with the non-passivated region then passivation will be unreliable. Under careful test conditions, passivity is possible but under conditions typical of most process machinery and structures, passivation will fail. This distinction in the quality of passivation is illustrated schematically in Figure 3.23.

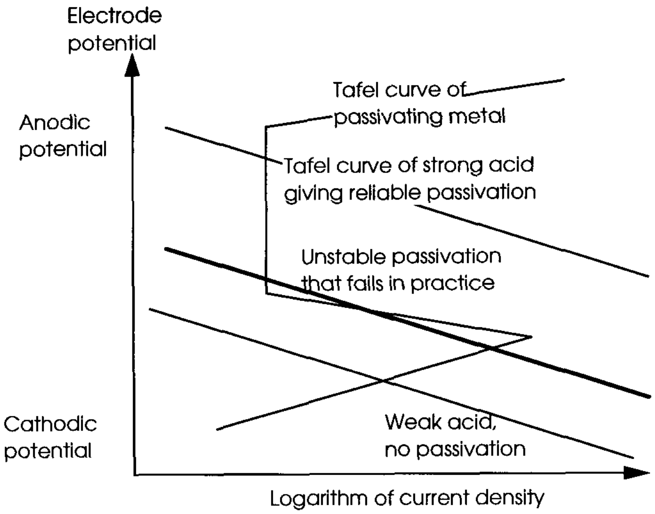


Figure 3.23 Discrimination between reliable and unreliable passivation in terms of the Tafel diagram. Adapted from *Fontana and Greene*.

The principle of *anodic protection* of metals is to engineer anodic passivation by the application of the correct anodic over voltage. The advantage of anodic protection over cathodic protection (which is the application of a cathodic over voltage) is lower electric current requirements. The instrumentation requirements of anodic protection are however more costly than cathodic protection where avoidance of excessive over voltages is not a strict necessity.

Velocity of flow of the corrosive medium can exert some unusual effects on passivity. If one considers a single electrode and nominal hydrogen electrode in a passivating medium, the diffusion limit of the hydrogen electrode usually dictates the overall corrosion current. If the electrochemical potential of the nominal electrode passes through the passivation range then it is possible for corrosion to cease at a critical velocity of flow, where the limit of diffusion current exceeds the maximum current at electrochemical potentials below the range of passivation. Passivation induced at high flow velocities is shown schematically by Tafel diagram in Figure 3.24. Coupled electrodes of dissimilar metals can also induce the same phenomenon of passivation at high fluid flow and this characteristic can be exploited in the design of impellers, pump parts etc.

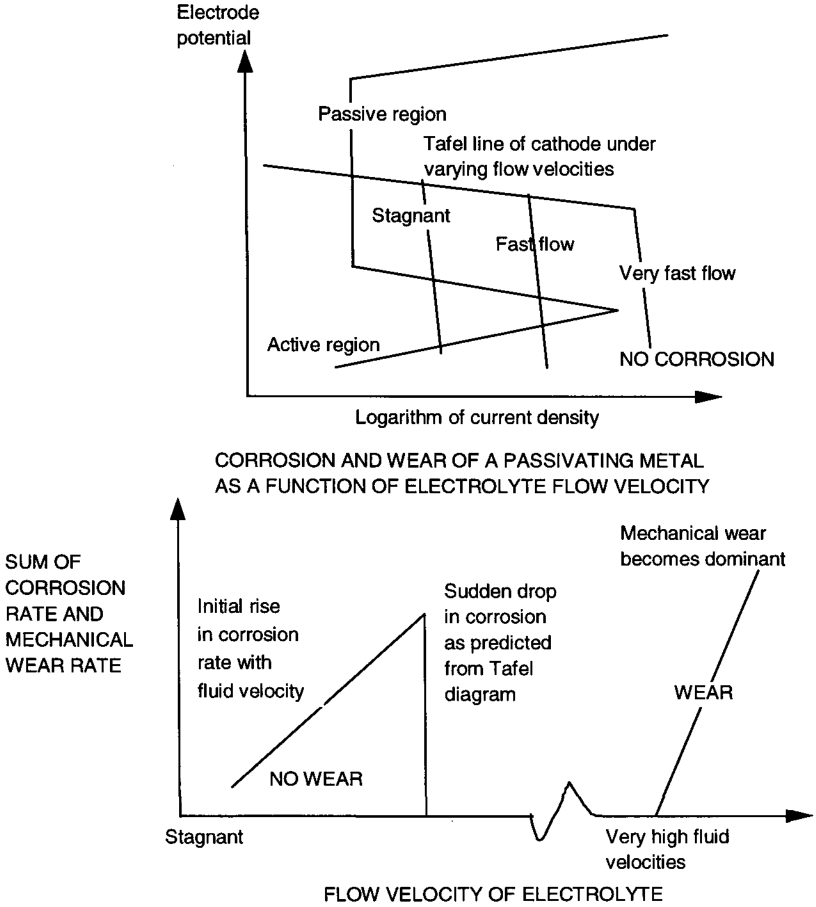


Figure 3.24. Tafel diagram illustrating passivation induced by fluid flow.(adapted from Fontana and Greene)

Passivators

Passivators are anodic inhibitors that shift the corrosion potential of the material or metal in the noble direction. The inhibitors can also aid uptake of oxygen by the surface in question to form a protective film on the surface. These inhibitors deactivate the anodic sites on the material surface. They provide passivation by several mechanisms such as 1) Forming a passive film on the surface in question, 2) preventing adsorption of aggressive anions (ex. Cl^-) by adsorption of competitive anions on the surface and 3) repassivation or film repair in the event of film damage. The passivators can further subdivided in to Direct Passivators and Indirect Passivators depending on the functionality of the corrosion inhibitor.

Direct passivators are those, which react with the surface of the material in question and form passive film. Typical direct passivators include metal Chromates, Nitrites, Tungstates and Molybdates. For the protection of steels, zinc is often used as the base metal to form e.g., zinc chromate, zinc nitrate. Zinc is readily available and confers its own additional protection on steels. Chromates and Nitrites are considered to be the best in terms of corrosion inhibition but it should be noted that metal chromates are highly toxic. These passivators are also called oxidizers.

Indirect passivators are not oxidizers but improve the adsorption of dissolved oxygen on to the surface to enable oxygen to actually passivate the material. Most commonly used compounds are either organic or inorganic in nature and these work only if dissolved oxygen is present in the environment. These passivators create high OH^- concentrations, where OH^- reacts with the H^+ ions in order to facilitate oxygen adsorption. Alkaline compounds are usually used such as, sodium hydroxide, sodium phosphate, Na_2HPO_4, Na_2SiO_3 and $Na_2B_4O_7$ (borax).

One of the important factor to be ensured with the use of passivation inhibitors is the concentration of the inhibitor should be maintained above the critical level everywhere in the system and so is not applicable to systems containing dirt, crevices, rust or lime scale in them.

Barrier Inhibitors

The second classification of inhibitors works on the principle of forming a protective layer on the surface separating the environment from the material. These are known as barrier inhibitors and form a protective layer between the material and environment by forming an adsorbed film or precipitating a compound scale on the surface. Based on the type of inhibitor used it can be subdivided in to three categories namely,

1. Organic Adsorption Inhibitors
2. Inorganic Precipitation Inhibitors and
3. Vapor Phase Inhibitors

Of the above inhibitors the organic inhibitors and vapor phase inhibitors fall in to the classification of adsorption film former and the inorganic inhibitor as indicated precipitates a compound scale on the surface.

Organic Inhibitors

This type of inhibitors forms a thin oily coating on the surface thereby protecting the surface from the environment. Organic inhibitors are used in acid environments where the acidity is a process requirement, such as a pickling bath. Organic inhibitors adsorb on to the surface by two mechanisms

1. Physical adsorption and
2. Chemisorption

Physical adsorption is limited to electrostatic attraction between a monolayer of polar organic molecules and the oxide film on a metal surface. Chemisorption involves bonding and charge transfer between the polar end of the organic molecule and the oxide anion or metal cation of the metal oxide layer. The chemisorption process is slower compared to the physical adsorption and most commonly used inhibitors include organic sulfides such as mercaptans, aromatic and heterocyclic amines, which have a dipole in the molecule. One of the problems associated with organic inhibitors is that they stimulate corrosion if not added in sufficient concentration. Another problem is the noxious nature of some of the organic compounds, some of which release intense odours (e.g., the mercaptans).

Inorganic Precipitation Inhibitors

Inorganic precipitation inhibitors react with the environment/ metal surface to precipitate insoluble products that give general coverage to the metal surface. The most commonly used inorganic inhibitors include those of the carbonate and phosphate salts, which form insoluble carbonate or phosphate films to protect the surface.

Vapor Phase Inhibitors

The vapor phase inhibitors are used to protect critically machined parts or precision mechanical instruments from corrosion. This type of inhibitor is released in small quantities in the surrounding atmosphere so as to form a condensed film on the critical parts to protect them against corrosion. For example amines are injected in to sour gas pipelines to react with one of the carrier gases to form protective sulfides or nitrites on steel. Vapor phase inhibitors are also useful in the protection of copper and bronze artefacts from atmospheric corrosion. Organo-nitrogen compounds such as benzotriazole and indazole are considered to be among the most effective protectors of copper and bronzes. There is a wide range of vapor phase inhibitors for steels, Guanadine chromate dicyclohexylamine benzoate is an example of an organo-metallic compound that is effective at very low vapour concentrations. The minimum vapour concentration for corrosion protection should be as low as possible to protect personnel and to conserve the inhibitor.

Cathodic Poisons

Poisons are another class of corrosion inhibitors that are often termed as cathodic passivators and inhibit reduction of hydrogen and oxygen at the cathode. This is achieved by addition of arsenic or phosphorous or antimony or bismuth compounds to restrict formation of molecular hydrogen at the cathodic sites. One of the major disadvantages of these cathode poisons are that these are toxic to humans and can cause hydrogen embattlement of blistering in metals. Oxygen reduction can also be poisoned by these inhibitors, in a similar manner to hydrogen reduction.

Scavengers

Scavengers are compounds, which react with the dissolved oxygen in the system and thereby eliminating them from the closed system. The scavengers are either neutral or alkaline compounds and include sulfites and phosphites. These sulfites or phosphites react with the dissolved oxygen in the closed system and form sulfates or phosphates and deprive of oxygen for cathode reaction thus stopping corrosion. Another scavenger used for high pressure and high temperature environments (e.g., boilers) is Hydrazine, which reacts with the oxygen to form nitrogen and water. These scavengers are not effective in acidic environment since they do not remove hydrogen ions.

Neutralizers

Neutralizers are compounds that reduce the acidity of a corrosive medium (the concentration the hydrogen ions), most neutralizers are alkalis which are added in sufficient quantity to remove most of the hydrogen ions.

3.2.4 Materials factors in aqueous corrosion

A major part of materials engineering involves the mixing of different materials to produce hybrid materials with superior properties. The physical scale of mixing inside the material structure can be of the order of atomic dimensions to produce metallic alloys. At a larger scale of mixing ranging from micrometers to several millimeters, composite materials are produced. Both alloys and composites involve the juxtaposition of materials with different electrochemical potentials. The unintended formation of innumerable electrochemical cells within the hybrid material is therefore inevitable. Alloys and composites are often subject to more intense corrosion problems than pure metals because of the electrochemical cells they contain within their structure.

Alloys are often processed with extensive heat treatment or even heated unintentionally as is the case of material adjacent to a welded joint. Such heating can cause segregation of alloying elements to produce a heterogeneous microstructure where previously a homogeneous microstructure was present. Microstructural heterogeneity can generate additional electrochemical cells inside the microstructure. It can also expose areas of the surface of the component or specimen to inadequate protection since there will be areas of metal surface depleted of protective alloying elements. Stainless steels are a classic example of this form of thermal damage since the passivity of stainless steel is dependent on a minimum concentration 12% chromium. If the concentration of chromium declines to 11% or less at any location then localized corrosion can occur.

Alloys

Alloys in current use contain two basic metallurgical features, which are, solid solution phases of two or more metals with possibly a non-metal, e.g. carbon and a network of inclusions to suppress plastic flow. Both these metallurgical features can influence corrosion processes. The classic example of corrosion problems caused by solid solution phases is the *dezincification* of brass. Brass is an alloy of copper and zinc and is composed of copper-rich and a zinc-rich phase. Zinc is a more electropositive or reactive metal than copper so that *galvanic coupling* (formation of an electrochemical cell) is possible where the zinc or zinc-rich phase is preferentially corroded. Severe dezincification occurs in hot water or boiler systems where brass tubes are used to convey the heated water. The original brass microstructure is degraded by dezincification to a porous structure consisting mainly of copper and copper oxide. It appears that during dezincification, the zinc is progressively ionized and solubilized by the reduction of water into hydrogen gas and hydroxide ions. Zinc is sufficiently electropositive to cause slow reduction of water at room temperature. Copper is not sufficiently electropositive to react with water and so remains largely uncorroded apart from oxidation by atmospheric oxygen. Dezincification is divided into two categories, *layer-type* dezincification and *plug-type* dezincification, which refer to extensive uniform dezincification and localized dezincification respectively. The occurrence of preferential corrosion of certain alloying elements involves the atom-scale structure of a surface as represented by the Terrace-Ledge-Kink model described earlier. It is possible for atoms of the more electropositive element, e.g. zinc to be removed from one atomic terrace or as adatoms on one part of the surface while adatoms and terraces of the less electropositive element, e.g. copper, gather on another part of the surface. Dezincification and its fundamental mechanism are illustrated schematically in Figure 3.25.

Dezincification is suppressed by the selection of brasses with smaller zinc contents, avoidance of cuprous and chloride ions, reduction in temperature and alloying with small amounts of arsenic or other elements. Arsenic, antimony and phosphorous are believed to form a film on the corroding brass surface, which shields the zinc from further corrosion.

High strength aluminum alloys depend on a carefully prepared network of inclusions for their mechanical properties. Inclusions of copper aluminide, $CuAl_2$ are usually precipitated from a solid solution of copper in aluminum to obtain the required strengthening. A corrosion problem sharply declines until much higher over voltages are imposed where current rises again. This anodic current - over voltage characteristic is illustrated schematically in Figure 3.22.

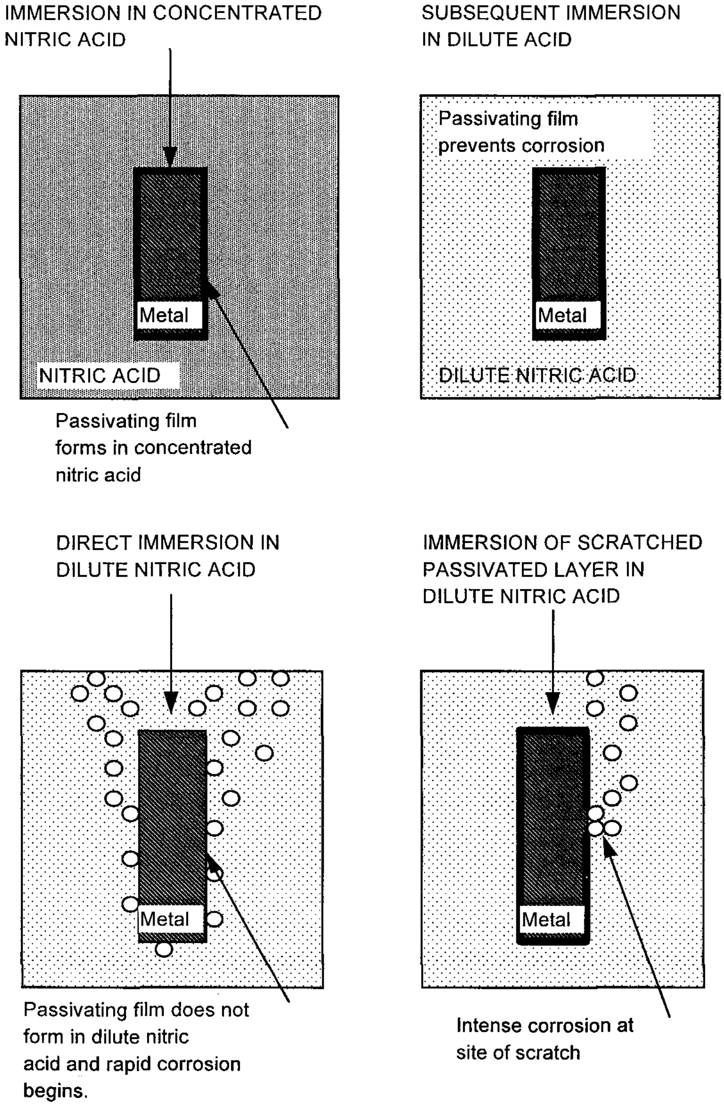


Figure 3.21 Faraday's demonstration of passivation of iron in concentrated nitric acid.

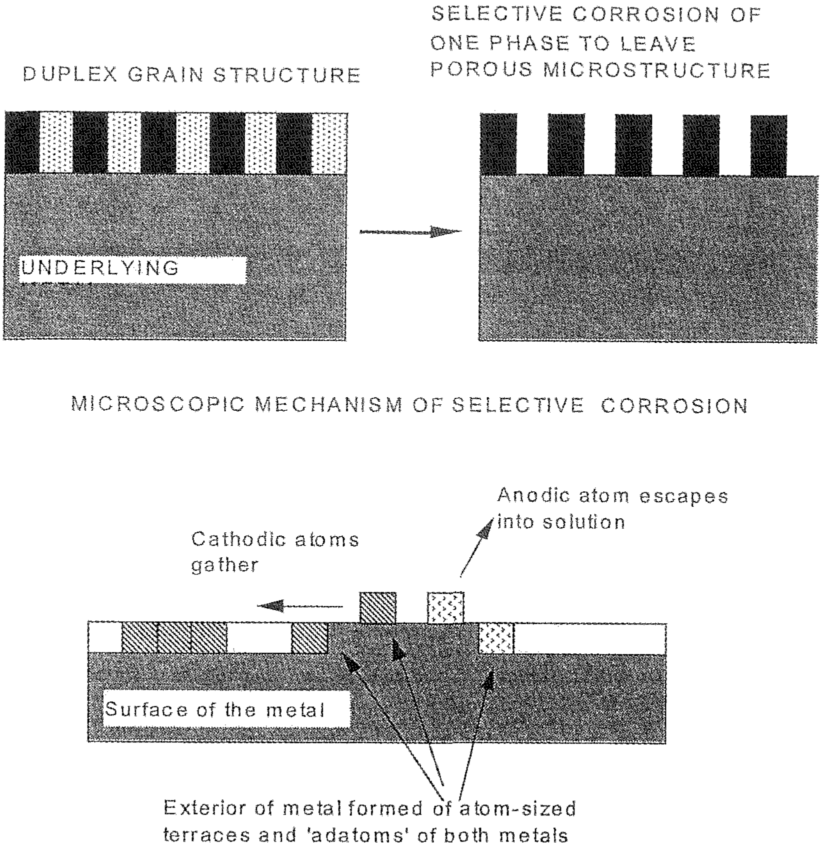
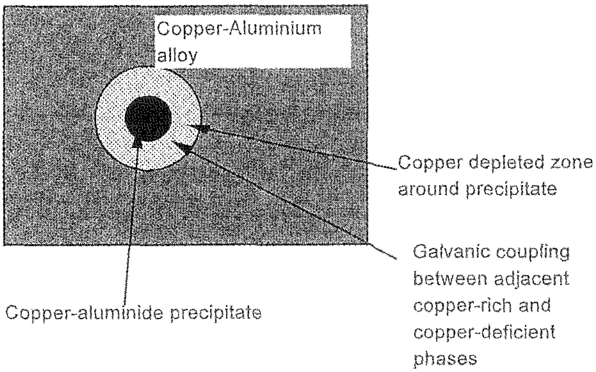


Figure 3.25 Dezincification of brass and its atomic mechanism.

occurs because the precipitation of copper aluminide causes the formation of regions of aluminum alloy matrix that are depleted in copper. Galvanic coupling then occurs between the copper-depleted aluminum alloy and unaffected alloy. The inclusions and adjacent copper depleted areas are located at the edges of the aluminum alloy grains. A form of differential corrosion known as *intergranular corrosion* then occurs where material close to the grain boundaries is preferentially corroded. Mechanical failure of the material then results due to the formation of weak corrosion products along grain boundaries. The mechanism of intergranular corrosion of aged copper- aluminum alloys is illustrated schematically in Figure 3.26.


GALVANIC CELLS FORMED BY PRECIPITATES IN ALUMINIUM ALLOY

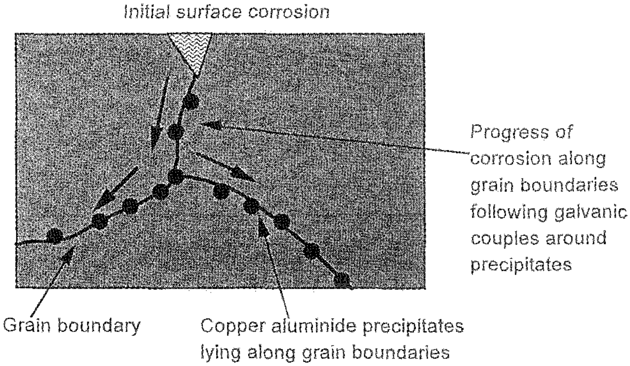

PROGRESS OF CORROSION BETWEEN PRECIPITATES

Figure 3.26 Intergranular corrosion mechanism of aged copper-aluminum alloys.

Substitution of copper aluminide precipitates with ferric aluminide, magnesium aluminide, magnesium silicide and magnesium zincide precipitates by appropriate alloying of the aluminum reduces the severity of intergranular corrosion. Ageing related problems of intergranular corrosion are also found in magnesium and copper-base alloys. Zinc alloys containing aluminum are also prone to this form of corrosion.

Not all inclusion formation in alloys is deliberate and may occur inadvertently where process heat is not adequately controlled. The corrosion resistance of stainless steels depends on a uniform or near-uniform concentration of chromium in solid solution with iron. Chromium is however a reactive metal and will precipitate from iron as chromium carbide if carbon is present in the iron. Welding of stainless steels is notorious for unintended heating of metal adjacent to the weld. Most of the chromium carbide formation occurs at a distance of a few mm away from the weld seam where the temperature reaches 700 to 800°C for a period of a few seconds. Areas of metal surface where significant chromium precipitation has occurred, lose their 'stainless' property and will corrode. When two plates of stainless steel are welded together and then immersed in corrosive acid, two linear zones of corroded metal then form parallel to either side of the weld seam Chromium carbides are normally precipitated at grain boundaries so causing a chromium depleted zone to form along the grain boundaries. Intergranular corrosion then proceeds along the chromium depleted zone. Remedies for *weld-decay* as this phenomenon is called, involve reheating the affected metal component to 1100°C followed by quenching to much lower temperatures. This practice causes chromium carbide decomposition to chromium and carbon at high temperatures without subsequent reformation. Other remedies are to lower the carbon content below 0.03% or add niobium and tantalum to the steel. Tantalum and niobium have stronger affinity for carbon than chromium and so prevent chromium carbide formation by combining with the carbon to form tantalum or niobium carbides. Tantalum and niobium are often referred to as *carbide stabilizers*. Niobium carbide is however decomposed by extreme heat particularly when the temperature reaches 1500°C or more as occurs in welding. Niobium carbide decomposition releases carbon for weld-decay directly beneath the weld-seam on the side of the sheet opposite to the weld. The released carbon then diffuses towards the chromium to cause chromium carbide formation. A highly localized form of corrosion known as *knife-line attack* then results. Knife-line attack is usually found during welding of thin sheet where cooling after completion of the weld is extremely rapid and the sheet is easily heated to melting temperature. The solution to this problem is to reheat the weld area to 1100°C to allow recombination of niobium and carbon.

Effect of work hardening

Metal alloys are not always used in the annealed state and work hardened alloys are important where higher strength is required. In general, work hardening accelerates corrosion because the higher levels of residual strain present in a work-hardened component are associated with a greater number of dislocations at the surface of the component. These surface dislocations act as favorable sites for corrosion reactions, which proceed faster as a result. The acceleration of corrosion reactions by plastic deformation of a material is known as *Mechanical Activation* and is observed for widely different types of materials and chemical reactions.

Composite materials

The most significant composite material in terms of quantity is reinforced concrete. In freshly solidified concrete, the steel reinforcing bars are protected from corrosion by the passivating effect of alkalis contained in the concrete. These alkalis are calcium hydroxide with smaller amounts of sodium hydroxide and potassium hydroxide; these cause the pH to reach 13 approximately. Long-term exposure to atmosphere causes this pH and consequent corrosion protection to decline since atmospheric carbon dioxide penetrates the concrete to react with the hydroxides and convert them to neutral carbonates. Exposure of concrete to salt water rapidly accelerates corrosion because the chloride ions in salt water quickly penetrate concrete to cause pitting of the steel reinforcement even under alkali conditions. The migration of chloride ions causes acidification as discussed above and severe corrosion results. The corrosion products of iron are less dense than iron or steel so that the volume of the un-corroded steel reinforcement bar and corrosion products increases with extent of corrosion. The increase in volume can only be accommodated by cracking of the concrete since concrete is brittle and weak in tension. Cracks allow more effective access by corrosive salt water so that corrosion becomes further accelerated. A destructive phenomenon, colloquially known as 'concrete cancer', then results.

Metal- metal composites, e.g. tantalum fibers in a copper matrix are also subject to significant corrosion problems. Galvanic coupling between tantalum and copper causes rapid corrosion of the copper matrix when the composite is immersed in a strong acid [Fontana and Greene]. To prevent severe corrosion of either the matrix or the reinforcing element, the electrochemical potentials of both metals should be as similar as possible. In practice this requirement imposes considerable limitations on materials selection and is one of the reasons why metal-metal composites have been largely substituted by metal-ceramic composites. Selective corrosion of either the matrix or reinforcement phase by a specific corrosive medium is also possible. This leads to a widening of the range of corrosive mediums that will attack the composite compared to either of the component metals alone. There is little or no scope for improving corrosion resistance, as is the case with alloying of metals. For example, with a copper-tantalum composite, copper but not tantalum is vulnerable to concentrated nitric acid while concentrated hydrofluoric acid attacks tantalum but not copper. A copper-tantalum composite is therefore vulnerable to both hydrofluoric and nitric acids [Fontana and Greene]. In order to render the copper-tantalum composite durable against hydrofluoric acid, it would be necessary to shield a core of composite material with an outer layer of copper only. This layered structure of a composite metal is analogous to the structure of concrete where steel reinforcement bars are protected from rusting by an outer layer of un-reinforced concrete.

3.3 Oxidative reactions of metals with oxygen, sulfur and other halogens

When a clean unrusted specimen of iron or steel is heated in air to red heat for several minutes, a *scale* or layer of reaction product will be observed on the steel surface after the specimen has been cooled down. Inspection of scale will reveal that it is uniform and appears to have some ability to cause passivation of the steel surface to limit further formation of the scale. If an attempt is made to scratch the surface with a screwdriver or chisel, it will become evident that the scale is far more brittle than the underlying steel. A corrosion process, i.e. chemical conversion of a metal to an inferior product can therefore be said to have occurred. This form of corrosion is often referred to as *oxidative corrosion* or *high temperature oxidation* of metals. Although, unlike electrochemical corrosion, which can occur at ambient or even sub-ambient temperatures, oxidative corrosion usually requires a high temperature for its occurrence, it is still regarded as a significant and severe form of corrosion. In certain cases, it is possible for oxidative corrosion to occur at ambient temperatures.

Among the chemical elements, oxygen is not the only known oxidant. Sulfur belongs to the same family of elements within the periodic table of elements and can in liquid or gaseous form initiate a directly analogous form of corrosion of metals. The halogens, fluorine, chlorine, bromine and iodine can also cause related corrosion which is often more severe than that caused by atmospheric oxygen.

Under dry conditions, the oxidation of most metals is prevented by the formation of adherent solid oxide films. Some metals have volatile oxides or more often sulfides and chlorides. These volatile corrosion products, which may be liquid, do not protect the metal surface against corrosion. Other metals such as sodium, potassium and magnesium produce a friable oxide, which continuously cracks to expose fresh metal surface to the air. For metals such as iron, chromium, copper, aluminum and other important engineering metals however, the oxide films are strong and durable. These oxide films range from a few nanometers in thickness at ambient temperatures to many micrometers in thickness at temperatures exceeding several hundred degrees Celsius. Very often, the oxide film reaches a limiting thickness after a period of initial rapid growth so that the oxide film can be said to have a characteristic thickness for practical purposes even though the oxide film does continue to grow at an infinitesimally slow rate. Where an oxide film is as thin as 5 nm, it is transparent so that the underlying metal can reflect light to produce the characteristic luster of a metal. When the oxide film thickness reaches 100 to 1000 nm, the oxide film becomes comparable in thickness to the wavelength of visible light. Interference between light reflected off the oxide surface and light reflected off the underlying metal creates colors. This is the reason for the tempering colors of polished quenched steels. Such thin oxide films are not usually a cause of direct corrosion problems although they exert a controlling influence on friction and wear phenomena of metals. Sulfide, chloride and other corrosion product films are believed to show similar characteristics to oxide films.

Whether a metal is protected from oxidative corrosion depends almost exclusively on the characteristics of the oxide film and its development. Basic chemical parameters of the metal such as electrochemical potential or heat of oxidation exert far less influence on oxidative corrosion than is observed for electrochemical corrosion. A metal will escape severe oxidation if the reaction product film is able to effectively exclude oxygen from the underlying metal. Where there is an oxide film, which prevents corrosion, it is still possible to induce severe corrosion by circumventing the oxide film. For example, aluminum is under typical conditions highly resistant to atmospheric oxidation. If a small quantity of mercury is placed on the aluminum and the portion of the aluminum surface is then scratched, rapid oxidation of the aluminum will result. Flakes of aluminum oxide will then form on the mercury and both metals will become heated by the exothermic reaction of oxidation. The reason for this phenomenon is that the mercury dissolves aluminum, which diffuses to the surface of the mercury. Oxidation of the aluminum then occurs without the protection of an aluminum oxide film, which cannot form on the surface of liquid mercury.

Another feature of metal oxidation, which is often overlooked, is that the heat of oxidation is usually dissipated in a metal structure that is much larger than an oxide film. At room temperatures, oxide films are usually of the order of 10 nm thick, while metal components have wall thickness of 1 mm or more. At temperatures of 1000°C, oxide scale thickness easily exceeds 10 μm so that the exothermic heat release becomes significant compared to the thermal capacity of the underlying metal. In such circumstances, a self-accelerating oxidative reaction is possible leading to a process that resembles burning. Such cases of large scale burning of metals have been reported in severe ship fires where the aluminum superstructure of the ship was observed to catch fire. At lower temperatures, with the continuing drive towards miniaturization of components, the need to examine the thermal capacity of the corroding metal becomes increasingly important. Electrical conductors consisting of very fine wires are a direct example of this problem. Uncontrolled oxidation by circumvention of an oxide film using mercury and heating of the underlying metal substrate are illustrated schematically in Figure 3.27.

Prevention of uncontrolled oxidation of metals dictate that mercury should not be transported in metal containers and that thin metal films coated on plastic substrates to make mirrors or lenses should be carefully protected against oxidation. Control of oxidation is a reason why gold , an inert metal, is often used for optical coatings.

Diffusional theory of oxidative corrosion

When a clean metal surface is first exposed to atmospheric oxygen or another oxidant such as liquid sulfur, the formation of a corrosion product is mainly controlled by the kinetics of chemical reactions between two materials in close contact. It is necessary for the oxygen (or other oxidant) to absorb on the metal surface, electron transfer to occur followed by bonding between the metal and oxidant. These processes are relatively rapid for most metals so that a thin monomolecular layer of corrosion product can cover a metal surface in a minute period of time. Under atmospheric pressure of oxygen, this period of time is of the order of microseconds for iron and steel. Unlike aqueous electrochemical reactions, the corrosion product is not solubilized and removed from the reaction surface by the corrosive medium. If the corrosion product is solid, it will instead accumulate on the surface as a film, the most common example being the oxide films discussed above. The film of

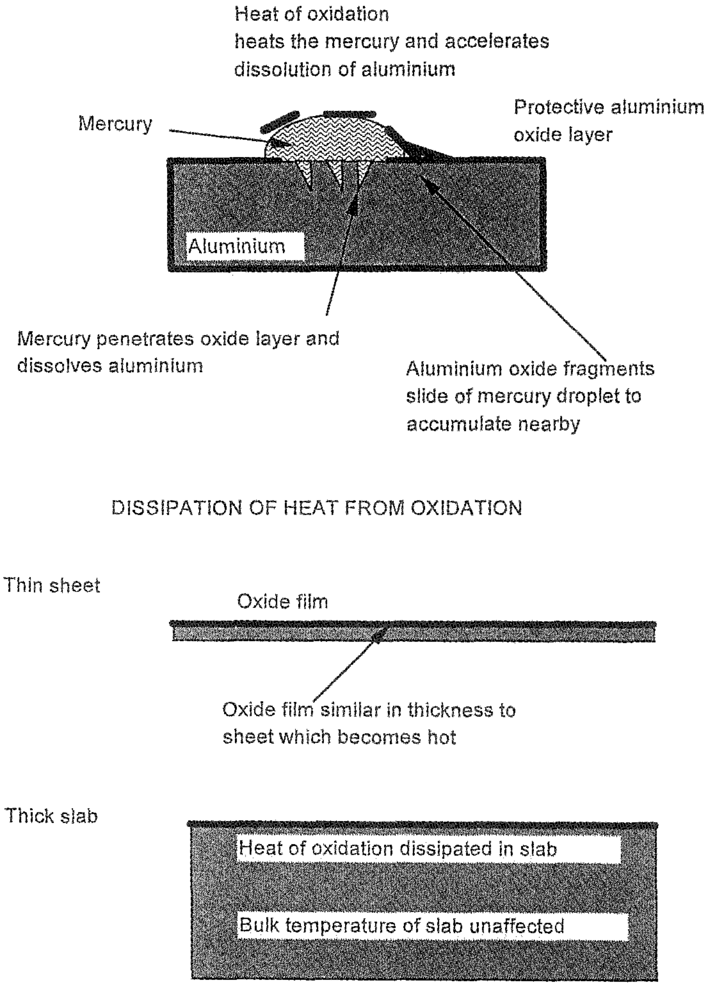


Figure 3.27 Mechanisms of uncontrolled oxidation of metals in atmospheric oxygen.

corrosion product exerts a fundamental influence on metal oxidation once a monomolecular film has been established on the surface of the metal. The presence of a corrosion product film alters every stage of the process of oxidation. For example, oxygen molecules or other oxidants can only absorb onto the corrosion product surface not the metal surface while electron transfer involves transport of electrons though the corrosion product film. It is also necessary for either ions of metal or ions of the oxidant to diffuse through the corrosion product film before bonding between the metal cations and corresponding anions can proceed. While the film thickness increases from just one monomolecular layer of corrosion product to several monomolecular layers in thickness, electron tunneling and spontaneous exchange of positions between metal and absorbed oxygen permit rapid growth of the corrosion product film to continue. This stage of growth allows the rapid formation of oxide films to thickness of 5 nm over a very wide range of temperature. Electron tunneling and the mutual change in location between oxygen and metal atoms is called *place exchange*. These are limited to very thin oxide film thickness of only 5 nm and beyond this thickness, diffusion of cations or anions become the controlling process for most corrosion product films. The processes of adsorption of e.g. oxygen molecules, their dissociation and ionization, the formation of a layer of metal cations at the interface between metal and corrosion product are not usually rate limiting. Under typical conditions therefore, corrosion rate is controlled by the diffusion of either anions or cations across the corrosion product film under the influence of a concentration gradient and the electric field developed between opposing layers of cations and anions. Initial oxidation process and subsequent control of oxidation by diffusion within the corrosion product film are illustrated schematically in Figure 3.28.

It is possible to derive an expression for corrosion product growth rate based on the assumptions of a uniform surface film with no voids between metal and corrosion product and no fissures in the surface film. The voids would hinder diffusion transfer by acting as an obstacle while fissures provide more direct access by the oxidant to the metal. The theory of oxide film growth involves the derivation of an expression for the flux of metal cations or oxidant anions through the corrosion product film where this flux is equivalent to the growth rate of the surface film. The flux of cations or anions is a function of Fick's law diffusion and the electrostatically driven movement of ions between vacancies in the defective crystalline lattice of the corrosion product. The corrosion product is very often non-stoichiometric in composition, i.e.. for iron oxides, the precise composition may be $Fe_{2.21}O_3$ instead of the stoichiometric composition Fe_2O_3. The non-stoichiometry causes the crystalline structure of the corrosion product to be highly defective, i.e. that there are numerous vacancies in its structure. The flux of cations or anions across the corrosion product is largely a result of 'hopping' of ions between vacancies. For corrosion product films up to 1 μm in thickness, electrostatically driven ion movement is a significant component of the

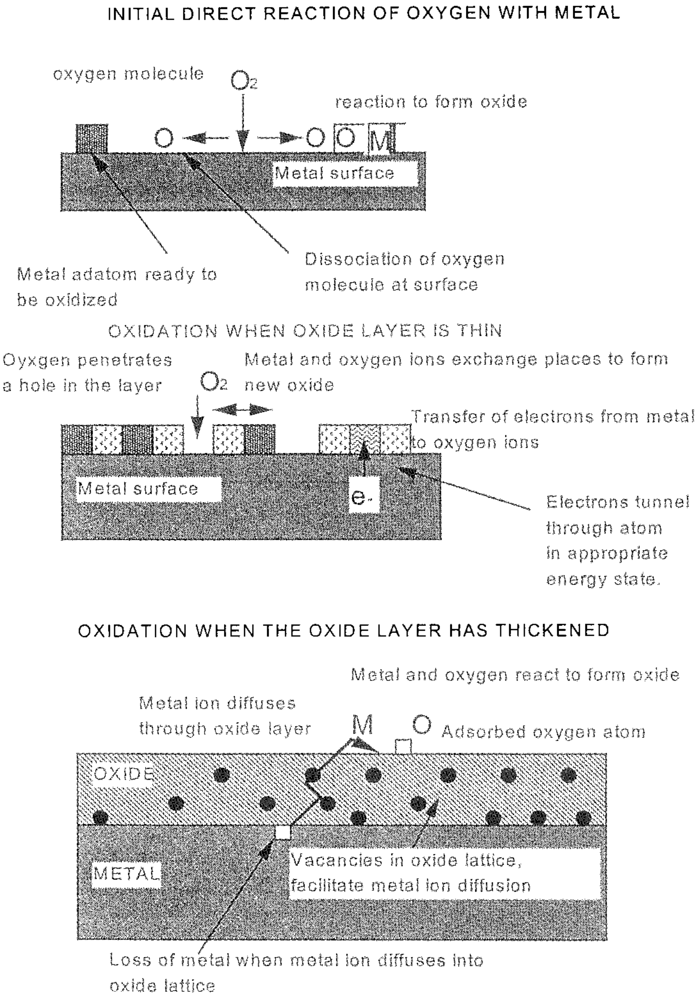


Figure 3.28. Corrosion product film formation on a metal surface. Initial direct oxidation with the metal and subsequent corrosion product film formation.

ionic flux across a corrosion product film. Analysis of corrosion problems involving film thickness less than 1 μm should allow for this characteristic of film growth. Beyond approximately 1 μm, Fick's law diffusion provides almost all of the ion flux. This feature of surface film growth simplifies considerably the theoretical modeling to allow the derivation of a very simple growth law known as the Wagner law or parabolic film growth law. The Wagner law of film growth states that the rate of film growth is proportional to a uniform linear concentration gradient of diffusing ions across the surface film. The concentration gradient is equal to the difference in concentration of diffusing ions (either cations or anions) between the surface of the un-corroded metal and the exterior surface of the corrosion product film. The concentration of diffusing ions at either the exterior surface of the corrosion product film or at the surface of the un-corroded metal is assumed to be constant. These assumptions inherent in the Wagner model of film growth are shown schematically in Figure 3.29.

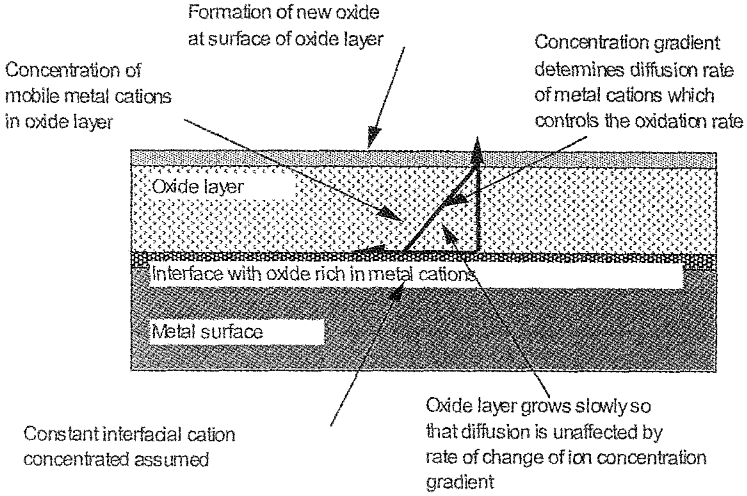


Figure 3.29 Wagner model of growth of corrosion product film by diffusion.

All of these assumptions about film uniformity, invariance of ion concentration are only an approximation to the actual more complex situation inside corrosion product films. In many cases however, the quality of approximation is good so that the Wagner model provides a useful prediction of surface film growth. The controlling equation of film growth according to the Wagner model is given by the following expression.

$$\delta D/\delta t = K/D$$

$$D^2 = K^*.\ t + C \qquad \{3.9\}$$

Where *D* is the corrosion product film thickness, *t* is time; *K* is a film growth coefficient, which is the product of the concentration difference in diffusing ions across the film and the diffusivity of ions within the corrosion product film. C is a constant, which is usually assumed to be zero since film thickness is nominally zero initially. A positive value of C may be assigned empirically to fit film growth data where electrostatically driven ionic flux has provided a significant initial boost to film growth.

The physical significance of the Wagner law of film growth is that it represents a minimum growth rate of the corrosion product film to which other growth mechanisms such as electron tunneling and electric field effects add on additional amounts of film growth. Electron tunneling and electric field effects are significant at low to moderate temperatures and cause the initial rapid growth in oxide film thickness to some apparently limiting value which is actually very slow oxide film growth. The Wagner film growth coefficient, K, attains large values at high temperatures to cause rapid film growth or corrosion. Parabolic or Wagner oxidation is most significant at high temperatures and is the prime cause of high temperature oxidative corrosion of metals. This rise in K is due to the activation energy of ionic diffusivity, which shows an Arrhennius law dependence on temperature. The difference in growth characteristics of parabolic (Wagner) oxidation and other forms of oxidation such as Low Temperature Oxidation is shown in Figure 3.30. Wagner or parabolic oxidation can be rapid but not as rapid as linear uncontrolled oxidation or uncontrolled burning of metals, which are also shown in Figure 3.30. The tendency of iron and steels to oxidize according to the Wagner law and not show a linear growth rate is a reason why iron and steel can be used without protection at high temperatures in air.

Most but not all oxide films tend to have lower ionic diffusivities than the corresponding sulfide films of the same metal. Higher ionic diffusivity in sulfides as opposed to oxides of the same metal, e.g. iron sulfide and iron oxide, is caused by the greater non-stoichiometry of sulfides. Non-stoichiometry is characterized by a defective crystalline structure with numerous vacancies for ionic diffusion. Sulfidation of metals such as iron is more rapid and therefore a more serious form of corrosion than oxidation by atmospheric oxygen at the same temperature. Direct experimental measurements of ionic diffusion coefficients are generally lower than the corresponding estimates of ionic diffusivity from film growth coefficients. This discrepancy in ionic diffusivities implies preferential paths for ion diffusion within the corrosion product film and

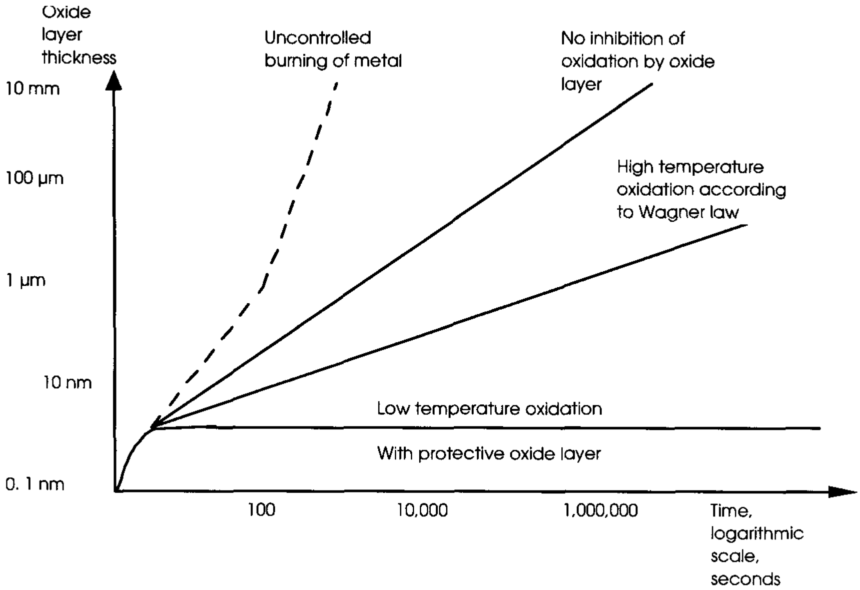


Figure 3.30. Growth characteristics of oxide films on metal surfaces.

it is generally believed that grain boundaries inside the corrosion product film facilitate rapid ion diffusion. Alloying of metals to produce corrosion products with a less defective crystalline structure and therefore low ionic diffusivity is currently being tested as a method of controlling corrosion.

3.3.1. Durability of corrosion product film and failure of Passivation

The principle constraint on metal oxidation is the passivating effect of the layer of corrosion product that forms on a metal surface. The durability of the corrosion product layer effectively controls the service life of the metal. Durability of the corrosion product layer depends on its mechanical and physical stability. The main causes of instability are mechanical failure and melting or volatilization. Although most theories of oxidation relate to attack of a metal by a single oxidizing agent e.g. oxygen or sulfur, combined attack by several oxidizing agents is much more common. In general, the combination of oxidizing agents intensifies corrosion by preventing a durable corrosion product layer from forming. Gases, which are not usually considered to be strong oxidizing agents, e.g. carbon dioxide, are found to cause significant corrosion of metals at elevated temperatures. Any combination of gas (except the noble gases such as argon) and metal should be tested experimentally for corrosion before it is assumed that no corrosion will occur. Corrosion does not always occur at the surface but may proceed into the interior of the metal. This *internal oxidation* causes brittle inclusions of corrosion product to form inside the metal thus causing embrittlement and fracture.

Mechanical detachment of corrosion product

There is no physical necessity that a corrosion product remains adhering to the underlying metal surface to passivate the metal even if it remains solid at elevated temperatures of oxidative corrosion. When a corrosion product film first forms on a corroding metal surface, the ions in the crystalline lattice of oxide or other corrosion product attempt to fit to the metal crystal lattice. The crystalline lattices of the metal and corrosion product are entirely different so that the crystalline lattices of both are strained. This phenomenon of mutual crystal lattice adaptation between metal and film is known as *epitaxy* and the resulting mechanical strains are known as *epitaxial strains.* Unless the crystal lattice of the corrosion product film is able to progressively accommodate the epitaxial strain as the film thickens, the interfacial stress between the corrosion product and the metal increases with film thickness until fracture results. This is because the fixed number of atom bonds at the metal-oxide interface has to compress an increasing number of atoms in the oxide film. A corrosion product film can remain attached to the underlying metal and provide passivation until a critical thickness is reached. Beyond this critical thickness, the corrosion product film will detach to initiate a new cycle of formation of corrosion product. In the case of magnesium, the oxide film does not detach completely, instead film material lying above a certain thickness is cracked and fissured to accommodate epitaxial stress. The cracks and fissures allow unrestricted access by atmospheric oxygen to a constant depth of oxide film. The oxide film therefore grows at a linear rate after an initial period of parabolic oxidation according to the Wagner model. This causes indefinite rapid oxidation instead of quasi-passivation according to the Wagner model. Epitaxy and its effects on corrosion are illustrated schematically in Figure 3.31.

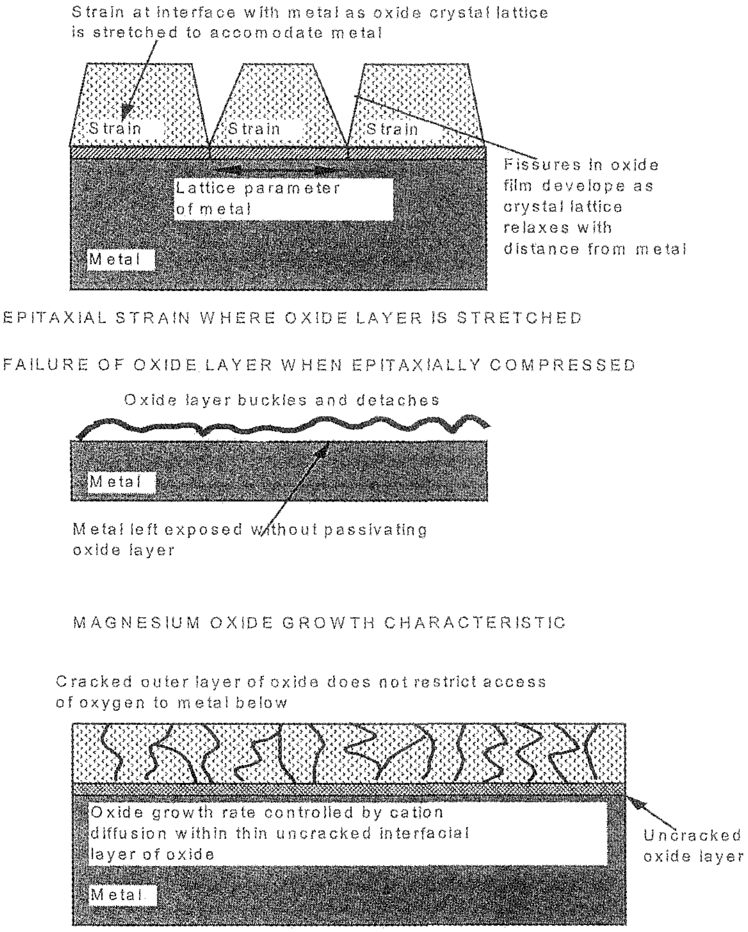


Figure 3.31 Effect on corrosion of epitaxy between a metal surface and overlying corrosion product film.

Even if there is a perfect fit between the crystal lattices of the metal and corrosion product, the dissipation of momentum by cations or anions diffusing through the corrosion product film causes a reaction stress according to Newton's 2nd law. In most cases, corrosion products grow by the outward diffusion of metal cations, which causes a tensile stress while anion diffusion causes a compressive stress. Such stresses are proportional to film growth rates hence rapidly growing films are inherently more highly stressed than self-passivating films. This growth rate dependency of ion diffusion induced stress enhances a tendency for corrosion products to differentiate between forming mechanically stable films that are self-passivating or forming rapidly growing and mechanically unstable films.

Most work on strains in corrosion product films has related to oxides as opposed to sulfides, chlorides and other corrosion products. One of the earliest theories about strains in oxide films and how the strain affects the durability of the oxide films was developed by two researchers, Pilling and Bedworth and published in 1923. This model was developed before metal cation diffusion was recognized as the most common controlling factor in oxide film growth. Pilling and Bedworth postulated that if the oxide film grew by oxygen penetrating the oxide film to form new metal oxide at the interface between metal and oxide, then the oxide was volumetrically constrained to fit the underlying metal. In other words, if the oxide volume were too small then the oxide film would be in tension and probably crack, as oxides are usually brittle. If the oxide volume were too large, the oxide would eventually buckle and detach from the surface. Pilling and Bedworth devised a ratio of the molar volume of oxide to the molar volume of metal as the criterion for tensile or compressive stresses in the oxide film. An expression for this ratio, which is known as the Pilling-Bedworth ratio, is given below.

$$\mathrm{PB} = (Mox.^{*}.\ \rho met)/(Mmet.^{*}.\ \rho ox) \qquad \{3.10\}$$

Where PB is the Pilling-Bedworth ratio, *Mox.* is the molar weight of oxide (or other corrosion product), *ρmet.* is the density of the corroding metal, *Mmet.* is the molar weight of the corroding metal and *ρox.* is the density of the oxide (or other corrosion product).

As can be seen, the subtleties of epitaxy and crystalline lattice structure are overlooked by this model, which gives an approximate guide to the stability of oxide films. According to the model, Pilling-Bedworth ratios close to unity indicate stable oxide films while values differing significantly from unity are unstable. Experimental data shows that while most oxides conform to this rule there are numerous exceptions. For example cadmium, which has a non-protective oxide film, has a Pilling-Bedworth ratio of 1.21 while the corresponding value for chromium, which has a protective oxide film, is 1.99. Experimental measurements show that if the value of the Pilling-Bedworth ratio is significantly below unity as it is for sodium, potassium and lithium or greater than two as is the case for tungsten then the corresponding oxide film is non-protective. When the value of the Pilling-Bedworth ratio for a specific metal and oxide lies between 1 and 2 then there is no reliable indication of oxide film durability. Estimates of the stability of sulfides and other corrosion products as surface films can be obtained from the Pilling-Bedworth ratio by substituting the density and molar weight of the oxide with the density and molar weight of the corrosion product.

Melting and volatilization

When the melting point of the corrosion product is exceeded, the corrosion product film will continuously flow away to offer no protection to the underlying metal. A thin liquid film of corrosion product then remains on the surface of the metal. Diffusion of ions through this liquid is very rapid and presents no limitation to corrosion unlike diffusion through a solid film. Very rapid oxidation known as *catastrophic oxidation* will occur if the corrosion product liquefies. Catastrophic oxidation may also involve self-accelerating oxidation driven by large amounts of energy from the exothermal oxidation reaction (see also the beginning of **3.2**). Fortunately, most oxides have a melting point comparable to or higher than the corresponding metal, the classic example being aluminum and aluminum oxide. However the oxides of some metals have lower melting points than the parent metal. If the melting point of the oxide is lower than the metal, then rapid oxidation by liquefaction of the oxide is possible. Figure 3.32 illustrates schematically the mechanism of catastrophic oxidation of a metal when the oxide is molten or evaporating as a volatile vapor.

The oxides of molybdenum (MoO_3) and of vanadium (V_2O_5) have melting points of 795°C and 674°C respectively, these temperature are lower than the melting point of molybdenum. Molybdenum trioxide (MoO_3) also forms a eutectic with molybdenum dioxide at 778°C. These metals and their alloys cannot be used for high temperature service in air but molybdenum does have good high temperature mechanical properties, which renders it suitable for high temperature vacuum service. Sulfides tend to have lower melting points than oxides of the same metal, for instance iron and FeS form a eutectic at 985°C and nickel

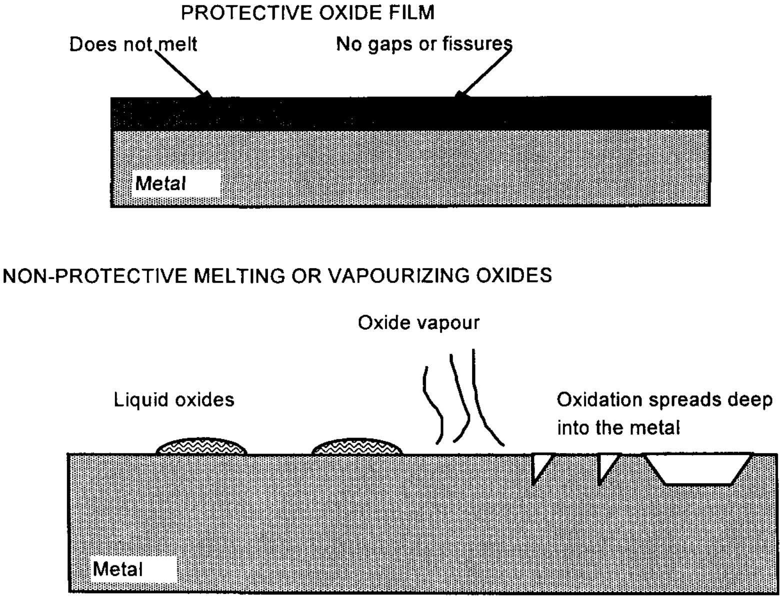


Figure 3.32 Catastrophic oxidation of a metal when its oxide film liquefies or evaporates.

forms a eutectic with Ni_3S_2 at 635°C. For iron and nickel, exposure to sulfur vapor beyond 985°C and 635°C respectively causes very rapid corrosion.

Rapid corrosion by chlorine containing gases, e.g. hydrogen chloride is also associated with volatilization of the metal chlorides formed. For example, ferrous chloride ($FeCl_2$) melts at 600°C and volatilizes at low temperatures. The chlorides of nickel and chromium ($NiCl_2$, $CrCl_2$ and $CrCl_3$) are comparatively less volatile than those of other metals. The lack of volatility enhances the resistance of these metals towards chlorine and hydrogen chloride.

In many cases, oxidation is not caused by pure oxygen but instead proceeds in a much more polluted environment. Fuel oils contain varying amounts of vanadium, which often becomes concentrated in residual oils from petroleum refining. The accepted method of disposal of such waste oils is to burn them. During combustion of the waste oil, the vanadium is oxidized to V_2O_5, which then forms a eutectic with sodium oxide, Na_2O, at 535°C. Sodium compounds, e.g. salt, are ubiquitous contaminants. Combustion temperatures easily exceed this eutectic and the molten liquid forms a flux to liquefy the oxides found on the metal components (usually steel) of the combustion system. Rapid corrosion of the combustion system then results.

Internal oxidation of metals

Internal oxidation is found in alloys of metals where a lesser amount of a base (reactive) metal is present in a more unreactive or *noble* metal. Preferential oxidation of the basic metal then proceeds from the surface of the alloy until a depth limited by access of oxygen. An external scale does not necessarily form and instead oxides are precipitated in the subsurface of the alloy. Alloys often possess a duplex microstructure where one of the phases is likely to have a greater solubility and diffusivity for oxygen than the other phase. Internal oxidation then proceeds rapidly along grains of one phase while the other phase remains relatively untouched. The mechanism of internal oxidation is shown schematically in Figure 3.33.

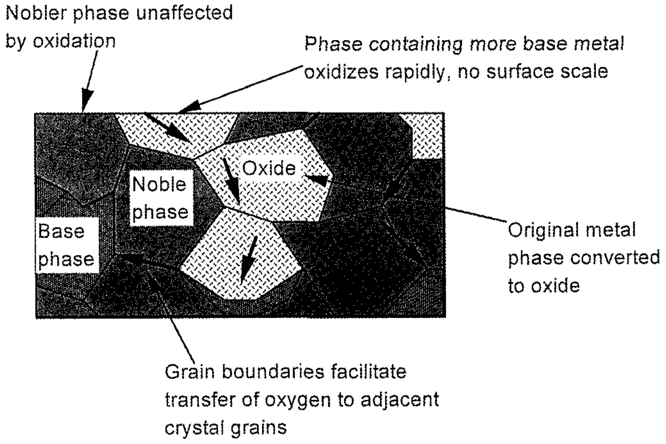


Figure 3.33 Mechanism of internal oxidation of metal alloys.

Examples of internal oxidation are found in cuprous and silver-base alloys, which contain basic metals such as beryllium, zinc or cadmium. Iron, nickel cobalt alloys are also affected by internal oxidation when small additions of chromium and aluminum are present. Chromium and aluminum are more reactive than iron and nickel and depend on a durable oxide film for passivation. Where the oxide film is unable to form because the concentration of aluminum or chromium in the alloy is too low, internal oxidation occurs. Internal oxidation can be prevented in this instance by increasing the concentration of chromium or aluminum to a level where a stable chromium or aluminum oxide film can form on the alloy surface. Nickel alloys containing aluminum are particularly vulnerable to internal oxidation with deep rod shaped zones of internal oxidation forming in linear shaped grains of the aluminum-rich phase. Internal oxidation occurs in iron, nickel and cobalt alloys simultaneously with external scale formation. Oxide precipitation in metals can raise hardness by the same mechanism as age hardening but also causes an undesirable increase in brittleness of the alloy.

Effect of combined oxidizing agents

In most cases of corrosion, e.g. boiler chimneys and chemical processing pipelines, oxygen is either combined with sulfur bearing gases or chlorine bearing gases. Severe corrosion results in both cases thus necessitating the use of special corrosion resistant materials.

Corrosion by sulfur dioxide containing gases causes a localized form of high temperature corrosion that in common with internal oxidation involves deep penetration of the metal. A very common corrosion problem involves hot combustion gases, which contain sulfur dioxide and oxygen. The combination of sulfur and oxygen is more corrosive than either acting alone. Sulfides decompose to oxides in the presence of oxygen so that corrosion proceeds with an external scale of metal oxides lying above a layer of sulfidized metal. Under steady conditions, pits are formed where metal sulfides fill the bottom of the pit and with a covering of metal oxides. If the combustion system is subjected to alternating cycles of hot sulfur dioxide and hot oxygen, internal Sulfidation of the metal proceeds simultaneously with external oxidation. These phenomena are usually found on Ni-Cr and Co-Cr alloys, which are widely used in combustion systems.

Chlorine and gaseous chlorine compounds, e.g. hydrogen chloride, cause accelerated metal oxidation when present in combination with oxygen. The phenomenon is termed *activated oxidation* and is very destructive of iron or steel. Instead of an adherent stable oxide layer, a fragmented non-protective oxide layer is formed on the metal surface. Combined attack of iron by chlorine and oxygen causes the formation of both ferrous chloride and iron oxides at the metal surface. The ferrous chlorides diffuses through the oxide layer and decomposes to form iron oxide and chlorine when it reaches the outer surface of the oxide layer. The interface between the oxide layer and the underlying iron or steel becomes disrupted by voids as a result of the escape of ferrous chloride. These voids cause the oxide layer to detach from the surface and become non-protective.

Exposure of metals to hydrocarbon gases ,hydrogen and carbon dioxide

When steel is exposed to hydrogen at a high temperature, carbon within the steel can react with the hydrogen to form methane. This reaction causes a loss of carbon within the steel known as *decarburization.* The solubility of hydrogen in steel allows hydrogen to penetrate deeply into steel, methane formation then occurs below the surface of the steel and produce voids filled with methane in the steel. The development of voids and the decarburization cause weakening and embrittlement of the steel. Conversely, when methane or another hydrocarbon gas is exposed to steel at the same high temperature, the reverse reaction, i.e. decomposition of hydrocarbon to hydrogen and carbon will occur. If alloying elements with strong carbide forming tendencies are present in the steel, e.g. chromium, the alloying element will react with the released carbon to form a carbide layer. This phenomenon is known as *carburization.* The carbide layer is usually too brittle to improve mechanical properties of the steel component. Severe carburization where not all the carbon is combined as carbide but instead a fraction of the carbon remains as graphite, is known as *metal dusting.* This is a descriptive term used to describe the powdery degradation of metal in hot hydrocarbon gas streams, which are commonly found in the petrochemical industries. It is possible to suppress both carburization and decarburization by adding a small quantity of hydrogen sulfide to the hot hydrogen or methane. The hydrogen sulfide releases a small quantity of adsorbed sulfur on the surface which blocks adsorption by hydrogen or methane on the steel surface prior to reaction. The quantity of hydrogen sulfide required is insufficient to cause sulfidization but this method of control is still relatively untried. Another method is to further alloy the steel with silicon so that a silicon dioxide layer forms on the steel surface at high temperatures. This silicon dioxide layer is believed to be an effective barrier to penetration by hydrogen or carbon. Other barrier layers such as alumina have also been found to offer some reduction in carburization. The mechanisms of decarburization and carburization are illustrated schematically in Figure 3.34.

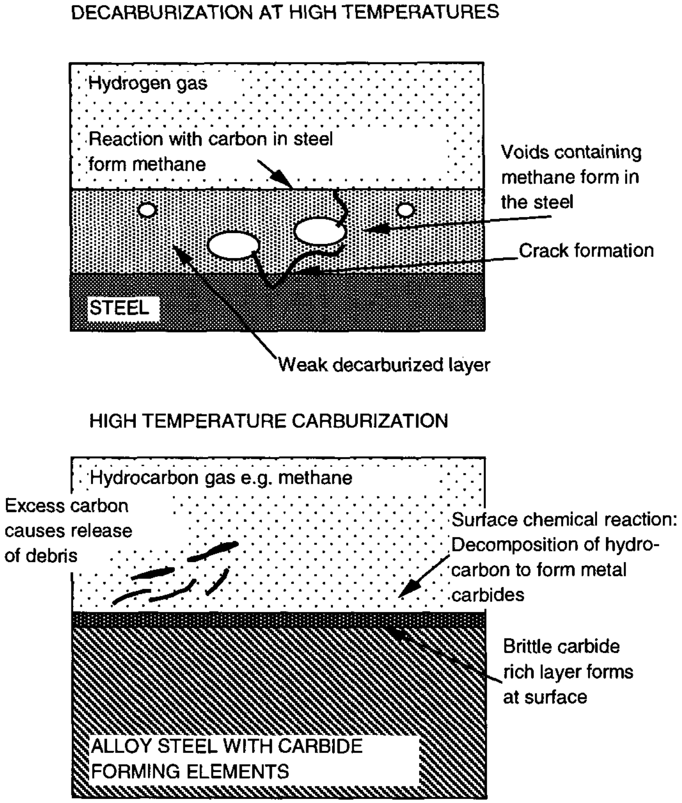


Figure 3.34. Mechanisms of decarburization and carburization.

Carbon dioxide is used as a coolant for nuclear power systems where it has the characteristic of promoting slow oxidation of steel according to the Wagner model for an initial period, which lasts for hundreds or even thousands of hours. After this initial period of slow oxidation, spalling on the oxide film is observed and fast linear oxidation occurs. Microscopic examinations of corroded specimens reveal that an outer layer of ferrous oxide (Fe_3O_4) is separated from the steel surface by a porous carbon rich layer where much of the carbon is present as carbides. Once this inner carbon-rich layer reaches a sufficient thickness, the outer oxide layer is unable to shield the underlying steel from oxygen and oxidation proceeds at a rapid rate.

Effect of mechanical stresses and strain on oxidation

Mechanical stresses and strains do not only accelerate aqueous corrosion but also accelerate the high temperature oxidation of metals. There are two main effects occurring. The first effect is of direct mechanical activation where a high level of surface dislocations promotes oxidation and lowers the temperature where limited oxidation (described by the logarithmic film growth law) changes to the parabolic oxidation growth law. For steels, work hardening reduces the minimum temperature for parabolic oxidation growth law to approximately 200°C. The second effect is that mechanical stress; particularly oscillating stress causes repeated fracture of the protective oxide layer. The so formed cracks facilitate access by oxygen to the metal and faster corrosion results. Stresses need not be necessarily generated by deliberately applied load but can also arise from restrictions to thermal expansion. Thus, thermal fatigue is associated with accelerated high temperature oxidation.

3.4 Softening and embrittlement of wood and polymers

Wood

Wood is highly resistant to most acids except concentrated strong acids such as fuming sulfuric acid, which can oxidize wood to carbon. Alkalis however rapidly degrade wood by chemical reaction with the hemicelluloses and lignin components of wood (for discussion of wood composition see below). Elevated temperature and high concentration of both acid and alkali promote more rapid degradation of wood compared to cool dilute solutions of acid or alkali. Softwoods are more resistant than hardwoods to alkalis because of the lower hemicelluloses content in softwoods. Pure water also causes strength loss by causing swelling of the wood.

Polymers

Most polymers are vulnerable to chemical attack by water and more corrosive media. Polymers vary widely in resistance to liquids, some polymers such as nylon will be degraded by water at room temperature while Ultra-High Molecular Weight Polyethylene UHMWPE is only corroded by hydrogen peroxide at temperatures close to 100°C. There are two basic forms of degradation (i) solvation and partial solvation of the polymer by a solvent which is usually organic (ii) chemical reaction leading to fracture of the polymer or in extreme cases conversion to a powder. Solvation and partial solvation is common with rubbers which when exposed to gasoline or lubricating oils will swell in size. This swelling is the first stage of solvation where solvent molecules from e.g. the gasoline, diffuse into the spaces between the elastomer molecular chains. With increasing aggressivity of the solvent or elevated temperature, the rubber is converted to a slimey product, which represents the stage where the solvent molecules are able to separate the elastomer molecular chains from each other. The mechanism of solvation both to a limited extent and extensive solvation is illustrated schematically in Figure 3.35.

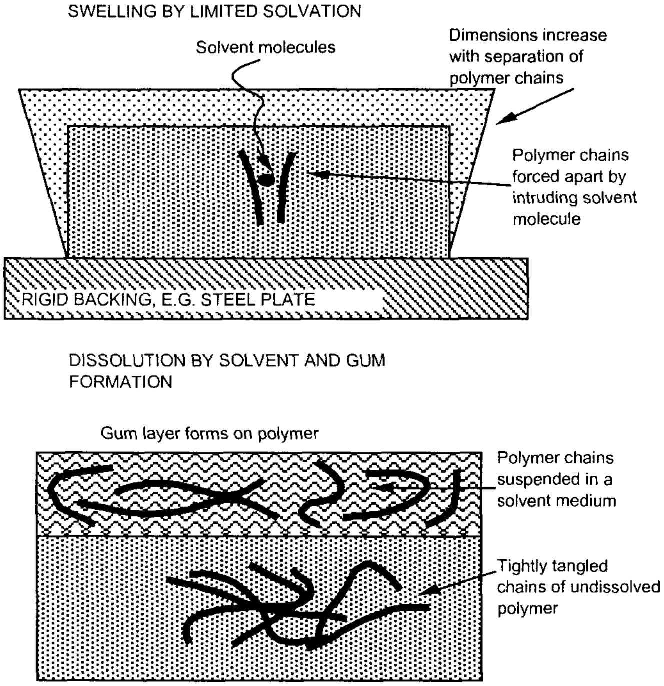


Figure 3.35 Solvation and partial solvation or swelling of polymers by a solvent.

Chemical reaction of a polymer with an aggressive reagent usually involves chain scission where the polymer chains are split to reduce the average molecular weight. One mechanism of chain scission is oxidative where the polymer chain is converted to a series of lower molecular weight oxidation products. The primary effect of such chemical degradation is an increase in brittleness of the polymer, which may lead to unexpected fracture or 'crazing'.

Crazing is the formation of a network of cracks over the surface of the polymer. In polymers, while the hydrogen-carbon bond is relatively stable against chemical attack, other bonds such as the amide bond between nitrogen and carbon are more vulnerable. This means that although a purely hydrocarbon polymer such as UHMWPE is relatively resistant against chemical attacks other more complex polymers such as nylon are much less resistant.

The bond between carbon and fluorine is however stronger than the carbon hydrogen bond which means that fluorocarbons such as polytetrafluoroethylene (PTFE) are extremely resistant to chemical attack. Other halides such as chlorine and bromine however do not form such strong bonds with carbon and so cannot be used to manufacture corrosion resistant polymers. In fact, organic chlorine compounds are used as sacrificial lubricant additives. The ends of polymer chains are the preferred reaction sites for many degradation reactions so that resistance to degradation increases with molecular weight of the polymer (molecular weight is inversely proportional to the number of chain-ends). Mechanical stress and wear are observed to accelerate chemical degradation where it is believed that the activation energies of degradation reactions are reduced by stress or wear.

Oxidation of polymers is usually inevitable since most polymers are used in open air. Oxidation proceeds by a chain reaction involving hydro peroxides as the intermediate compounds that accelerates the chemical reactions of oxidation. Oxidation tends to produce more peroxides so that oxidation of polymers tends to be self-accelerating or autocatalytic. Polymers are rarely found in a pure state but instead contain traces of polymerization additives and other compounds. The polymerization additives are often metallic which reduce the induction period of oxidation and accelerate the final rate. The induction period of oxidation is the initial period where no apparent oxidation occurs. The crystalline state of a polymer affects oxidation; amorphous polymers have more inter-atomic space to allow oxygen diffusion than crystalline polymers. Oxidation of amorphous polymers is therefore more rapid than oxidation of the corresponding crystalline polymers. Size also has an effect on oxidation where thin polymer films are more rapidly degraded by oxidation than thick bulky objects. The degradation processes of depolymerization, side group elimination and random chain scission are shown schematically in Figure 3.36.

Polymer degradation is not always destructive since a major problem is the persistence of many plastic products beyond their useful lifetime. The classic example of this problem is the plastic shopping bag. The shopper uses these plastic bags for at the most a few hours after purchase of the goods. After use, the bags are often carelessly disposed of to accumulate as litter for many years. In this context, the biodegradable plastic bag is being actively developed where a polymer that degrades after a few weeks is sought. Polyethylene and polyamides (nylons) are extremely resistant to biodegradation and other polymers such as polycaprolactone have been proposed as a degradable material. A technique of introducing biodegradability is to add subgroups to the polymer chain such as amino acid groups, which increase the ease of digestion of the polymers by bacteria. Bacteria in the soil and elsewhere are seen as the primary agents for biological disposal of polymers, which are generally unsuitable as food for animals or fertilizer for plants. In the field of medicine, degradable polymers are extremely useful to provide temporary support for regrowth of tissue. For example, cartilage in a damaged orthopedic joint can be regrown inside a degradable polymer mesh. Polylactate is often used as this purpose where it provides support for the implanted cartilage while it grows and attaches itself to the surrounding tissue. The Polylactate is slowly degraded by the enzymes within the human body and eventually disappears.

Polymeric materials usually contain not only the polymer but also numerous additives, such as anti-oxidants and plasticizers. When these additives are lost by prolonged exposure to solvents, such as water or organic compounds, the polymer will start to degrade rapidly. This process is known as leaching and commonly affects Poly Vinyl Chloride (PVC).

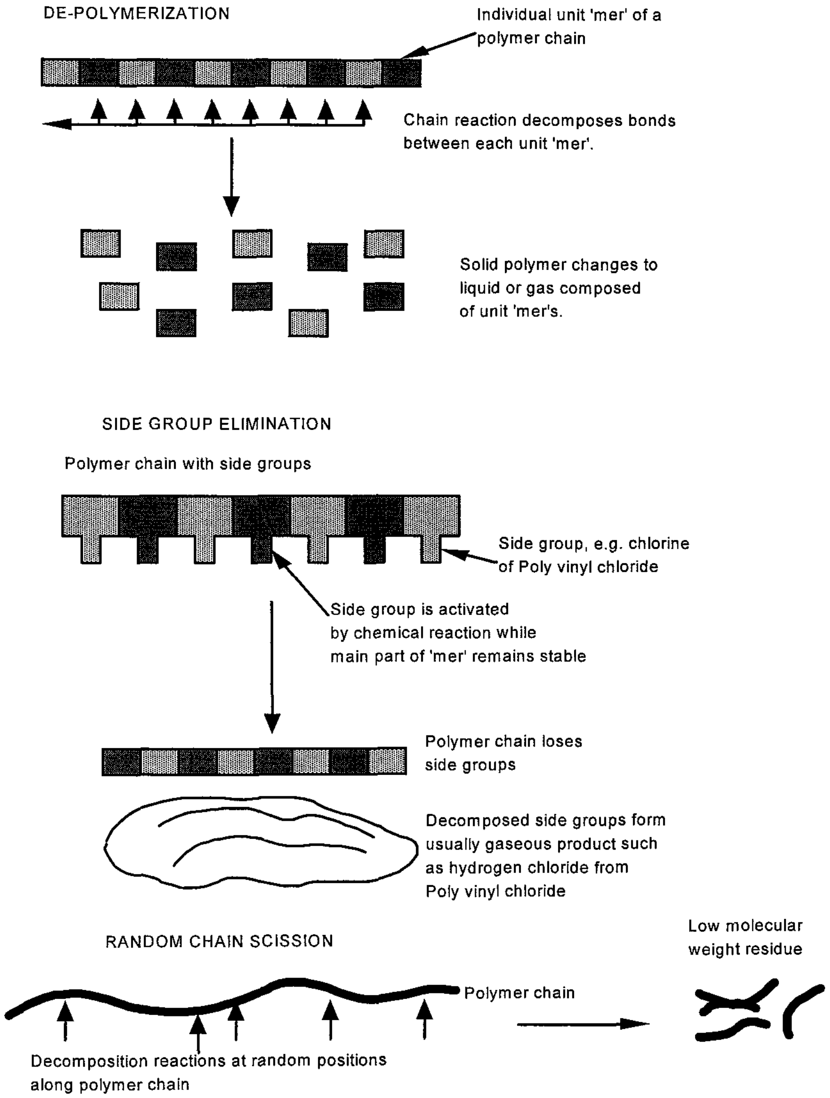


Figure 3.36 Degradation processes in polymers: de-polymerization, side group elimination and random chain scission.

3.5 Damage to cement and concrete, glass and engineering ceramics

Cement, ceramics and glass are not entirely immune to chemical attack despite their much better resistance to water than steel and other commonly used metals. Corrosive damage to these materials can be sufficiently severe to prevent the material from performing its required function.

Cement

Cement and its derivative material concrete, suffer from several degradation problems. There is direct chemical attack of cement analogous to corrosion of a metal, chemically induced swelling of cement, debonding of cement structure by lubricants and loss of corrosion protection of metal reinforcement for concrete.

Cement is vulnerable to any acid since it is mostly formed of hydrated calcium silicate and calcium aluminate which are decomposed by acids to form a calcium salt of that acid. Such acid attack occurs because even though calcium silicate is strictly a salt of silicic acid, silicic acid is a very weak acid, which does not form a durable compound with most metals. Cement also contains some residual calcium hydroxide from its manufacture; this alkali is not present in sufficient quantities to neutralize any invading acid. Strong inorganic acids such as sulfuric, hydrochloric and nitric acid cause immediate severe corrosion. Milder organic acids, e.g. acetic acid, cause a slower corrosion, which may allow a useful service lifetime of the concrete. Fatty acids which are a form of organic acids, will however tend to react with the calcium compounds to form soaps (soap is the salt of a fatty acid). The soap compounds are able to lubricate the grains of cement and cause the cement to soften and weaken. Alkalis only attack concrete when present as concentrated hot aqueous solutions. Lubricating oils (without fatty acid additives) and other organic compounds do not degrade concrete with the exception of creosote and phenol.

Cement contains water and raising the water content can cause swelling. Substituting the hydrated calcium aluminate, which is found in cement with another calcium compound that contains more water, can raise water content. The common cause of such water content increase is dissolved sulfate salts, particularly calcium sulfate. Calcium sulfate reacts with the hydrated calcium aluminate to form calcium-aluminum trisulphate by the following reaction:

$$3CaO.Al_2O_3 + 3(CaSO_4.2H_2O) + 26\ H_2O \ \text{--->}\ 3CaO.Al_2O_3.3CaSO_4.32H_2O$$

As can be seen there are 32 water molecules present in each complex of alumina, calcium oxide and calcium sulfate. This accumulation of water means that the compound formed has a greater volume than the original calcium aluminate and swelling of the cement then results. Swelling usually coincides with cracking, as cement is brittle. The mechanism of swelling is illustrated schematically in Figure 3.37.

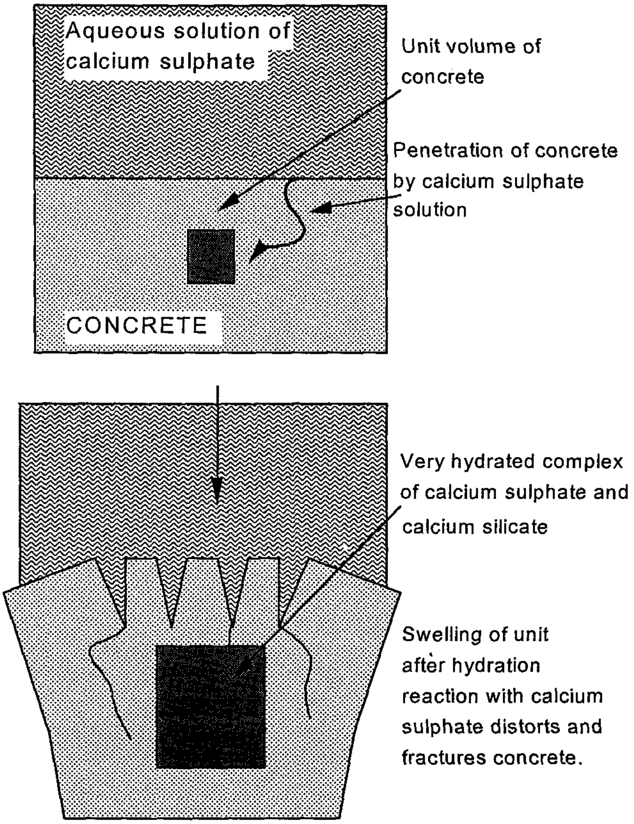


Figure 3.37 Mechanism of swelling in concrete immersed in calcium sulfate solution.

In fresh cement, the residual calcium hydroxide maintains a strong alkalinity, which protects steel reinforcement from aqueous corrosion. However, eventual impregnation of the cement by atmospheric carbon dioxide causes the calcium hydroxide to calcium carbonate. Calcium carbonate formation is accompanied by a decline in pH to 9 or less which allows steel to corrode in water. Calcium carbonate formation occurs when atmospheric carbon dioxide concentration exceeds 300 ppm. Once the steel rusts, the problem of 'concrete cancer' as discussed above will become evident. With increasing levels of atmospheric carbon dioxide, this problem can be expected to become more severe. It has been found however that the rate of steel reinforcement corrosion is controlled by the capillary diffusion of oxygen and chloride to the steel surface [Buenfeld]. The addition of additives to the concrete with the purpose of blocking capillary diffusion, can provide a useful postponement of 'concrete cancer' [Buenfeld].

Glass

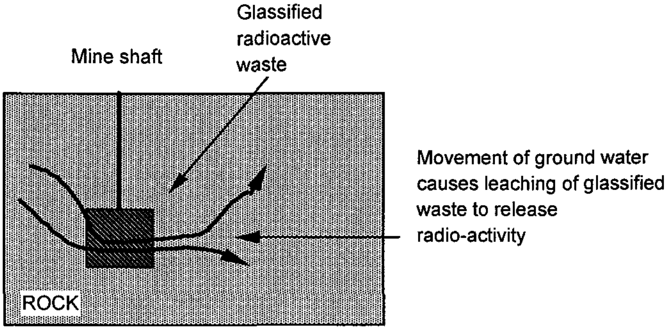
Glass is an amorphous mixture of silicic acid, sodium, calcium, aluminum or boron silicates. It is made by heating silica with a mixture of alkalis, such as sodium hydroxide, calcium hydroxide, aluminum hydroxide or boric oxide. The proportions of the different alkalis, silica and additional clay (which is mostly aluminum silicate) vary according to manufacture and application. Corrosion resistance reduces with increase in the proportion of alkalis used in glass manufacture. Glass has an amorphous molecular structure, which appears to enhance corrosion resistance. Glass does however suffer from corrosion as evidenced by ancient glass artefacts, which display loss of transparency, crazing (network of surface cracks) and other signs of materials degradation.

Glass is resistant to most acids except hydrofluoric acid but not to strong alkalis. A mixture of hydrofluoric acid and nitric acid is particularly aggressive to glass because the nitric acid dissolves the reaction products from hydrofluoric acid. Corrosion damage by alkalis is accelerated by both elevation in concentration of the alkali (higher pH) and increase in temperature. Most other substances do not cause severe corrosion of glass, e.g. organic compounds, strong acids at ambient temperatures, this is a major reason for the use of glass vessels to store chemicals in laboratories (the other reason is the low cost of glass). Hot acids will attack glass and even weak acids will eventually cause some slight surface damage to glass (see below).

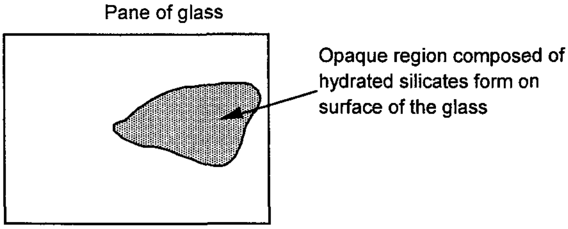
With the exception of some storage applications and translucent security windows, the functionality of glass depends on transparency. Even a limited superficial degradation of glass may cause loss of transparency. Prolonged exposure to moisture is the most common cause of transparency; the duration of water exposure required is very prolonged so that in most applications, misting of the glass does not occur. This transparency loss is caused by water hydrating the alkali silicates in glass to form hydrated alkali silicate and silica gel. An insoluble and opaque or semi-opaque mixture of hydrated alkali silicate and silica gel accumulates on the surface of the glass, which is where the greatest access to water can be found. A common example of glass misting occurs where glass sheets are stored together without an intervening layer of paper to absorb residual water.

Glasses are being studied as a means of containing high-level nuclear waste by incorporating radioactive atoms into the amorphous molecular structure of glass. Leaching of the glass by groundwater is a basic problem as this is the mechanism for the solubilization of radioactive atoms into groundwater. In common with optical and window glasses, it is found that high temperatures and extremes of pH, i.e. $pH < 1$ or $pH > 9$ promote leaching of the glass to release silicic acid and radioactive material. Fortunately most groundwaters have a pH between 5 and 9 so that leaching can be controlled.

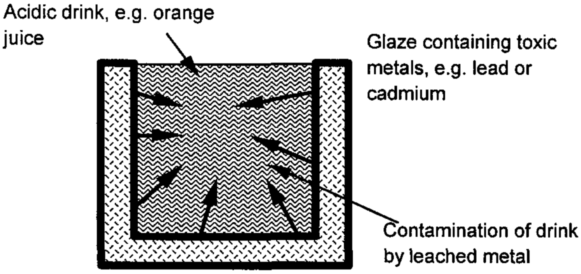
Glasses are routinely used as enamel coatings on metal objects to prevent corrosion where the most common example enameled steel kitchenware. Ceramics are also coated with glass in the form of a glaze and both glazes and enamel coatings are subject to similar problems of acid and alkali attack. Many types of enamel contain lead and cadmium oxides as a coloring additive and it is found that even weak organic acids such as citric acid or steam can dissolve the lead and cadmium oxides. This can cause toxicity problems for kitchenware, in one case, several people suffered acute lead poisoning after drinking orange juice that had been stored in a pot coated with a glaze containing lead oxide. Where an enamel or glazed container must store foods for human consumption, even a small loss of chemical elements from glass should be carefully evaluated for health risks. Corrosion problems caused by leaching and hydration of glass are illustrated schematically in Figure 3.38.



LEAKAGE FROM RADIO-ACTIVE WASTE RESPOSITORIES



LOSS OF TRANSPARENCY IN GLASS BY HYDRATION



TOXIFICATION OF DRINKS BY LEACHING OF GLAZES ON CONTAINER WALLS

Figure 3.38 Examples of corrosion problems caused by leaching or hydration of glasses.

Engineering Ceramics

Engineering ceramics can be conveniently classified as either oxide ceramics, e.g. alumina or partially stabilized zirconia and non-oxide ceramics such as silicon nitride and silicon carbide. Both oxide and non-oxide ceramics are highly corrosion resistant when compared to base metals such as iron. It is possible for some corrosion of ceramics to occur, particularly at high temperatures, which is where ceramics are often used in preference to metals.

Alumina and other oxide ceramics are vulnerable to hot concentrated solutions of strong acids such as sulfuric acid and hydrofluoric acid and in particular nitric acid. Nitric acid causes the most rapid damage to ceramics while phosphoric acid, which is not usually classified, as a 'strong acid' is also very destructive to oxide ceramics. Damage to oxide ceramics by strong acids is the result of chemical reaction of the oxide with the acid to form a salt. Non-oxide ceramics such as silicon carbide are mostly resistant to acids except where the ceramic is impure and in the case of silicon carbide contains significant amounts of silicon. Reaction bonded silicon carbide usually contains free silicon and is therefore vulnerable. Nitric acid rapidly degrades silicon carbide containing silicon so silicon carbide cannot be recommended for exposure to nitric acid, as it is technically difficult to manufacture silicon carbide components that do not contain silicon. The silicon in silicon carbide is usually found at the contact points between grains of silicon carbide where a silicon carbide component is typically formed by sintering. When the silicon is dissolved by acid, bonding between silicon carbide grains is broken and the component disintegrates. Loss of bonding between silicon grains after degradation by nitric acid is illustrated schematically in Figure 3.39.

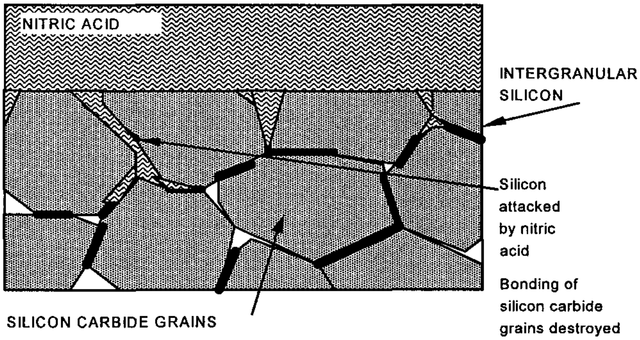


Figure 3.39 Acid attack of silicon impurities in silicon carbide.

Silicon nitride is also very resistant to most acids except hydrofluoric acid provided that little free silicon is present. Processing of ceramics causes significant variations in impurity levels so it is not possible to generalize on the durability of a specific ceramic without considering whether it is formed by e.g. reaction bonding or Hot Isostatic Pressing (HIPing).

Exposure to water causes eventual hydration of the ceramic; the main problem associated with hydration is corrosion fracture of the ceramic. Hydration of material around a crack tip accelerates crack growth rate since hydroxides are usually weaker than oxides of the same metal. All engineering ceramics are porous which allows corrosive liquids and water to penetrate the interior of a ceramic component. Corrosion damage may not therefore be entirely superficial but involve considerable subsurface damage of the ceramic. Partially stabilized zirconia (PSZ) and Tetragonal Zirconia Polycrystals (TZP) are prone to a form of water accelerated ageing at temperatures between 200 to 300°C in air which severely reduces the toughness of these ceramics. The precise causes of this ageing are unclear but its effect is to cause spontaneous transformation of the tetragonal phase to the monoclinic phase by zirconia. The tetragonal phase is denser than the monoclinic phase so the phase transformation causes swelling and cracking of the ceramic. The tetragonal phase is also tougher than the monoclinic phase in TZP while the density change of the tetragonal to monoclinic phase transition is exploited in PSZ to provide a crack closure force and so generate toughness. The tetragonal to monoclinic transformation in zirconia degrades both TZP and PSZ from valuable materials with high toughness to essentially useless brittle materials. Damage by water is critical to the successful application of super conducting ceramics. The multi metal oxide of yttrium, barium and copper: $YBa_2Cu_3O_{7-x}$, which has a very high limiting temperature of superconductivity, sustains a reduction in this limiting temperature from 93 °K to 70°K or less when exposed to humid air. Much more severe degradation of this super conducting oxide occurs when exposed to liquid water where non-conductive barium hydroxide is formed. Water-induced loss of superconductivity is dependent on a minimum temperature greater than 30°C so that low temperature storage of the superconductor (.i.e. close to its operating temperature) does not appear to present significant materials degradation problems. The various modes of water damage to ceramics both structural, i.e. PSZ and TZP or super conductive are illustrated schematically in Figure 3.40.

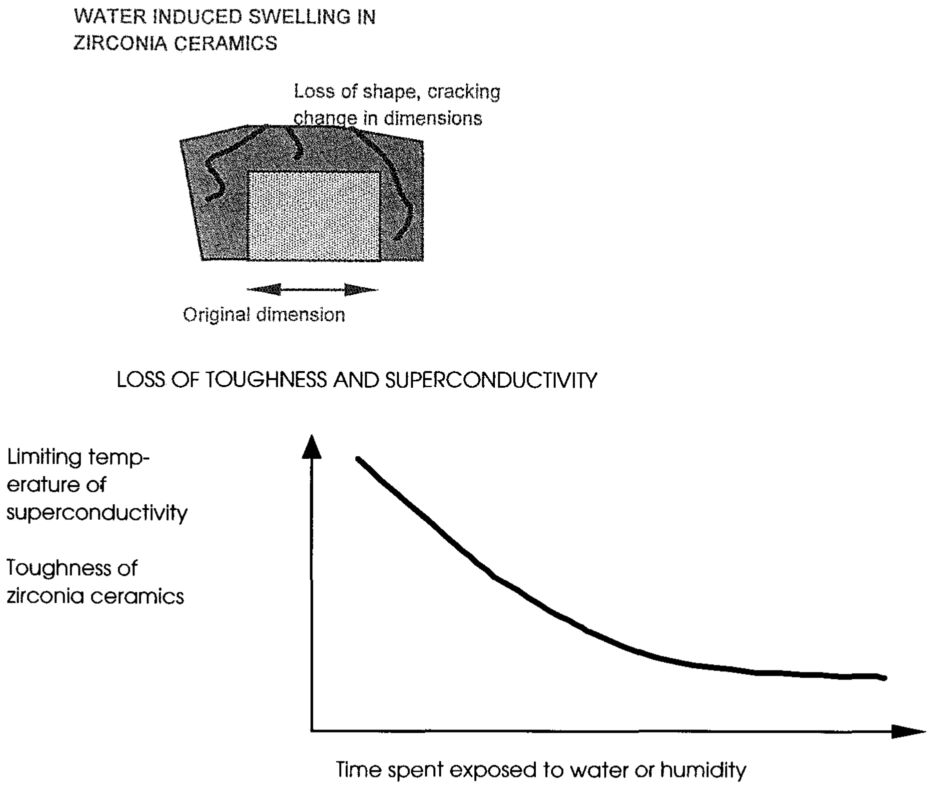


Figure 3.40 Water damage to zirconia ceramics and super conductive ceramics.

Non-oxide ceramics may be oxidized in the presence of atmospheric oxygen. Reaction Bonded Silicon Nitride (RBSN) is particularly vulnerable to oxidation because of its porosity, which lies between 15 and 20 vol. %. Rapid oxidation of RBSN is observed between 1000 and 1200°C where creep of RBSN is also found to occur. The creep is caused by the replacement of creep-resistant silicon nitride by the weaker silicon dioxide.

Silicon carbide is more resistant to oxidation than silicon nitride and is often used for protection against oxidation at high temperatures. Silicon carbide is however vulnerable to corrosive damage at temperatures between 1200°C and 1400°C by reducing atmospheres of e.g. hydrogen. It is believed that the passivating coating of silicon dioxide that forms on silicon carbide is reduced to the more volatile silicon monoxide. The silicon monoxide then evaporates to expose fresh silicon carbide without the protection of a passivating layer. Both silicon carbide and silicon nitride can suffer from rapid oxidation in a process directly analogous to the catastrophic oxidation of metals when heated in an oxidizing atmosphere containing sodium salts and sulfur. Combustion of sulfur-rich fuels in salt laden air is the most significant example of rapid oxidation, which presents problems in the use of ceramic turbine blades. The high temperatures of combustion cause common salt to react with sulfur dioxide (formed by oxidation of sulfur impurities in the fuel) and water to form sodium sulfate and hydrochloric acid by the reaction:

$$2NaCl(g) + SO_2(g) + 1/2\,O_2(g) + H_2O(g) \text{---}{>} Na_2SO_4(l) + 2\,HCl(g)$$

Where '(g)' denotes gaseous phase and '(l)' denotes liquid phase. The gaseous hydrochloric acid is harmlessly dispersed in the exhaust gas but the liquid sodium sulfate remains on the surface of the silicon carbide. When silicon carbide oxidizes to form a protective silica scale, sodium sulfate decomposes to form sodium oxide, which then reacts to form a low melting point sodium silicate with the silica by the following reactions:

$$Na_2SO_4(l) \text{----}{>} Na_2O(s) + SO_3(g)$$

$$Na_2O(s) + 2SiO_2(s) \text{----}{>} (Na_2O).2(SiO_2)(l)$$

The molten silicate then flows away to continuously expose unoxidized silicon carbide for rapid corrosion. Fortunately, this problem is most severe for low sulfur fuel of 0.05% wt. sulfur or less, with high sulfur fuel, e.g. 0.5% wt. sulfur, the decomposition of sodium sulfate to form sodium oxide is impeded by the higher concentration of sulfur trioxide. The resulting mixture of sodium sulfate and sodium oxide does not effectively dissolve the silica, which remains protective of the silicon carbide. In this manner the 'hot corrosion' as this form of materials degradation is often called differs between ceramics and metals where increasing amounts of sulfur cause corresponding rises in corrosion. Free carbon in the silicon carbide, probably from incomplete chemical reaction between silicon and carbon during the manufacture of silicon carbide, promotes the hot corrosion of silicon carbide by the combustion of high sulfur fuels. It is believed that carbon facilitates the decomposition of sodium sulfate to form sodium oxide by the following reaction path.

$$Na_2SO_4 + 2\,C \text{---}{>} Na_2S + 2\,CO_2$$

$Na_2S + 3Na_2SO_4$ ---> $4\ Na_2O + 4\ SO_2$

The sodium oxide then reacts with silica on the surface of silicon carbide to form the low melting point sodium silicate.

With the exception of hot corrosion that tends to cause loss of a large evenly distributed portion of ceramic, damage to ceramics by acids and other reagents causes porosity in ceramics. The structure of ceramics after corrosion is distinctly different from metals, which tend to form a discrete surface layer of reaction products. In most cases, corrosion occurs under immersion in a chemically active liquid, e.g. an acid, which can dissolve the reaction products of corrosion. Small holes, which are either pores that are deep and narrow, or shallower and broader pits, form on the surface of the ceramic. As corrosion progresses, the pits or pores become numerous and the porosity of the ceramic increases until only a skeletal structure of the ceramic remains. Cracking of the ceramic during corrosion is more common than with metals as the brittleness of the ceramic prevents it from accommodating the strains induced by damage to the ceramic microstructure. The process of pore and pit formation in ceramics by corrosion is often referred to as 'leaching'. Leaching is illustrated schematically in Figure 3.41.

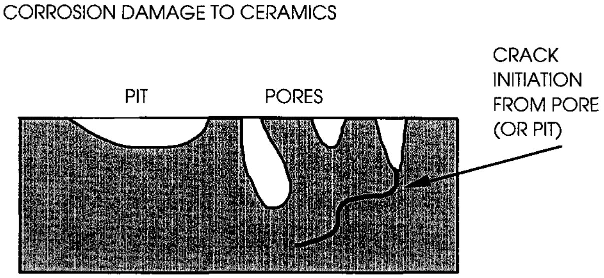


Figure 3.41. Hole and pit formation in ceramics during corrosion by a liquid.

The porous corroded layer has very little mechanical strength even though a significant amount of ceramic is left uncorroded and the porous corroded layer becomes an extremely favorable crack initiation site for fracture. Corroded ceramics are therefore more brittle than the corresponding uncorroded ceramic.

3.6 Dissolution of metals and ceramics in liquid metals, alkalis and salts

Many industrial processes necessitate contact between a solid metal or ceramic and molten metals, salts or alkalis. Molten salt heat treatment baths represent a very common example of such contact. The corrosion problems entailed are sufficiently severe to restrict the use of many processes and systems.

Dissolution in liquid metals

A solid metal, which is typically steel, is often used to contain a molten metal for the purpose of casting or preparing a metal-matrix composite. There is usually at least some limited degree of solubility between most combinations of solid and liquid metals so that the liquid metal will progressively dissolve any solid metal it contacts. This is the basic cause of *liquid metal corrosion* or dissolution of metals in liquid metals. In the context of dies for liquid metals, this form of corrosion is often referred to as *washout*. The solid metal can directly dissolve into the liquid metal in a manner analogous to dissolving of sugar in water or an inter-metallic compound can be formed on the surface of the solid metal. The intermetallic compound progressively grows in thickness and forms protuberances or 'buds' at the interface with the liquid metal. These protuberances are released into the liquid metal once they reach a critical size to cause a continuous loss of intermetallic compound. Solid metal does not directly contact the liquid metal but is instead converted to intermetallic compound before being lost in the liquid metal. The mechanism of liquid metal corrosion is illustrated schematically in Figure 3.42.

Liquid metal corrosion is intensified by flow of the liquid metal and the mechanisms of liquid corrosion vary with temperature. For example, titanium directly dissolves in liquid aluminum at 800°C but forms an intermediate layer, which may be an intermetallic compound in semi-liquid aluminum at 700°C. Most liquid- solid metal combinations fortunately do not appear to result in rapid corrosion. Liquid metals which cause rapid corrosion of most rapid metals, i.e. are 'aggressive', include liquid aluminum and zinc. Gallium is generally considered to be most aggressive of the liquid metals. It is known to rapidly corrode zinc and aluminum at 30 degrees Celsius and has been found to react with other metals at higher temperatures. For instance, thick layers of the inter-metallic compounds, $FeGa_3$, Ni_2Ga_3 and $CrGa_4$ were formed by

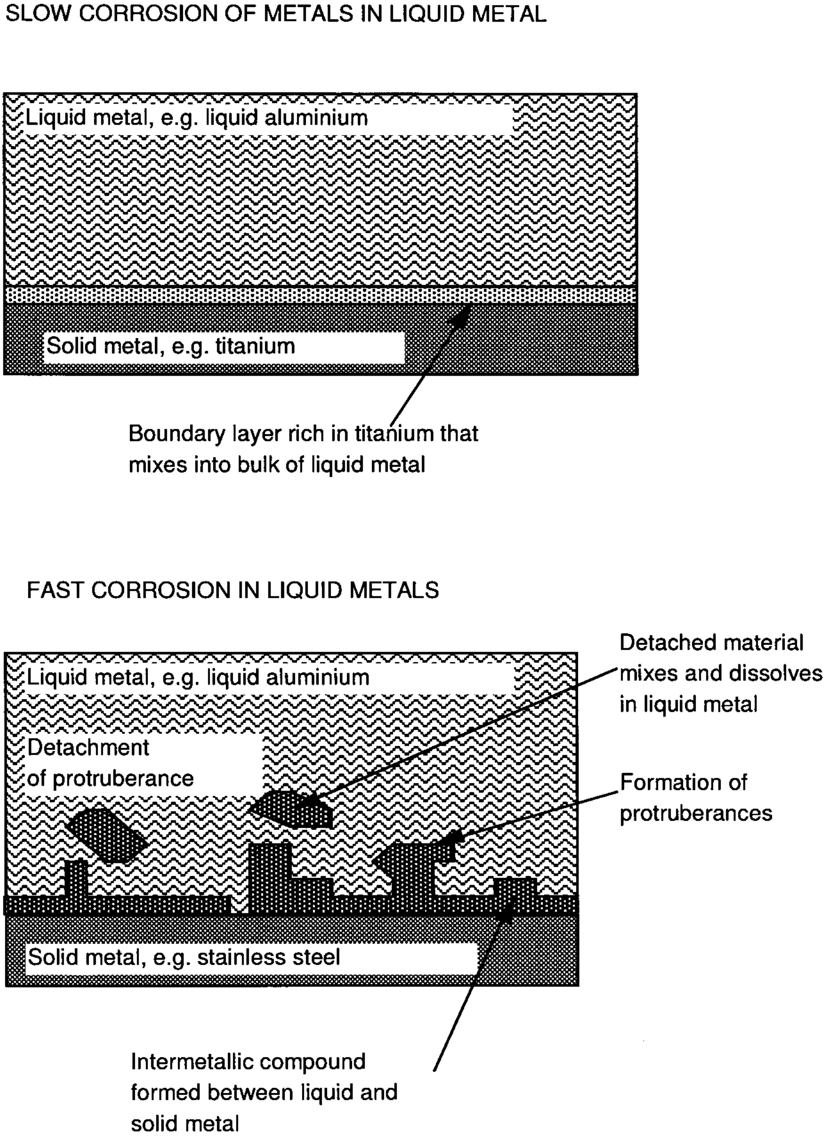


Figure 3.42 Mechanisms of corrosion of a solid metal immersed in a liquid metal.

reaction between gallium and iron, nickel and chromium respectively at 400 degrees Celsius [Barbier and Blanc]. The depth of the gallium inter-metallic compound layer grows at a constant rate without any apparent limiting thickness. Molten lead, bismuth and tin cause stress corrosion cracking of steels alloyed with chromium and nickel. Pure nickel, cupronickel alloys or Monel are found to have much better resistance to cracking under similar conditions. Molten lithium but not sodium or potassium is also considered aggressive. As a general rule, liquid corrosion of steels and other metals by liquid metals correlates well with solubility of the primary alloying component (for steels this is iron) in the liquid metal.

Oxidative corrosion by liquid salts and alkalis

Corrosion by liquid salts and alkalis is a form of oxidative corrosion where the salt facilitates reaction between the metal and atmospheric oxygen. If atmospheric oxygen is absent, then corrosion is much less for the majority of molten salts.

Atmospheric oxygen is dissolved in the molten salt where it either diffuses through the molten salt to reach the underlying metal or the oxygen reacts with the molten salt to form peroxide and superoxide ions which then aggressively corrode the metal. Molten sulphates typically show limited interaction with oxygen while molten carbonates actively interact with oxygen. A eutectic mixture of potassium and lithium carbonate is being developed as the electrochemical medium for a fuel cell but the corrosive nature of molten carbonate presents much difficulty in the selection of fuel cell materials. The metal reacts with the oxygen, peroxide, superoxide ions to form a metal oxide layer. With prolonged corrosion, mixed metal oxide forms where the corroded metal, e.g. iron forms e.g. lithium iron oxides in molten lithium carbonate

A critical feature of molten salt corrosion is that in some cases, the metal oxide is dissolved by the molten salt to cause *catastrophic oxidation.* An important example is the dissolution of chromium oxide as potassium chromate in molten potassium carbonate. In almost all cases of molten salt corrosion, the corrosion product layer formed on the metal is porous and does not create an effective barrier to further contact between metal and oxygen or oxygen radicals (peroxide, superoxide). The mechanism of liquid salt corrosion is illustrated schematically in Figure 3.43.

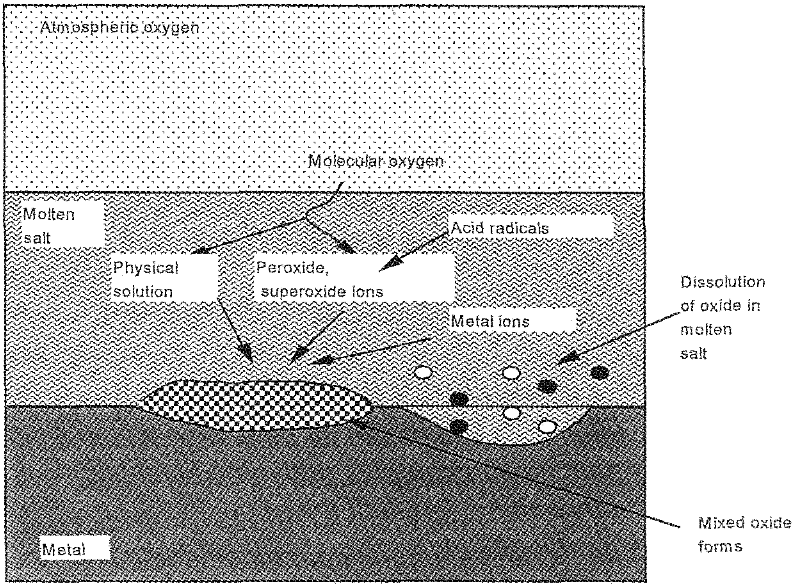


Figure 3.43 Mechanism of corrosion of metals by molten salts.

Liquid sodium hydroxide corrodes steels but nickel and nickel-base alloys are resistant to liquid sodium hydroxide because the nickel oxide surface film effectively controls oxidation.

Silicon carbide, which is resistant to most chemical reagents (e.g., acids), is rapidly etched by molten salts such as e.g, molten sodium carbonate and sodium sulfate. It is found that basic salts where a strongly alkali metal such as sodium is linked with a weak acid such as carbonic acid cause the most rapid etching of silicon carbide.

3.7 Biochemical and biological modes of materials degradation

There are several modes of biochemical and biological materials degradation each of which are different in mechanism and site of degradation. Aqueous solutions of proteins, the enzymes of bacteria, fungi and invertebrates are the most important agents of materials degradation. Viruses and vertebrates are not usually directly involved in materials degradation problems with the exception of mice or rats eating plastic and impact by animals and birds on structures. Consumption of plastic by rats and mice was found to cause electrical cable failure in aircraft. A solution to this problem was to carefully control eating of food inside the aircraft as mice and rats were attracted to the aircraft by waste food left in the aircraft. Once the waste food was exhausted, the mice and rats, which were unable to escape from the aircraft, became hungry and ate the plastic sheaths of electrical cables. Another aircraft problem caused by vertebrates is the impact of birds against aircraft engines during flight. Large birds e.g. vultures, which weigh several kg can reach heights of several thousand meters to occasionally collide with aircraft. Fracture of turbine blades often results if the bird hits the gas turbines that power the aircraft. In other cases, birds cluster around airports in search of waste food or animals living in the grass fields surrounding the airport runways. It is often necessary to deliberately scare the birds away with e.g. recordings of predator birds. Road vehicles which usually have a radiator made of soft copper located at the front of the vehicle are vulnerable to radiator damage when traveling at full speed from impact against kangaroos, deer and other large animals

Acceleration and initiation of aqueous corrosion by proteins

Proteins are commonly found as aqueous solutions, e.g. blood contains hemoglobin and many other proteins. Protein solutions are usually more corrosive than water but the underlying mechanism is still not well understood. It is believed that proteins absorb on a metal surface to form a film, which alters the pH immediately above the metal surface. The metal then suffers accelerated corrosion as a result of this change in pH. Examples of protein-induced corrosion are pitting of stainless steels after contact with aqueous solutions of globulins, which are found in the human body. Surgical implants e.g. the 'artificial knee' have to be made of special corrosion resistant alloys to prevent any risk of protein-induced corrosion inside the human body. Uncontrolled corrosion of surgical implants would result in unacceptably high metal ion concentrations inside the human body.

Acceleration of aqueous corrosion by microbes, biofilms and other unicellular organisms

Bacteria and other unicellular organisms depend on a wide range of substances for their nourishment. Bacteria have two modes of *metabolism, aerobic* and *anaerobic* where metabolism is the chemical reactions that generate life. Aerobic refers to a metabolism that is dependent on atmospheric oxygen and anaerobic describes a metabolism that occurs in the absence of oxygen. A familiar example of bacterial metabolism is the process of rotting or organic decomposition, which produces a range of volatile products that generate odors and liquid products that create slime. The liquid products are often organic acids which are the formed by anaerobic metabolization of sugars. Such acids can be produced in sufficient concentration to reach a pH 4 or less. This strength of acid can cause corrosion of galvanized coatings as well as steels. Corrosion induced by organic acid is a common problem in the food and agricultural processing industries.

In locations where common sources of food in the form of dead animal and plant tissue are not available, certain species of bacteria have evolved an ability to metabolize as food from some inorganic substances. Sulfur and sulfates are the inorganic materials mostly commonly involved in corrosion problems. Two species of bacteria cause anaerobic and aerobic metabolization of sulfur and sulfur compounds, these are the sulfate-reducing bacteria, which is called Desulfovibrio desulfuricans and the sulfur oxidizing bacteria of which thiobaccillus thiooxidants is a prominent variety. These bacteria are commonly found in the soil and cause corrosion of buried metal structures such as pipelines. Sulfate reducing bacteria exist by the reduction of sulfate ions in the presence of hydrogen to produce sulfide ions and water, the reaction is shown below:

$$SO_4^{--} + 4\,H_2 = S^{--} + 4\,H_2O$$

The hydrogen is formed from the adjacent decomposition of cellulose and sugars or from cathodic corrosion reactions. The sulfide ion accelerates corrosion by suppressing cathodic reactions and accelerating the anodic dissolution of metal. If ferrous ions are present, an insoluble ferrous sulfide is produced as a deposit. Sulfur oxidizing bacteria convert sulfur in the presence of oxygen and water to sulfuric acid by the following reaction:

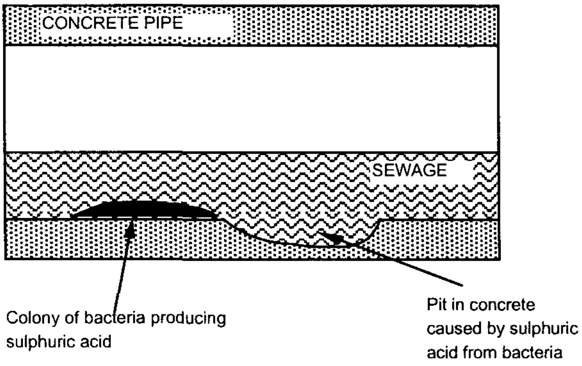
$$2\,S + 3\,O_2 + 2\,H_2O = 2H_2SO_4$$

The sulfur is obtained either in elemental form from sources of mineral sulfur or as sulfur compounds from oil fields or from sewage. The synthesis of sulfuric acid is a modified form of the above reaction where sulfur compounds are the source of sulfur. The sulfur oxidizing bacteria prefer acidic conditions for growth and produce as much as 5 wt % of sulfuric acid in water. This concentration of sulfuric acid is equivalent to a pH of approximately 2, which is sufficient to rapidly corrode steels and concrete.

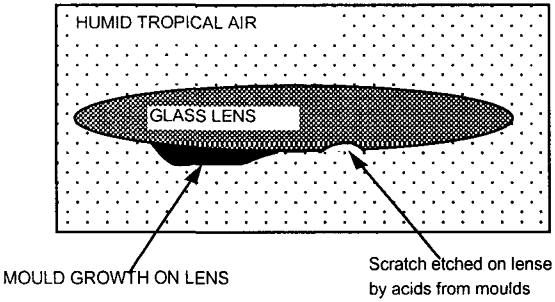
Sulfate reducing and sulfur oxidizing bacteria often coexist in the same location where metabolizing conditions vary from anaerobic to aerobic in a cyclic manner. For example when soil is soaked by heavy rain, oxygen is excluded from the subsurface of the soil so that sulfate-reducing bacteria grow. When the soil dries out after a prolonged period without rain, oxygen is able to reach the lower levels of soil and sulfur oxidizing bacteria then grow. Continuous corrosion of buried steel structures then results. Where nitrogen is available in large quantities in the soil as is the case below fields supplied with ammonia fertilizers, different species of bacteria synthesize nitric acid by a process similar to the production of sulfuric acid. Nitric acid corrosion of buried metal structures can then result.

Accumulation of marine organisms on metal surfaces exposed to seawater, moulds and slime adhering to metal surfaces intensify aqueous corrosion by providing crevices in the space between the adhering organism and the metal surface. Once a mould or other stationary organism is established on the metal surface, anodic dissolution proceeds to cause rapid corrosion. Biofilms of bacteria can also block the release of electric charge from the substrate to the liquid medium. This causes galvanic voltage differences to develop which lead to localized corrosion or pitting. Metal matrix composites (MMC's), e.g. aluminum reinforced by alumina particles are severely affected by biofilm-induced corrosion. In salt-water corrosion problems, the marine bacteria Pseudomonas forms biofilms on aluminum MMC's which cause intense localized corrosion, pitting and severe loss of tensile strength. The corrosion is more severe for the MMC than the corresponding aluminum alloy because the process of mixing molten aluminum and alumina particles introduces many new corrosion sites into the MMC.

Fungi can also affect glass by colonizing its surface and secreting organic acids, which etch the surface of the glass. This form of damage is relatively mild and mainly significant to the care of optical equipment in tropical climates. In tropical climates, long periods of relative humidity in excess of 90% are common and this encourages fungi to grow on glass. The surfaces of lenses and mirrors can be become roughened by fungus induced acid etching, which corrupts the optical quality of the lenses or mirrors. Damage to artificial structures by bacteria and moulds are illustrated schematically in Figure 3.44.



DAMAGE TO CONCRETE BY BACTERIA



DAMAGE TO LENSES BY MOULDS IN A TROPICAL CLIMATE

Figure 3.44. Biological causes of corrosion damage to artificial structures.

Even where a biofilm does not cause significant damage to its substrate, the biofilm may grow large enough to impair the mechanical function of a system. A common example such impairment is found in water pipes where large colonies of P.Aeruginosa block the flow of water. Biofilms can also form in medical implants such as catheters, where the biofilms become a source of persistent infection in the body.

Rotting and other forms of damage to wood

Wood is a natural material and a wide variety of living organisms consume wood as a food. The process of wood being formed in trees and then later destroyed by other life forms is part of a self-sustaining ecological life cycle that involves continuous recycling of organic materials. It is fortunate that the principle organic component of wood, cellulose, can only be digested by a small number of living organisms such as Fungi and invertebrates, e.g. termites, which are the principle agents of wood destruction. Birds and rodents may also cut into wood to form a shelter but this destruction is minor compared to that of fungi and invertebrates. Preservation of wood in a manufactured article such as a roof or a bridge represents an effort to remove the wood from the natural process of recycling.

Fungi cause wood to rot and are the prime cause of damage to wood in temperature climates. Fungi act by metabolizing the cellulose and other components of wood in order to obtain energy and organic materials from it. Fungal attack on wood leaves dry powdery debris that is useless for structural purposes. A key factor controlling fungi is moisture; very dry conditions with a relative humidity of less than 20% prevent the growth of fungi. Low humidity is the prime reason for the survival of wooden furniture and other articles in the tombs of Egyptian pharaohs, which are located in dry desert conditions. Fungal attack on wood causes both brown and white rot, which differ in the fungal species and the component of wood that is being degraded. In brown rot, the fungi consumes only cellulose and related compounds while with white rot, both cellulose and lignin is consumed. Cellulose is a polymer of glucose while lignin is a complex aromatic substance.

Protective measures for wood against fungal attack include paint coverings, which function by excluding moisture and fungi spores. Impregnation of wood by poisonous substances such as creosote and arsenic compounds to kill fungal spores is useful for outdoor applications where paints are unreliable. Impregnation of wood by arsenic compounds for preservation has been known to cause toxicity hazards when the wood was later burnt. Immersion in water prevents fungi growth as the supply of oxygen to the fungi is restricted. Numerous other organisms in water and especially salt-water consume wood, such as, borer worms (e.g. Teredo worms), so that protective measures are still required. Teredo and types of 'borers' are either a type of crustacean or mollusk. They damage wood by boring deep holes into the wood while consuming it as food. Most water-borne worms are found in salt-water and immersion of wood in fresh water slowly kills salt-water worms.

Termites, which are also known as white ants, can cause rapid destruction of artificial wood structures. There are about 1,000 species of termites, which are naturally found in tropical and sub-tropical climates where they feed on wood, leaves and plant litter. The destructive termites feed of wood and are divided into species, which build semi-permanent nests on the ground, and species which nest in the wood itself. The latter type of termite is more mobile and has been spread by human transport of wood far from the country of origin. Termites attack wood by chewing it into a mushy paste to be consumed either as food or used as a building material in their nests. The nests provide an enclosed humid environment essential for reproduction. Most termites have difficulty in digesting cellulose and rely either on bacteria in their digestive tracts or external colonies of fungi, which are fed with, chewed wood and later harvested. Termites are hardy and it is necessary to suppress them by chemical sprays of usually chlorinated hydrocarbons, which are toxic to humans. When fumigating a house for termites, it is necessary to evacuate the house during spraying and immediately afterwards. A new non-chemical approach is to try and exclude the termites from the house by laying a covering of fine nickel mesh beneath the foundations of the house. The graticules of the mesh are small enough to prevent passage of the termites from the ground to the roof, window frames and doors of the house. The selection of nickel for the mesh is to prevent corrosion of the mesh during the lifetime of the house. With the use of this mesh, it is still necessary to ensure that termites nesting in wood are not allowed to enter the property by attachment to a portable wooden article.

Other insects that attack wood are more common in temperature climates. These insects are known as the furniture beetle, the deathwatch beetle, the powder post beetle and the house longhorn beetle. These insects form nests inside the wood while slowly consuming the wood. A characteristic shared with marine borers and termites is that damage to the interior of the wood is usually accompanies by little surface damage apart from small entrance and exit holes. Wood can be almost totally consumed to leave just an external shell before damage becomes very noticeable. It is necessary therefore to carefully inspect wooden items to ensure that insect attack is absent. This discrete mode of attack is very different from fungal attack, which starts from the surface and spreads inwards. The interior nature of damage to wood by termites and other insects is illustrated schematically in Figure 3.45

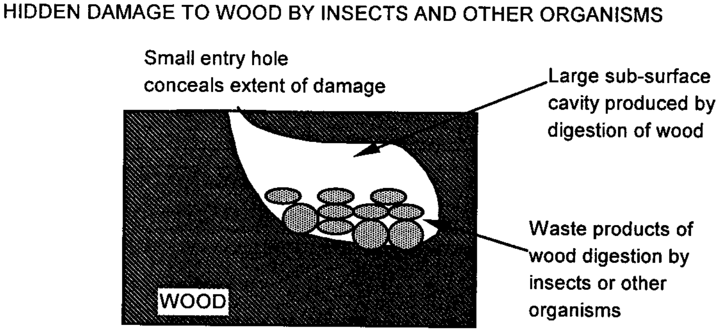


Figure 3.45 Interior damage of wooden structures by insects.

Wood is not always a passive victim of materials degradation and live wood from trees can cause damage to man-made structures. A classic example is the cutting of subterranean telephone cables by the West Australian Christmas tree (Nuytsia floribunda, family Loranthaceae). The West Australian Christmas tree, which is so called because of its brilliant yellow floral display in December and January, is the largest known variety of parasitic plant and is related to the mistletoe. In common with mistletoe, the West Australian Christmas tree does not form its own complete root structure so its sends out underground suckers, which attach themselves to the roots of other trees to extract water and nutrients. Telephone cables in the vicinity of West Australian Christmas trees become the target of the roots, which form collars of woody tissue around the cables. In an effort to try and extract water from the telephone cable, the collar exerts sufficient pressure on the cable to cut it, which results in cessation of telephone communication (Australian Telecommunications Authority).

3.8 Corrosion resistant materials

Aqueous corrosion

In general, organic polymers offer superior resistance than metals with the exception of certain polymers such as nylon. Ultra-High Molecular Weight Polyethylene is highly resistant to water but is expensive and difficult to mould. Where metals have to be used because of strength considerations or conduction of electrical current, stainless steels are suitable. Stainless steels are however vulnerable to pitting by aqueous chloride ions which are commonly found in salt water and ground water. Aluminum alloys vary widely in corrosion resistance and some alloys are durable against exposure to both fresh and salt water. Ceramics are resistant to water at moderate temperatures but some loss in strength should be expected from fracture induced by water.

High temperature oxidative corrosion

Nickel based alloys enhanced with chromium are widely used for resistance to high temperature oxidative corrosion. An example of an oxidation resistant alloy for turbine blades is Ni, 0.11% C, 16% Cr, 8.5% Co, 1.75% Mo, 2.6% W, 1.75% Ta, 0.9% Nb, 3.4% Al, 3.4% Ti and 0.05% Zr. This alloy is known as IN738LC and is used in applications where the economic gain obtained by high temperature service can justify the cost of the alloy.

A newer mode of conferring resistance to oxidative corrosion is to deposit intermetallic compounds on the surface of the metal component. For nickel-based alloys used as turbine blades, the blades are covered with nickel aluminide diffusion coatings. A disadvantage of intermetallic compounds is their greater brittleness than the original metal. Surface originated fatigue cracks are readily initiated in the intermetallic compound coating to cause early fracture of the component. Titanium aluminides are currently being developed for high temperature based on their resistance to high temperature oxidative corrosion combined with low density and strength at high temperatures.

Corrosion by liquid metals and salts

Less is known about resistant materials for corrosion by liquid metals and salts than for other forms of corrosion. Most metals are susceptible to some degree of liquid metal corrosion because both the corroded material and corroding liquid are metals. There is however a wide variation in resistance between metals. For instance, titanium is found to be considerably more resistant than stainless steel to solvation by molten aluminum at high temperatures. Inter-metallic compounds and ceramics are currently being studied for resistance to liquid metal corrosion because of their lower solubility in liquid metals.

For liquid salts, the high temperatures involved restrict usable metals to the known high temperature alloys, which usually are nickel based. These alloys are found to suffer from severe corrosion problems so inter-metallic compounds and ceramics are currently being considered as protective coatings. For instance, titanium nitride and titanium carbide coatings appear to offer useful protection some molten sodium and lithium carbonates. Borides and silicides of metals may also be useful as protective coatings.

3.9 Summary

Chemical degradation of materials covers a wide range of different mechanisms because of the large number of combinations of materials and corrosive agents. Water and aqueous solutions are perhaps the most important corrosive media for metals and probably for ceramics at low temperatures. Organic polymers are mostly resistant to water but are also susceptible to solvation damage by organic liquids such as lubricating oil and solvents. Ceramics are also susceptible to strong acids such as nitric acid which react with the ceramic to form salts of the acid, e.g. nitrates and sulfates.

High temperatures bring their own range of degradation mechanisms. Strongly adhering oxide layers that form on the surface of the metal effectively protect some metals. Other metals lack such oxide films and rapidly oxidize at high temperatures. Oxide films can also melt to expose unprotected metal while gaseous sulfur and chlorine compounds react to form low melting point sulfides and chlorides. Ceramics are not totally resistant to damage at high temperatures despite their higher melting points and intrinsically low chemical reactivity. Oxide ceramics tend to be more resistant than nitrides and carbides, which oxidize at high temperatures.

Biological organisms can cause chemical degradation by releasing aggressive biochemicals. Bacteria are the prime agents of this bio-corrosion, which may occur anywhere that water is present. Proteins accelerate the aqueous corrosion of metals and some enzymes rapidly attack polyurethane. Biochemical organisms also directly attack materials as a source of food or material to construct a nest.

RECOMMENDED READING.

Corrosion of metals

M.G. Fontana and N.D. Greene, Corrosion Engineering, publ. McGraw-Hill, 1978, New York, USA. *An invaluable discussion of corrosion, which gives adequate discussion to the detail of the subject, particularly corrosion theory.*

Corrosion Mechanisms, ed. F. Mansfield, 1987, Marcel Dekker, Inc., New York, USA.

Pitting corrosion of metals, Z. Szklarska-Smialowska, 1986, publ. Nat. Assoc. Corrosion Engineers, Houston, Texas, USA.

R.N. Parkins, Corrosion Processes, 1982, publ. Applied Science, Barking, United Kingdom.

U.R. Evans, The corrosion and oxidation of metals: scientific principles and practical applications, publ. Edward Arnold, 1960, United Kingdom. *A classic text about the degradation of metals, contains extensive and detailed information about industrial corrosion problems.*

G. Tsoumis, Science and Technology of Wood, publ. Van Nostrand Reinhold, New York, 1991.

D. Knoefel, Corrosion of building materials, translated by R.M.E. Diamant, publ. Van Nostrand Reinhold, New York, USA, 1978.

A. McLeish, Underwater concreting and repair, publ. Edward Arnold, 1994, London, Great Britain.

N. Buenfeld, Private communication, 2001.

V.S. Sastri, Corrosion Inhibitors, Principles and Applications, publ. John Wiley and Sons, 1998.

D.E. Clark and B.K. Zoitos (editors), Corrosion of glass, ceramics and ceramic superconductors, publ. Noyes, New Jersey, USA, 1992.
Specific reference to the following chapters.
Chapter 7, G.G. Wicks, Nuclear waste glasses: corrosion behaviour and field tests, pp. 218-260.
Chapter 12, R.A. Eppler, Corrosion of glazes and enamels, pp. 372-391.
Chapter 15, J. P. Bennett, Corrosion of ceramic construction materials in acidic environments, pp. 455-479.
Chapter 16 , P. Vaidyanathan,Degradation of ceramic cutting tools, pp. 481-490.
Chapter 17, Degradation of TZP ceramics in humid atmospheres, pp. 492-512.
Chapter 18, N.S. Jacobson, J.L. Smialek and D.S. Fox, High temperature corrosion of engineering ceramics, pp. 514-543.
Chapter 19, A. Barkatt, H. Hojaji and K.A. Michael, Corrosion of ceramic superconductors, an overview, pp. 548-575.

T. Kelen, Polymer degradation, publ. Van Nostrand Reinhold, 1983.

Marine biofilm corrosion.

R.U. Vaidya, D.P. Butt, L.E. Hersman and A.K. Zurek, Effect of microbiologically influenced corrosion on the tensile stress-strain response of aluminium and alumina-particle reinforced aluminium composite, Corrosion, 1997, Vol. 53, No. 2, pp. 136-141.

High temperature materials.

B. Kleinpass, K.-H. Lang, D. Lohe and E. Macherauch, Thermal-mechanical fatigue behaviour of NiCr22Co12Mo9, pp. 327-337, Proc. Symp. Fatigue under Thermal and Mechanical Loading, editors J. Bressers, L. Remy, M. Steen and J.L. Valles, 22-24 May 1995, Petten, The Netherlands, publ. Kluwer Academic Publishers, 1996, Dordrecht, The Netherlands.

Keijzer.M, Hermes.K, Vanderput,P.J.J.M.,Dewit.J.H.W., Schoonman.J., "A search for suitable coating materials on separator plates for molten carbonate fuel cells", Corrosion Science, 1997, Vol. 39, Iss.3, Pp. 483-494.

Liquid-metal corrosion

F. Barbier and J. Blanc, Corrosion of martensitic and austenitic steels in liquid gallium, Journal of Materials Research, Vol. 14, No. 3, 1999, pp. 737-744.

A.W. Batchelor, N.P. Hung and T.K. Lee, Wear of metal stirring rods in molten aluminium and suspensions of alumina particles in molten aluminium, Tribology International, Vol. 29, 1996, pp. 41-50.

CHAPTER 4

MATERIALS DEGRADATION INDUCED BY HEAT AND OTHER FORMS OF ENERGY

4.1 Introduction

Heating and radiation affect materials in a variety of ways, ranging from simple heating to localized attack by radiation. Extreme heat will melt or decompose any material but lower levels of heat can also degrade materials in the solid state. Materials are not invulnerable to radiation; light and ultra-violet radiation affect the surfaces of materials while nuclear radiation can embrittle a metal or other material. The main categories of degradation by heat and radiation are classified schematically in Figure 4.1.

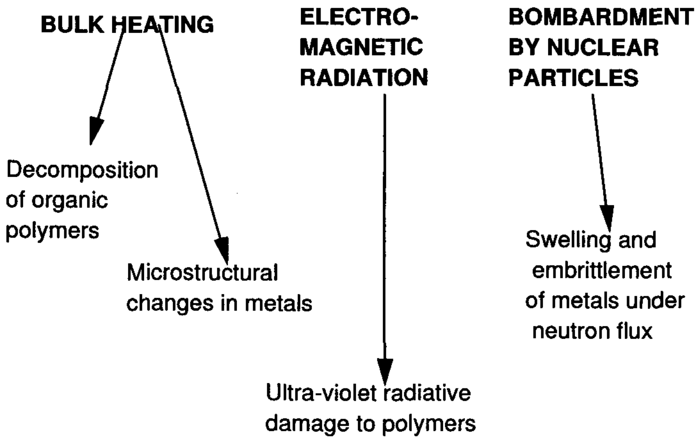


Figure 4.1 Classification of damage to materials by the heat and radiation.

The forms of materials degradation by radiation and heat that are shown schematically in Figure 4.1 are discussed below in this chapter.

4.2 Thermal degradation of materials

All materials are vulnerable to thermal degradation even where other factors such as corrosion or wear are absent. Increases in temperature generally lead to a decline in hardness of materials, which is usually counterbalanced by an increase in ductility. There are exceptions to the trend of declining hardness with temperature where a material may show a limited rise in hardness over a narrow temperature range but the effect is usually minor. Electrical conductivity declines for metals but is increased for non-metals, e.g. ceramics at elevated temperatures. Most of the changes to physical properties are reversible, e.g. hardness except where microstructural change has occurred. Thermally induced microstructural changes of metals, viz. grain growth, recrystallization, ageing, phase transformation, thermosetting in polymers and loss of toughening mechanisms in ceramics such as Partially Stabilised Zirconia are discussed in detail in Materials Science texts. These topics are not discussed further here, instead the more extreme effects of heat such as decomposition of polymers are described.

Among engineering materials, polymers are prone to thermal decomposition unlike metals and most ceramics. Important engineering polymers include wood, which is a natural polymer composite as well as synthetic polymers.

Wood

Wood is resistant to heat even though it is made of combustible organic compounds. When the wood is heated uniformly with sufficient allowance given for the low thermal conductivity of wood, then rapid decomposition occurs at temperatures of 200°C or higher. Slow decomposition is observed for temperatures above 100°C. Decomposition products are mostly low molecular weight organic compounds such as carbon monoxide, methane, formic and acetic acids. Some higher molecular weight tars are also released. 100 or 200 °C is a relatively low temperature compared to many engineering applications but it is found in practice that wood is extremely resistant to heat. When wood is exposed to fire, the low thermal conductivities of wood and the charcoal layer that forms on burning limit damage to the surface of the wood. The interior of a wooden object remains at a temperature much less than combustion temperatures until almost all of the wood is consumed by fire. A wooden beam or other structural member can maintain mechanical strength in a fire for much longer periods than an equivalent steel beam. The higher thermal conductivity of steel ensures that the steel beam is uniformly heated to a high temperature, where it fails by creep or fracture. The role of thermal insulation in protecting wood from damage by fire is illustrated schematically in Figure 4.2.

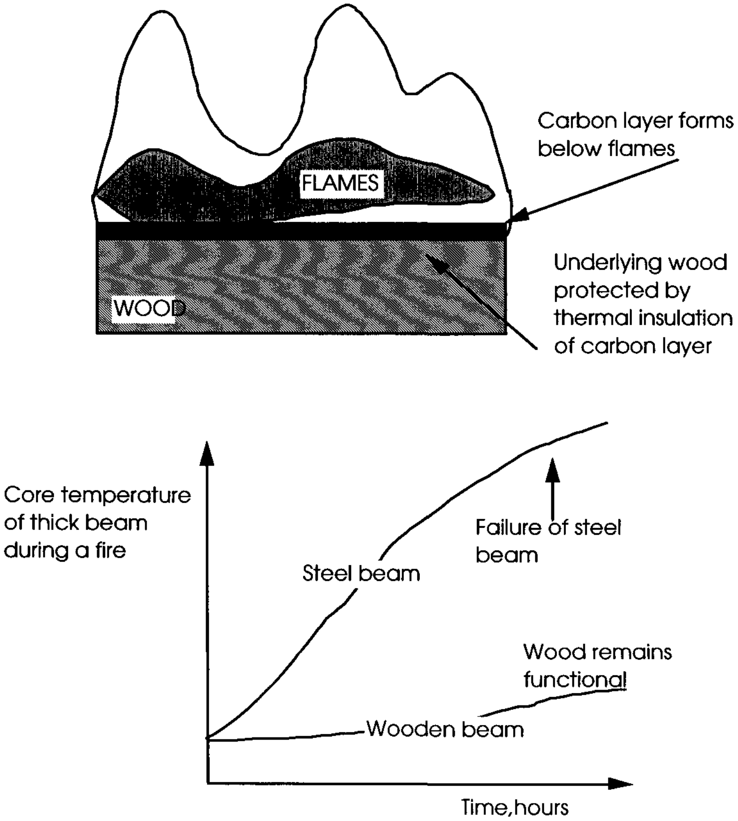


Figure 4.2 Role of thermal insulation in protecting wood from damage by fire.

Synthetic polymers

Polymers are sensitive to high temperatures and will progressively decompose to leave a mechanically inferior solid residue and possibly emit some vapor. One of the most sensitive or 'thermolabile' polymers is Polyvinyl chloride (PVC) which releases hydrogen chloride at temperatures of 150°C or greater. The basic categories of degradation process are depolymerization where the polymer reverts to its original monomers, random chain scission and decomposition of each 'mer' or unit of the polymer to a simpler form. The conversion of a 'mer' to a simpler form is usually caused by the loss of molecular side groups. Most polymer molecules consist of a carbon chain with various branches or 'side groups' composed of hydrogen, carbon and other elements, e.g. chlorine. An example of this is the loss of methanol from polymethacrylate at temperatures exceeding 320°C. The dominant mode of degradation of a specific polymer at any given temperature is determined by the speed of chemical reactions involved in any particular degradation mode. Most polymers therefore show different several modes of degradation when the temperature is progressively increased to a high level. Depolymerization is particularly destructive because the solid polymer reverts to a state approximating to its original liquid resin. These three modes of polymer degradation, depolymerization, loss of side groups and chain scission are similar to the molecular oxidation mechanisms illustrated and described in Chapter 3.

Polymers differ widely in susceptibility to a particular mode of degradation, PVC is prone to decomposition of the basic molecular unit as discussed above while polystyrene, polymethylmethacrylate and polyaldehydes are prone to depolymerization. A lower heat of polymerization favors depolymerization as the activation energy of depolymerization is the total of the activation energy of polymerisation propagation and polymerisation heat. The polymerisation heat is much larger than the activation energy of propagation and therefore controls the likelihood of depolymerization. Polyethylene (PE) provides a good example of a polymer that suffers from random chain scission at high temperatures with negligible depolymerization. Random chain scission causes an increase in brittleness and lowering in mechanical strength of the polymer. Polymers with side groups that contain elements other than carbon and hydrogen are vulnerable to loss of these additional elements. Examples of polymers showing this mode of degradation other than PVC are polyacrylonitrile, which decomposes to release hydrogen cyanide at temperatures higher than 400°C. High temperature decomposition of polymers can therefore release highly toxic vapors such as hydrogen cyanide. High temperature degradation of polymers can be accelerated by the presence of air in a process known as thermo-oxidative degradation, e.g. the evolution of HCl from PVC is accelerated by atmospheric oxygen.

4.2.1 Materials degradation at very low temperatures

A reduction in temperature inhibits the chemical reactions that cause many materials degradation problems. Furthermore, when the temperature falls below 0°C, freezing of water removes sources of liquid water for electrochemical corrosion. Severe forms of materials degradation are less likely at low temperatures than at high temperatures.

However, very low ambient and process temperatures still pose some problems of materials degradation. Steel and many other metals become brittle at low temperatures thus necessitating the use of special alloys that are usually based on nickel. Nickel has the advantage of maintaining toughness at low temperatures. Non-metals are also prone to low temperature brittleness. The bitumen in tarmac is observed to crack when there is a rapid reduction in climatic temperature. If the temperature falls rapidly then tensile strain develops in the bitumen due to thermal shrinkage and there is insufficient time for visco-elastic relaxation.

The formation of ice with an associated volumetric expansion (swelling) can cause a road surface to buckle because of lifting by subsurface ice. The widespread use of salt to melt snow greatly accelerates electrochemical corrosion of steel and other metals since the snow and salt create an aqueous sodium chloride solution.

Indirect causes of materials degradation can also occur when the temperature is sufficiently cold to cause solidification or wax crystal formation in lubricating oils. Automobiles and trucks that are exposed to severe winter weather can suffer severe wear inside their engines if the lubricating oil is unable to flow from the engine sump to the moving parts of the engine. Wax crystals can accumulate in oil supply pipes and block the flow of vitally needed lubricating oil.

4.3 Photochemical degradation of polymers

When light is shone on a polymer, some of the atoms within the polymer absorb the energy of the photons and become excited. The absorbed photon energy can be sufficient for atoms to spontaneously dissociate from bonds with neighboring atoms. Such dissociation causes localized decomposition of the polymer resulting in degradation phenomena such as chain scission. Excitement by light and dissociation of atoms is shown schematically in Figure 4.3.

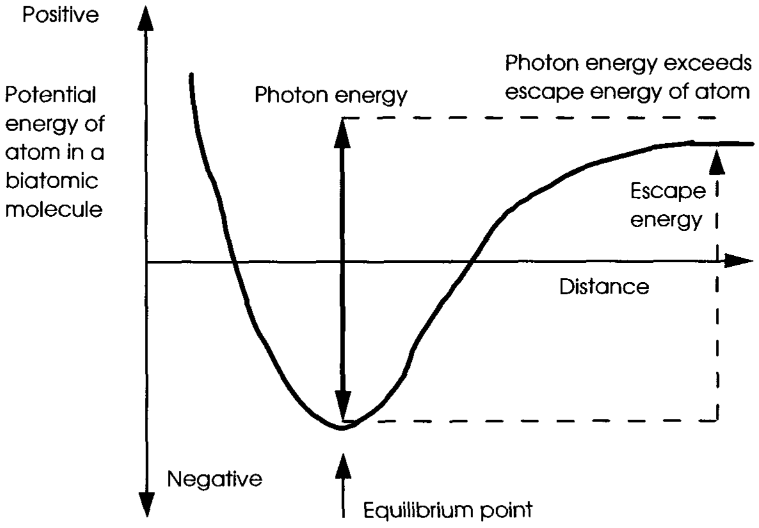


Figure 4.3 Mechanism of atom dissociation by photon energy.

Ultra-violet light (UV) has sufficient photon energy to initiate chain scission, which is why polymers exposed to sunlight eventually suffer from increased brittleness. Specific combinations of atoms from different chemical elements are particularly prone to photochemical reaction such as the carbonyl bond or carbon-oxygen bond and the carbon-chlorine bond. Where a carbonyl bond is located in the middle of a carbon chain, the carbon chain is split to release carbon monoxide. For the carbon-chlorine bond, direct decomposition of the carbon chlorine bond occurs to release free chlorine, which can then become either hydrogen chloride or chlorine gas. Photochemical degradation is accelerated in the presence of air resulting in photo-oxidation. Photo-oxidation is considered to be similar to high temperature oxidation of polymers except that it can occur at ambient temperatures. A similar process of the generation of radicals is involved (transient molecules with incomplete bonds and extra electrons) that propagate the oxidation reaction from one polymer molecule to another. A fundamental difference however between photo-oxidation and thermal oxidation is that the photo-oxidation products absorb UV far more effectively than the original polymer so that photo-oxidation is auto catalytic and self-accelerating. Polymers also contain numerous minor ingredients such as organic compounds with carbonyl and hydroperoxy groups, which are used in the manufacture of the polymer. These compounds can absorb the UV light to start the release of radicals that cause oxidation. Polymers vary widely in composition and amount of these minor ingredients with a corresponding variation in sensitivity to light.

4.4 High energy radiation damage of materials

Much of the materials used to construct nuclear power systems are exposed to high-energy radiation. The most damaging mode of radiation is neutron flux where the neutrons penetrate a material to alter its microstructure. Fluxes of high-energy ions from fission reactions or helium ions from fusion reactions also cause damage by mechanisms similar to those caused by neutron flux but are less common. Metals are mostly used for nuclear power systems so that knowledge of irradiation damage of ceramics and polymers is more limited. The level of radiation commonly experienced by engineering materials varies widely. It is found that low levels of neutron radiation generate a mild strengthening effect on metals and some polymers. At higher radiation levels however, the effect on materials is destructive with the scale of damage directly proportional to the amount of radiation received.

When neutrons penetrate a material they contain energy that is either completely dissipated as damage to the crystalline structure of the material or else the neutrons pass through the material with only a partial loss of energy. If a material structure of sufficient thickness is considered then all of the neutron energy is dissipated within the material. The basic mechanism of neutron energy dissipation in metals is dislodgement of atoms to form vacancies and dislocations. The metal, e.g. steel, then becomes harder with a higher tensile strength after neutron irradiation but also becomes more brittle. These changes in mechanical properties, which are known as radiation embrittlement, are the result of abnormally high levels of vacancies and dislocations in the material, which impede plastic flow. Swelling of metals also occurs as a defective crystalline structure necessarily has a lower density than a perfect crystal. Radiation therefore causes metal components to exceed dimensional tolerances with drastic consequences for shafts and other moving components. Precipitate formation is often observed in a microstructure after irradiation and this phenomenon can be likened to over aging. After prolonged irradiation, vacancies formed by neutron collision coalesce to form voids. These voids function as sites of stress concentration for premature fracture of the material. The microstructural composition of a metal can also be altered by neutron flux. Many alloys are supersaturated solutions that are thermodynamically unstable. Neutron irradiation provides energy and vacancy sites for atoms to diffuse from regions of high chemical potential energy to low potential energy. Two or more different solute elements in an alloy may combine to form precipitates to cause the over aging mentioned above. Creep of metals is also initiated by radiation to cause a phenomenon known as *radiation creep.* Damage to the crystalline lattice by neutron irradiation is shown schematically in Figure 4.4.

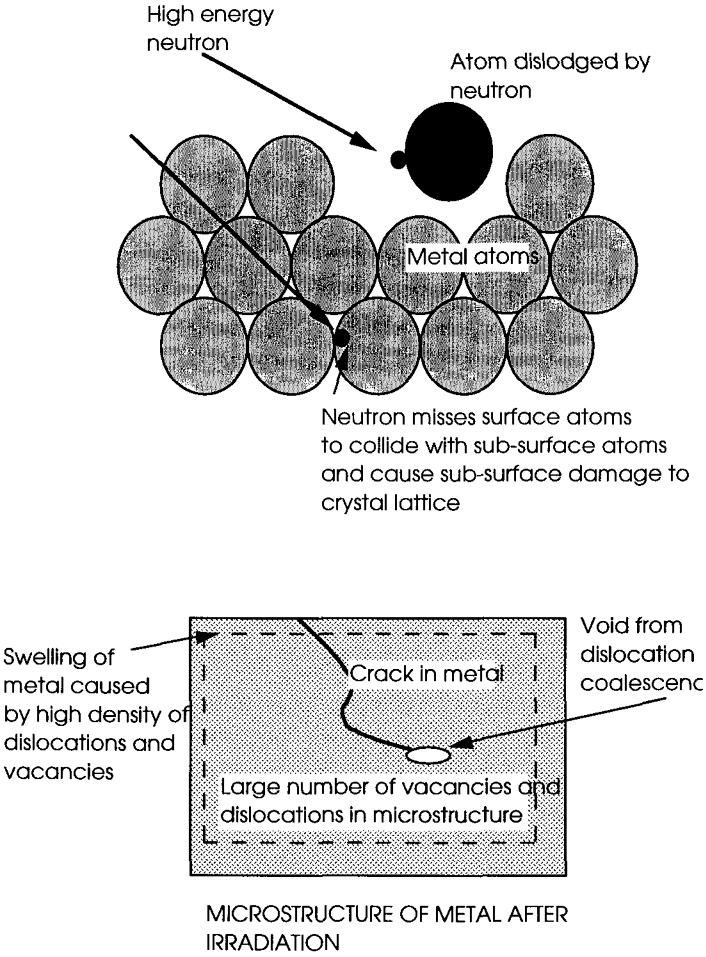


Figure 4.4. Damage mechanisms in metals when subjected to high energy neutron irradiation.

The effect of neutron radiation on metals is not entirely destructive as the loss of ductility is accompanied by an increase in tensile strength. Radiation damage declines at high temperatures where continuous annealing suppresses any accumulation of dislocations and voids. Annealing of metals can be used to correct prior radiation damage.

Ceramics are also vulnerable to radiation damage, especially when the ceramic has covalent bonding, e.g. silicon nitride and silicon carbide. The covalent bonds can be altered by neutron bombardment with a resultant effect on the ceramic microstructure. Ionically bonded ceramics, e.g. oxide ceramics suffer less damage from neutron bombardment and respond in the same way as metals, i.e. accumulate dislocations and vacancies to later form cracks. Annealing is also effective for ionic ceramics to curb damage by neutrons.

The direct effect of radiation on polymers is to generate reactive radicals at the ends of the hydrocarbon chains that form the polymer. High-energy radiation modes such as gamma radiation and neutron radiation can cause chain-end radical formation. Bonding between the radicals then occurs to cause *cross-linking* between chains. Cross-linking inhibits motion between chains to raise the tensile strength but reduce toughness. Rubbers lose their elasticity or rubbery properties before softening at much higher radiation levels. High-energy radiation can also cause *scission* of chains to reduce the overall molecular weight. The formation of a surface layer of reduced molecular weight on Ultra High Molecular Weight Polyethylene (UHMWPE) is the result of irradiation by gamma rays that are used to sterilize the UHMWPE. Sterilized UHMWPE is used for orthopedic implants (artificial hips and knees) where there is a compromise between adequate sterilization and loss of mechanical properties.

4.5 Summary

Thermal energy and other forms of energy are found to exert a variety of destructive effects on all materials. Even visible light and ultra-violet radiation can damage materials such as polymers. Damage can occur at temperatures much less than the melting or sublimation point of the material. Destructive effects include changes in dimension, loss of material, and loss of strength. Nuclear radiation is especially destructive towards materials and this should be considered when selecting materials for handling radioactivity.

RECOMMENDED READING.

Damage by high energy radiation. Proc. 11th ASTM Int. Symp. Effects of radiation on materials, Scottsdale, Arizona, USA, 28-30 June

1982, edit. H.R. Brager and J.S. Perrin, publ. ASTM, Philadelphia, USA.

Thermal and photochemical degradation of polymers.
T. Kelen, Polymer degradation, publ. Van Nostrand Reinhold, New York, USA, 1983.

Thermal degradation of wood.
G. Tsoumis, Science and Technology of Wood, publ. Van Nostrand Reinhold, New York, 1991.

CHAPTER 5

DUPLEX CAUSES OF MATERIALS DEGRADATION

5.1 Introduction

The interaction between two or more independent causes of materials degradation results in a duplex mode of materials degradation. The materials degradation rate whether expressed in loss of material or loss of mechanical strength is often faster than the sum of each form of materials degradation acting separately. For this reason, duplex modes of materials degradation are of great practical importance. Synergism between unrelated forms of materials degradation are not immediately obvious and have in some instances been only recently recognized. For example, in minerals processing, the shiny or polished surfaces of tools processing mineral slurries delayed recognition of the fact that corrosion and wear acted synergistically to rapidly destroy the tools.

There are many forms of synergism between independent modes of materials degradation, the most significant of these are illustrated schematically in Figure 5.1.

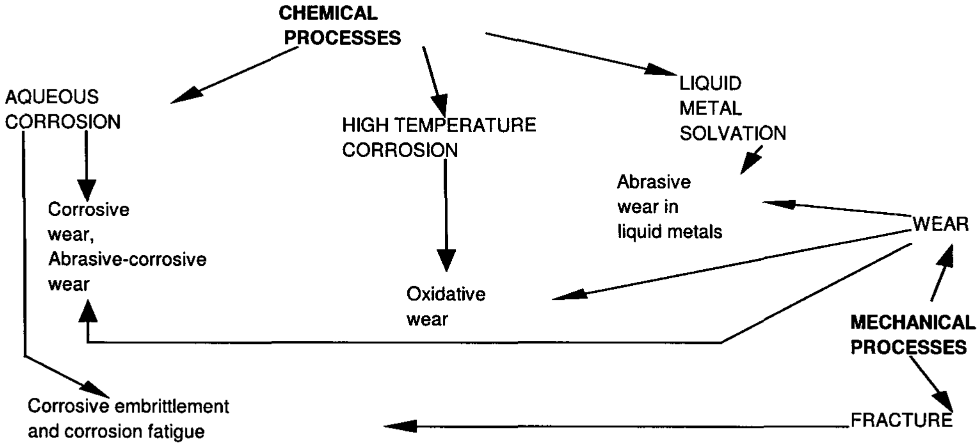


Figure 5.1 The different forms of destructive synergism between independent modes of materials degradation.

The primary source of destructive synergism is the interaction between chemically and mechanically driven materials degradation. Aqueous corrosion, high temperature corrosion and liquid metal solvation combine (separately) with wear and fracture to cause corrosive wear, corrosive-abrasive wear, oxidative wear, corrosion fatigue and embrittlement. There are numerous examples of these duplex modes of materials degradation to be found in minerals processing plant, petrochemical refineries, outdoor steel structures and other industrial plant. All the phenomena shown schematically in Figure 5.2 are discussed in this chapter.

5.2 Wear in a corrosive or chemically active environment

In a terrestrial environment which is often artificially manipulated, e.g. irrigation by a chemical reagent, wear is nearly always subject to some chemical influence. Atmospheric oxygen is very difficult to exclude from a wearing contact and so is present unless substituted by some other gas or liquid. In the presence of frictional heat or exposure of un-oxidized metal by adhesive wear, atmospheric oxygen reacts with the metal to modify wear. Where oxygen is excluded, another reactive substance is often substituted. For example, oil refining requires supplies of concentrated sulfuric acid, which has to be pumped.

Manufacture of polyvinyl chloride involves compression of chlorine gas. Both the pump and compressor depend on seals, which are necessarily wetted by sulfuric acid and chlorine gas respectively. Concentrated sulfuric acid and chlorine gas are both strong oxidants, which can be expected to have a strong effect on the wear of metals, polymers and other substances. Interpretation of many practical wear problems necessarily involve knowledge of *tribochemistry* that relates to the interaction between chemical and mechanical factors in wear.

5.2.1 Corrosive and corrosive-abrasive wear

Corrosion in the context of engineering materials involves the transformation of a metal into an oxide, hydroxide, sulfide, chloride or other oxidants. In most cases, the mechanical properties of the corrosion product are grossly inferior compared to the original metal, a classic example being common rust and steel. In the presence of wear, corrosion has the effect of transforming a usually wear resistant metal into a weak and brittle corrosion product that is rapidly removed by wear. A rapid wear known as corrosive wear then results, which, for reasons to be described, has the deceptive characteristics of combining rapid wear with reduced friction. There are two significant forms of wear as influenced by corrosion: true corrosive wear between two materials in sliding contact when immersed in a corrosive agent, corrosive-abrasive wear where abrasion occurs in a corrosive media. Similarities and differences between these two forms of materials degradation are discussed below.

Corrosive wear

Corrosive wear is usually modeled in terms of a sliding contact between steels and cast irons where sulfuric acid or another strong acid is present in varying concentrations. Depending on the concentration of the acid, one of four possibilities listed below may occur:

1. A durable and thin corrosion product film forms, which effectively suppresses adhesive wear to maintain moderate wear and friction.

2. A mechanically weak corrosion product film forms, which is rapidly removed by wear. The film of corrosion product functions as a sacrificial solid lubricant to provide reduced friction at the expense of rapid wear.

3. Localized wear of the corrosion product film may occur to expose small areas of the worn surface. Galvanic coupling will intensify corrosion in the exposed areas with the surface covered by film to cause pitting in the exposed areas. The weakening of the worn surface that is caused by pitting will accelerate wear.

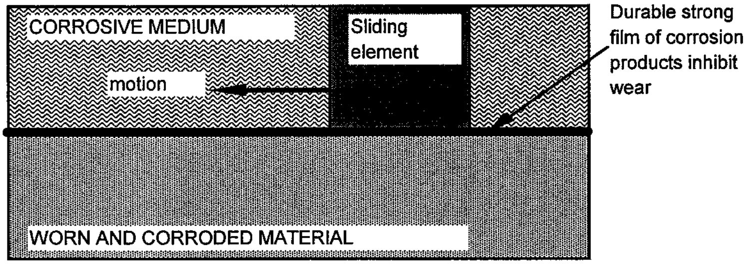
4. Corrosion and wear act independently to produce a rate of materials degradation that is the sum of separate corrosion and wear on different specimens of the same material at the same test conditions.

These four modes of corrosion-wear interaction are illustrated schematically in Figure 5.2

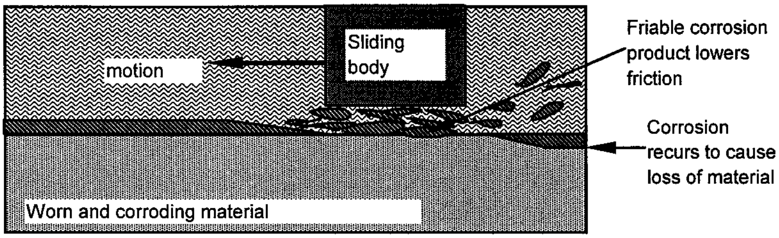
The first three modes of the four modes of corrosive wear illustrated in Fig. 5.2 can be observed during the wear of steel in sulfuric acid. The fourth mode is relatively uncommon because of the catalytic effect that exposure of un-oxidized metal and plastic deformation of surface layers exerts on corrosion. When the concentration of sulfuric acid is varied from very dilute to highly concentrate, the mechanism of wear of the steel changes from one of the three modes to another. Very high concentrations of sulfuric acid are less corrosive to metals than low concentrations because there is insufficient water to enable dissociation of the sulfuric acid molecule into an anionic radical and two hydronium ions. As discussed in chapter 3, the hydronium ions generate acidity in water. At approximately 90% weight concentration of sulfuric acid, a durable film of iron oxides is developed on the worn steel surface so that friction and wear are less than in plain water. As the concentration is reduced from 90%, pitting and film corrosion at low concentrations occur with lowered friction but elevated wear compared to plain water.

It is observed that the most common form of corrosive wear for all but extremely corrosive reagents such as sulfuric acid is the second mode involving a short-lifetime film of corrosion product. A model of corrosive wear involving a mechanically weak film that is regularly removed involving cycling initiation and cessation of parabolic film growth is illustrated in Figure 5.3.

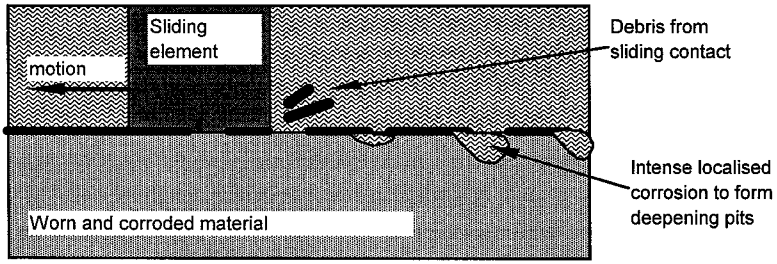
MODE 1, Protective film formation.



MODE 2, Corrosion product acts as sacrifical lubricant



MODE 3, Localised corrosion at worn areas of material



MODE 4, Corrosion and wear act independently

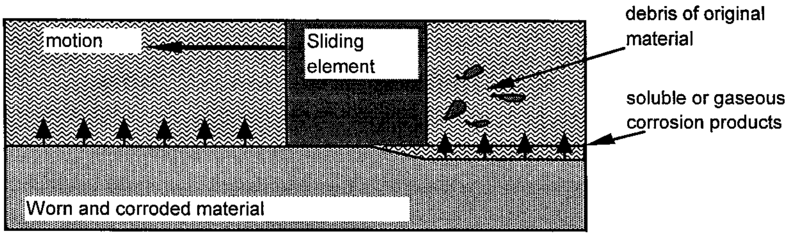


Figure 5.2 Modes of interaction between corrosion and wear.

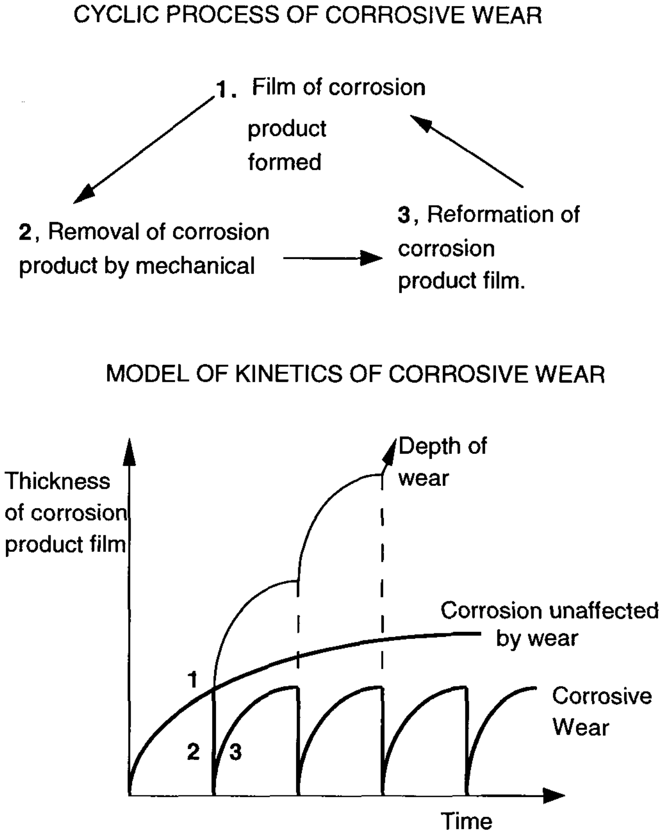
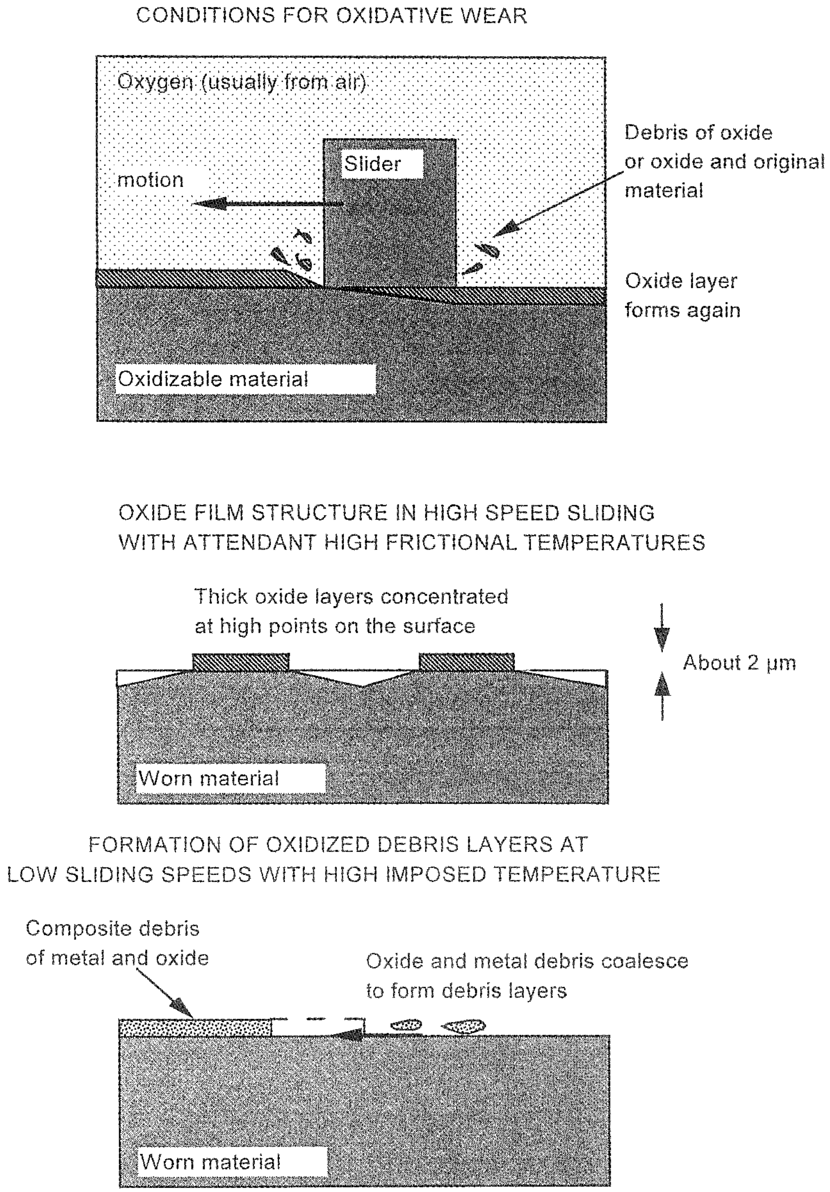


Figure 5.3. Corrosive wear by cyclic removal and formation of passivating films: mechanism and effect on corrosion kinetics.

Corrosive wear occurs in a variety of situations, the most obvious being the case cited above of metals exposed to strongly corrosive fluids. Less corrosive fluids than strong acids can also cause corrosive wear if conditions are favorable. A very common case is the corrosive wear by air of steels where sliding speeds are sufficiently high to raise frictional temperatures to the point where rapid oxidation occurs but not so high as to reach the melting point of the metal. This form of corrosive wear is known as *oxidative wear.* When most of the temperature rise is caused by frictional heat, thick layers about 1 µm thick of almost pure iron oxide form on the worn surface. Wear is controlled by the rate of detachment of these thick oxide layers. If sliding speeds are reduced to suppress the frictional temperature rise and external heating is used to raise the temperature, a different form of oxidative wear results. Since asperity contacts under low sliding speeds are not provided with sufficient energy to cause significant oxidation within the duration of each asperity contact, adhesive wear and debris production precedes the formation of oxides. Layers of partially oxidized debris then accumulate on the worn surface to eventually control friction and wear. The debris is oxidized in stages while it is plastically deformed between the wearing surfaces to produce a debris layer with complex structure. These mechanisms of oxidative wear are illustrated in Figure 5.4.

A third benign form of oxidative wear occurs under rolling between steels in atmospheric conditions where the lack of sliding precludes direct wear of the contacting steel surfaces. Oxide films are cyclically formed by atmospheric corrosion to be subsequently removed by rolling contact. The oxide films prevent direct contact between the steel rolling elements and suppress any adhesive wear that might otherwise occur. Rolling bearings and railway wheels are the most common sites for oxidative wear in rolling contact. Wear depths of the order of a few millimeters are acceptable on railway wheels so that corrosive wear permits a long lifetime for the railway wheels. For rolling bearings, which lack lubrication, by oil, only a much smaller wear depth of 100 μm or less can be tolerated. Steel rolling bearings, which are forced to operate without oil, as may occur in emergency operation will rapidly fail even though the amount of wear is much smaller than for railway wheels.

Another form of corrosive wear that is of practical importance is corrosive wear by sulfur or chlorine based thermo-labile (easily decomposed) lubricant additives. These lubricant additives, which are known as Extreme Pressure additives or 'EP' additives, ideally function by mild corrosion of mode 1 as listed above. Chemical activity of the additive results from their thermo-lability, which releases reactive sulfur and chlorine or other oxidizing elements. If the additives are too active for a given situation then corrosive wear of mode 2 results. Very often however, corrosive wear by these additives is accepted as the inevitable lack of precision in control of lubrication would render a wearing contact liable to severe adhesive wear should the chemical activity of the EP additives be insufficient.


CONDITIONS FOR OXIDATIVE WEAR

OXIDE FILM STRUCTURE IN HIGH SPEED SLIDING WITH ATTENDANT HIGH FRICTIONAL TEMPERATURES

FORMATION OF OXIDIZED DEBRIS LAYERS AT LOW SLIDING SPEEDS WITH HIGH IMPOSED TEMPERATURE

Figure 5.4 Mechanisms of corrosive wear by air, i.e., ***Oxidative wear***.

Corrosive-abrasive wear

Corrosive-abrasive wear is in most instances the cyclic removal of corrosion production films by the abrasion of grits that contact the worn surface. As mentioned above, this form of wear was slow to be recognized and mineral processing engineers persisted for many years with very hard but corrodible materials before cheaper, softer but non-corrodible polymers were substituted to give lower wear rates and longer service lifetimes. A common example is the replacement of steel piping by Ultra-High Molecular Weight Polyethylene or High Molecular weight Polyethylene piping for coal slurry processing. This form of wear usually occurs under what is called 'low-stress abrasive wear' where the grits are allowed to slide and roll past the wearing surface without being crushed. The mechanism of corrosive-abrasive wear is illustrated schematically in Figure 5.5.

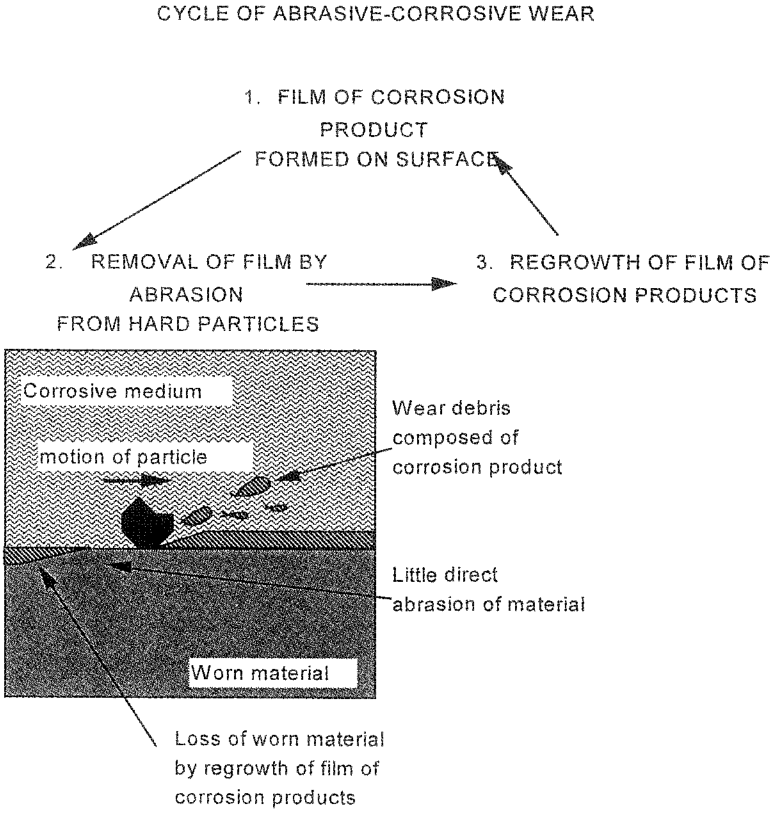


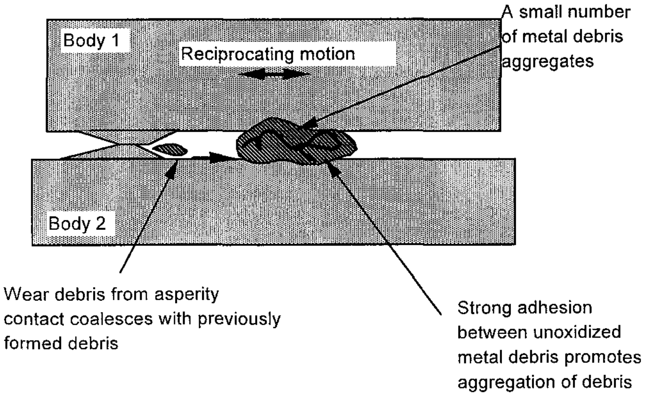
Figure 5.5 Mechanism of corrosive-abrasive wear.

Corrosive-abrasive wear exhibits synergism i.e. a greater materials degradation rate than corrosion and abrasive wear acting separately, when the rate of materials degradation by corrosion is comparable to or greater than the rate of wear when both occur separately. In other words, slow wear when combined with rapid corrosion produces an even more rapid corrosive-abrasive wear whereas moderate corrosion does not significantly accelerate a rapid wear process. This result can be interpreted in terms of the model of cyclic removal of corrosion product films illustrated above. Even a very slow removal of corrosion product films by very moderate wear will have a significant effect on the materials degradation rate whereas corrosion product films never become established on a rapidly wearing surface. The magnitude of the synergism between corrosion and abrasion can be high since, mild abrasive wear that is very slow on a strong metal may be far more rapid on weaker corrosion product films.

5.2.2 *Oxidative and corrosive effects on fretting*

Most sites of fretting are exposed to atmospheric oxygen, which significantly modifies the process of fretting wear. Fretting between steel components in air is accompanied by the release of fine red-colored iron oxide powder from the worn contact once approximately 100 cycles of fretting have elapsed. Microscopic studies of the fretting wear scars reveal that the debris layers on the wear scar surface are heavily oxidized and it is observed that the fretting wear rate in air at ambient temperatures is higher than in vacuum. Reasons for this acceleration of wear are not fully understood and one explanation is that since iron oxide is harder than iron, it acts as an abrasive against the fretted surfaces. The iron oxide powder is probably too fine to act as an abrasive as there is a minimum size of particle abrasivity, which is typically 10 μm. An alternative model of the action of oxygen is that oxidized debris is more friable and brittle than metal debris so that it is more easily expelled from the fretting contact. Fretting wear, which is controlled by the loss of debris from the debris layer, is accelerated by the faster loss of oxidized debris. In the absence of oxygen, metal debris would tend to adhere to the fretting surfaces thus prolonging its retention time in the fretting interface. This model of fretting, which involves formation of wear particles by adhesive and fatigue wear, followed by their prompt removal after oxidation is illustrated schematically in Figure 5.6.

FRETTING OF A BASE METAL IN VACUUM OR AN INERT GAS



FRETTING OF A BASE METAL IN AIR

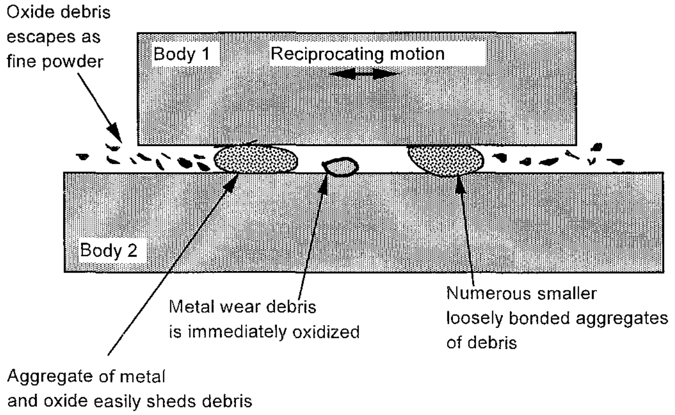


Figure 5.6. Model of acceleration of fretting wear by atmospheric oxygen.

Immersion of the fretting site in a lubricant can suppress fretting wear by two mechanisms; reduces the supply of atmospheric oxygen and suppresses adhesive wear that forms the wear debris.

5.2.3 *Abrasive wear in liquid metals*

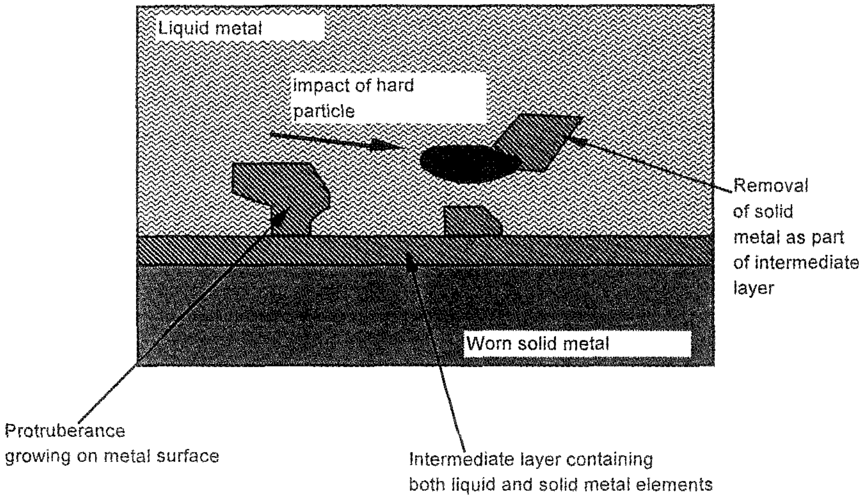
The addition of hard abrasive particles such as alumina grits to a molten metal to create a metal matrix composite is observed to accelerate the wear of stirring rods that are used to homogenize the mixture of grits and liquid metal. Abrasive grits are believed to accelerate the wear of stirring rods by one of two mechanisms. Where a metal stirring rod forms an intermediate layer composed of metal from both the stirring rod and liquid metal, abrasive grits accelerate the detachment of protuberances that extend from this intermediate layer. In most cases of stirring, the grits impact against the rod too slowly to cause direct erosive wear of the stirring rod or the intermediate layer. If a metal does not form a distinct transition layer, abrasive grits may accelerate wear or dissolution of metal by disrupting the boundary layer that surrounds the surface of the stirring rod. This boundary layer constitutes a diffusion barrier to the transport of metal atoms from the stirring rod to the bulk of the liquid metal and disruption of the boundary layer by abrasive grits will accelerate diffusion based transfer of metal atoms. Abrasive detachment of protuberances and disruption of the boundary layer as mechanisms of accelerated wear by grits in liquid metal are illustrated schematically in Figure 5.7.

Wear of stirring rods is undesirable because of loss of the stirring rods and because of contamination of the liquid metal which affects the material properties of the cast metal. Control of such wear appears to involve either use of wear resistant metals such as titanium as opposed to stainless steel, which wears rapidly in liquid aluminum. The use of protective coatings does not appear to have been investigated yet.

5.3 Corrosion fatigue and corrosion embrittlement (SCC)

Corrosion fatigue and the related phenomenon, corrosion embrittlement are fatigue and fracture that are accelerated by corrosion or exposure to corrosive media where there is no apparent corrosion. An important example of corrosion fatigue is the failure of steel welds on offshore structures after prolonged

RAPID FRACTURE OF PROTRUBERANCES FORMED
ON SOLID METAL SURFACE



DISRUPTION OF BOUNDARY LAYER AROUND SOLID METAL

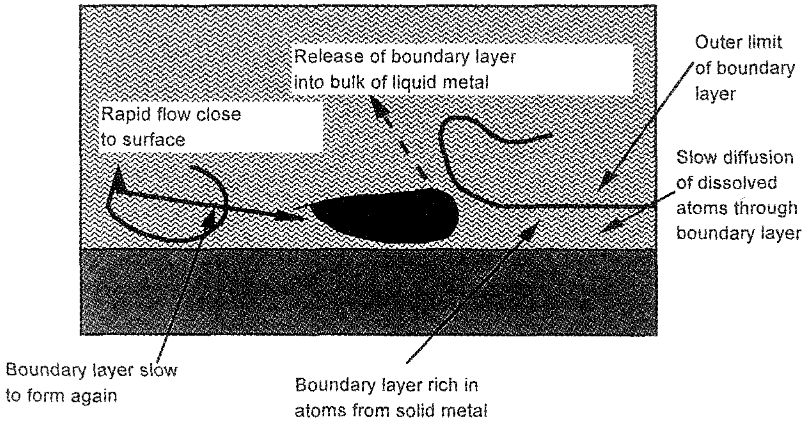


Figure 5.7. Mechanisms of wear by abrasive grits in liquid metal, (i) accelerated detachments of protuberances from the intermediate layer, (ii) disruption of the boundary layer by abrasive grits.

impact by sea waves. It is found that corrosion fatigue denies steel any fatigue limit and that the average lifetime for a given stress is significantly reduced by immersion in seawater. Corrosion fracture is often found in chemical refineries where a strong metal structure that is only moderately loaded fails suddenly upon exposure to a corrosive chemical such as an acid or sulfur compound.

Corrosion fatigue and corrosion fracture are caused by the same underlying physical phenomena but have slightly different characteristics. Corrosion fracture is also known as Stress Corrosion Cracking (SCC) a name, which recognizes the fact that fracture only occurs after a period of exposure to a specific substance, by a component that is loaded to above a critical minimum stress. Examples of corrosion fatigue and fracture are illustrated schematically in Figure 5.8.

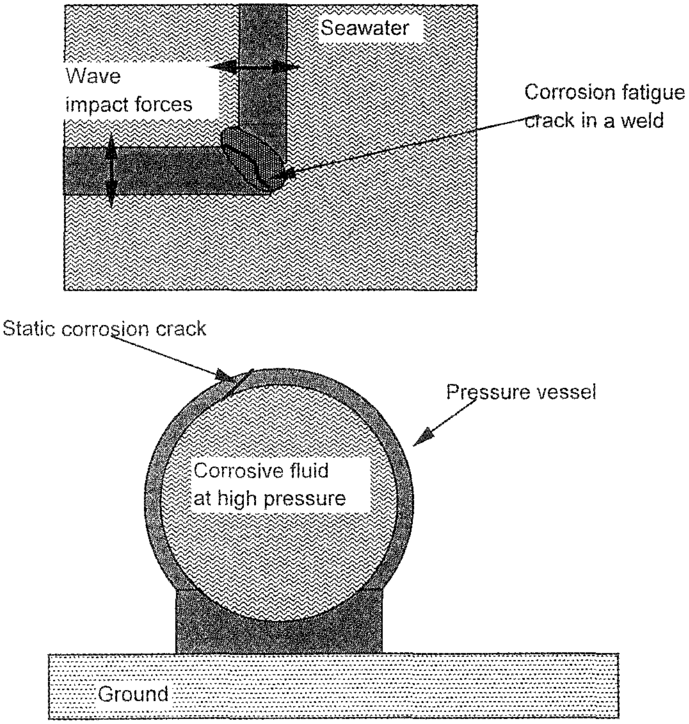


Figure 5.8. Examples of corrosion fatigue and corrosion fracture.

In most cases of corrosion fatigue and fracture, failure of the material occurs before corrosion has significantly reduced the cross-sectional area of critically loaded parts of a component. Bulk degradation of materials by corrosion is therefore not often a significant factor in this mode of fracture. It is still the case however the structures have and will fail due to loss by corrosion of load carrying material. The critical role of corrosion in fatigue and SCC is at the crack tip where crack propagation occurs. According to a model developed by Uhlig who made many important contributions to corrosion science, SCC and Corrosion fatigue are controlled by microscopic interactions between the corrosive fluid and the closed end of the crack. The crack tip is in most cases extremely sharp, which means that the tip is enclosed by only a few atoms or molecules in the material. The bonding strength between atoms in a material is controlled by the electron density between adjacent atoms; when an extraneous molecular species comes into close proximity, the electron density is altered. In almost all instances, the electron density between atoms enclosing the crack tip is reduced by a corrosive fluid or gas. The electrons align with the closely contacting atom or molecule of corrosive substance as part of the preliminary stages of chemical reaction. Even if the reaction does not proceed to completion, the decline in electron density between atoms of the material causes a localised weakening around the crack tip. Crack propagation is therefore accelerated or proceeds at stresses that would be too low to cause crack propagation if the corrosion substance was absent. The role of electron density changes in acceleration of crack growth by corrosion is illustrated schematically in Figure 5.9.

It is also believed that the passivating film will be absent around the crack tip, giving rise to three consequences. The first is as described by Uhlig, direct interaction between the chemical substance and the material to be fractured. The second consequence is enhanced corrosion at crack tip by anodic corrosion. Such corrosion will however tend to blunt the crack tip so reducing stress intensity and crack growth. It is still not precisely known how anodic corrosion would accelerate crack growth. The third consequence is that corrosion products usually have a lower

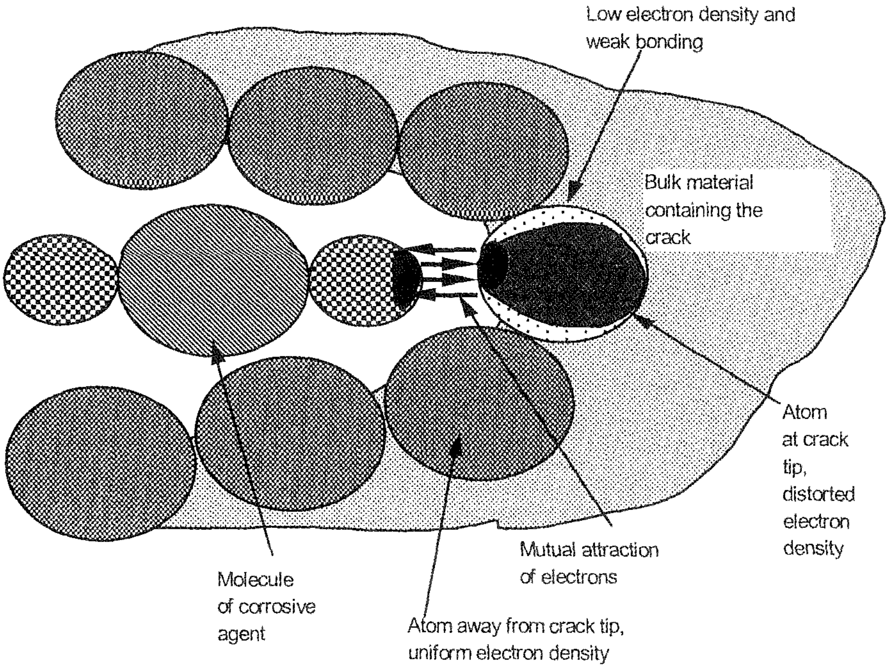


Figure 5.9 Mechanism of weakening of material around a crack tip by molecules or atoms of a corrosive substance.

density than the original material (typically a metal) and these products fill up the crack space. Tensile stresses are generated around the crack due to compression of solid corrosion products trapped between the opposing crack surfaces. Crack enlargement by compressed corrosion products is shown schematically in Figure 5.10.

Another mechanism of corrosion fatigue is the role of corrosion pits as crack initiators. Stainless steels are particularly vulnerable to pitting corrosion; stress concentrations are generated around the pits causing the local stress intensity to exceed the critical value for crack initiation. Acceleration of fatigue crack growth by corrosion is most significant at the initial stages crack growth where corrosion provides a source of small cracks. Even when a pit is not formed directly by corrosion, accelerated initiation of fatigue can occur by corrosion of the microscopic extruded lips of metal that form in the early stages of fatigue, these are known as slip bands.

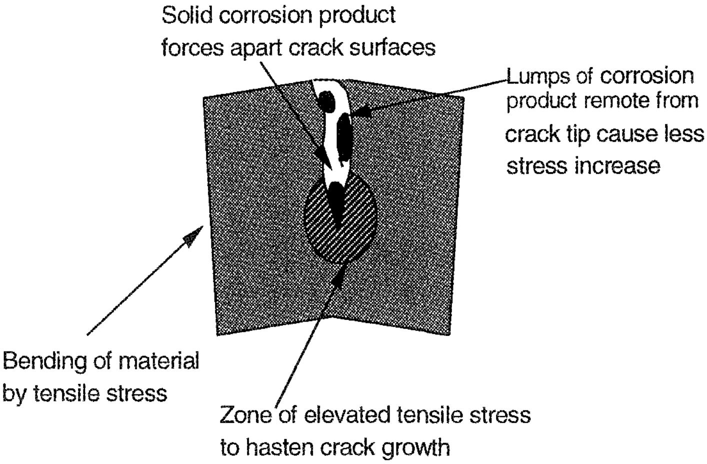


Figure 5.10. Crack enlargement by trapped corrosion products.

Extrusion of the slip band exposes 'nascent' metal, which is unprotected by any passivating film and therefore promptly attacked by any corrosive agent present. It is believed that a fatigue crack is initiated from the microscopic pit formed by such corrosion.

When the crack is large, corrosion does not significantly accelerate crack growth compared to dry conditions. Stress corrosion cracking is not accelerated by pit formation and instead requires an 'incubation period' of exposure to an aggressive chemical before fracture occurs.

Crack growth is clearly related to the possibility of chemical reaction between the material and corrosive substance. It follows therefore that inert materials are more resistant to corrosion fatigue and fracture provided that mechanical strength is not less than the equivalent non-inert material. Many metals, which are passive under static corrosion such as stainless steel and zirconium, are vulnerable to corrosion fatigue or stress corrosion cracking. This is because mechanical stress exposed small fissures in the passivating film, which become highly anodic to the rest of the surface and intense localized corrosion results. The role of surface engineering in coating vulnerable materials with resistant metals or ceramics or tough, low elastic modulus polymers is potentially a very valuable means of controlling this problem.

Stress corrosion cracking and corrosion fatigue are also believed to be related to hydrogen embrittlement as many of the chemical substances involved can release hydrogen upon reaction with the attacked metal. Water, ammonia and acids all conform to this pattern of chemical interaction. Other aggressive elements are sulfur, chlorine and phosphorous, which appear to cause cracking by direct attack at the crack tip rather than by diffusion into the metal as does hydrogen.

5.4 Summary

Interactions between independent degradation mechanisms usually produce a more destructive effect than either acting alone. Instances where one degradation mechanism suppresses another are unfortunately very rare. The most important combined mechanisms of materials degradation involve corrosion with wear and corrosion with stress. Wear also interacts with oscillating stress to produce a very rapid form of mechanical fatigue. The combined causes of materials degradation hinder identification of the specific mechanisms involved in any given problem. If one factor is not recognized, e.g. wear, then it is difficult to find the optimum means of control.

RECOMMENDED READING.

Stress-Corrosion Cracking, Russell H. Jones, editor, publ. ASM, 1992, Ohio, USA.

Stress-corrosion cracking, edited by R.H. Jones, publ. ASM international, 1993, Ohio, USA.

Environment Assisted Fatigue, edit. P. Scott and R.A. Cottis, publ. Mechanical Engineering Publications, 1990, London, United Kingdom.

PART 2: SURFACE ENGINEERING

Overview

In Part 2 of this book, the diverse technologies of surface engineering are described and compared. There are two basic types of surface engineering, providing the substrate with a protective coating and modifying the surface to improve a critical characteristic, e.g. corrosion resistance. Almost every known method of joining two materials together has been considered for adaptation as a method of surface coating. Surface modification is rapidly developing beyond traditional simple processes such as flame hardening to far more advanced solutions to wear, corrosion and fatigue problems. The processes involved in surface coating and modification together with their scope of application are discussed below.

CHAPTER 6

DISCRETE COATINGS

6.1 Introduction

Discrete coatings range from the familiar paint, varnish and enamel to advanced coatings, which can only be produced under vacuum with a sophisticated electrical control system. The main types of coating are list schematically in Figure 6.1.

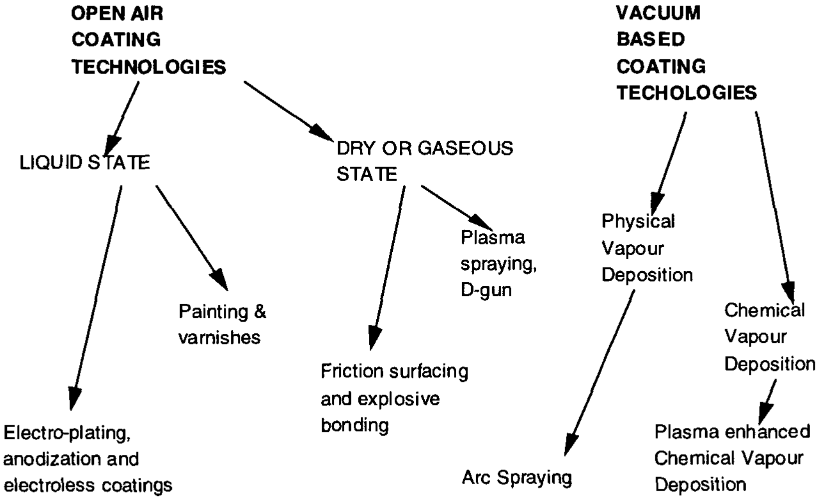


Figure 6.1 Classification of the different types of discrete coating technologies.

Despite the seemingly complex and disparate range of coating technologies, all of these coatings serve the same objective, which is to physically separate the vulnerable substrate material of a component from external damage processes. An engineer should know at least in general terms, answers to such questions as: Which coating offers a particular form of protection, e.g. wear or corrosion? How is the coating to be generated? What resources of material and equipment are required for the coating process? The chapter below is intended as a guide to answering these questions.

6.2 Coatings of organic compounds

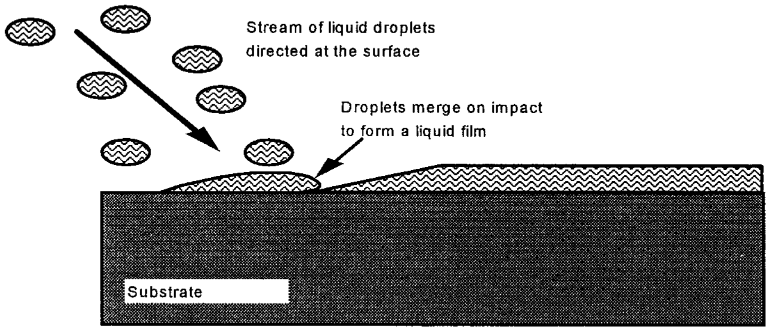
Paints, varnishes, layers of epoxy resins comprise coatings of organic compounds. Paints and other organic coatings are deposited on a surface in a liquid state to form a liquid film that is distributed over a surface-by-surface tension forces. Polymerization reactions in this liquid film that are initiated by contact with e.g. atmospheric oxygen, ultra-violet light then converts the liquid film to a more durable solid film. The mechanism of paint film formation is shown schematically in Figure 6.2.

An organic coating consists of several components some of which function during the formation of a solid coating and others, which provide special properties to the coating during service. The ingredients of a paint or liquid precursor of the organic coating contains can be summarized in the equation given below:

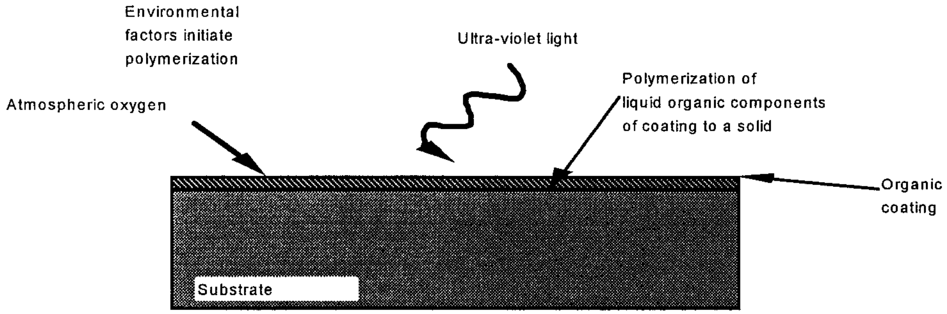
Paint = Solvent + Binder resin + Pigments + Extender + Dryers + Special purpose additives.

Solvent

The purpose of the solvent is to provide liquid medium during storage and deposition. The solvent can be an organic fluid such as white spirit, water or a mixture of both. Other organic solvents used are xylene, toluene, ketones, alcohols and halogenated compounds. The health hazards and fire risks associated with organic solvents have lead paint technologists to extend the use of water as a solvent from traditional simple water-based paints to more advanced paints. During coating, the solvent evaporates after a short period of exposure to air and serves no further purpose to the organic coating.



LIQUID STATE DEPOSITION OF COATING



POLYMERIZATION TO FORM SOLID COATING

Figure 6.2 Mechanism of paint film formation.

Binder resin

The binder resin polymerizes or 'cures' to form a solid matrix for the organic coating. Phenolic resins, cellulose derivatives, melamine derivatives, epoxy resins and elastomeric resins are all used as binder resins. The type of resin used has a strong influence on the mechanical strength and durability of the organic coating and paints are usually classified according to the resin used. The different types of commonly used resins are listed below in Table 6.1 together with their manufacturing process and key characteristics.

Resin Type	Manufacture and characteristics
Polyester and alkyds	Formed by reaction between polyfunctional organic acids and polyhydric alcohols. Stable in storage, cheap and versatile, hence high usage.
Amino	Formed by reaction between formaldehyde and an amino compound, e.g. urea. A second reaction with an alcohol produces the final resin. Used in combination with other resins, e.g. alkyd to produce hard coatings with good resistance to chemicals.
Acrylic	Formed by polymerization of methyl and ethyl methacrylate. Combines moderate cost with excellent durability in outdoors applications.
Polyurethane	Polymer of urethane, which is the product of isocyanate and alcohol. Tough, abrasion and chemical resistant, cures at low temperatures. Degraded by water.
Vinyl	Vinyl acetate, vinyl toluene and vinyl chloride. Used for water-based emulsion paints. Vinyl coatings are resistant to ultra-violet light and water, suitable for outdoor applications
Phenolic	Formed by reaction between phenol and formaldehyde or other aldehydes. Good resistance to chemicals, water and organic solvents.
Cellulose derivatives	Cellulose ethers, e.g. sodium carboxymethyl cellulose. Can be used as a basic resin but more commonly used to modify rheology, stability of water based paints, e.g. acrylic paints.

Oleoresinous	Formed by heating a mixture of unsaturated oils (plant origin) and natural or synthetic resins to form a solid that is soluble in organic solvents An obsolescent formulation but still useful.
Epoxy	Formed by reaction between diphenyol propane and epichlorhydrin in alkaline conditions. Hard, resistant to most chemicals and solvents. Cracks if substrate is deformed. Suitable for elevated temperatures. Strong bonding to most substrates.
Elastomer	Chlorinated rubber coatings. Tough and corrosion resistant. Fast curing time and good durability in storage. Poor surface finish and difficult to apply by brush. Softened by animal fats and oils. Latex (source of elastomer) often combined with other resins, e. g. acrylic.

Table 6.1 Categories of paints and their function.

Pigments

Pigments serve to provide color to the paint but may also provide an anti-corrosion effect. Inorganic metal salts, e.g. zinc phosphate and zinc chromate is widely used as pigments. Titanium dioxide is used for white color in paints because of its brilliant whiteness and graphite is used for black. Organic compounds are however used to produce brilliant primary colors, as inorganic compounds do not provide a high quality of color. Extenders are used to give bulk to paint while reducing its unit cost, which can be high, if expensive pigments such as titanium dioxide are used.

Extenders, dryers and special purpose ingredients

Commonly used extenders are kaolin (white clay), mica, talc, calcium carbonate (chalk). Dryers are added in small quantities to the paint to control the drying and polymerization process. Dryers are usually fatty acid esters of cobalt, manganese, lead, vanadium and other metals. Special purpose ingredients can be fungicides to prevent moulds from growing on the paint surface.

6.2.1 Utility of available organic coatings for different conditions

The surface area of material protected by paints and other organic coatings far exceed the surface area of any other coating method. The range of size painted components is also much larger than most other coating technologies as paints are used to protect large structures such as bridges as well as tiny electronic components. Paints and varnishes can be effectively deposited by manual brushing or spraying which lowers costs and enlarges the range of components that can be so coated. There is no requirement for specialized and restrictive technology as is the case with vacuum deposition for example. Paints are widely used even where a different type of coating would provide a better technical quality of materials protection because paints are cheaper and more convenient to use. Paints and varnishes do not normally react with a substrate, which is typically metal, so that thin sheets of metal can be coated without change of mechanical properties. Painting is normally performed at ambient temperature so that thermally induced microstructure changes of the substrate cannot occur. Other coating methods, such as 'hot-dipping' in liquid metal can degrade thin sheets of metal by thermal stress and liquid metal corrosion.

Paints and other organic coatings serve a wide variety of purposes additional to protection of materials. An important use is decorative or graphic. Bold lettering; brightly colored vehicles and buildings are examples of the graphic and decorative nature of paint. The primary purpose of organic coatings is corrosion protection where decorative considerations are of little or no importance. A common example of this type of coating is the underbody paint of a car. A more recent development is wear-resistant coatings of strong polymers. While polymer coatings cannot be expected to have the wear resistance of hard metal coatings, advances in polymer technology have achieved significant advances in resistance to abrasive wear or corrosive-abrasive wear in a slurry. There are also special purpose organic coatings, e.g. paints blended with anti-septic or biocide for use on hospital walls and ceilings or low friction 'slippery ' paints for security needs. In this last case, the low friction paint can prevent a thief from climbing up the outside drainpipe. The range of applications of organic coatings can be divided into 6 basic categories, decoration, protection of wood, vehicles and aircraft, industrial use, marine use and anti-corrosion where each category denotes the primary purpose or context in which the coating is used.

Industrial use coatings

Industrial use coatings are typically used to prevent a metal component from corrosion in a severe environment. An example of the use of such coatings is the protection of sewage processing or oil refinery equipment. In this case, resistance of the organic coating to solvation by organic compounds, extremes of acidity or alkalinity and temperature are critical.

Wood

Wood is prone to damage by moulds and the use of coatings (paints) suitable for metal causes mould induced blistering of the paint. It is important to use coatings that inhibit moulds.

Marine use

In marine conditions, resistance to salt water corrosion is paramount. The organic coating should also be resistant to the mechanical stresses of impact by water in the 'splash zone' at the surface of the sea. A biocidal effect on adherent marine organisms such as barnacles is also extremely desirable.

Vehicles and aircraft

Vehicles and aircraft require resistance to chipping or blistering from impact by solid and liquid particles. When a vehicle is traveling at full speed, road stones may hit the organic coating of the vehicle with a velocity of approximately 30 m/s. The comparable coating of an aeroplane traveling through cloud is subjected to impact by water droplets at speeds of 200 m/s or more. Good resistance to ageing by ultra-violet light and a high quality of surface finish are also important. Supersonic planes sustain considerable kinetic heating from the air so that temperature resistance to approximately 100°C is needed.

Anti-corrosion coatings

Anti-corrosion coatings refer to organic coatings, which do not merely act as passive separators of metal substrate and corrosive reagents but also actively suppress corrosion. These coatings usually contain inorganic compounds such as zinc chromate, which suppress the electrochemical corrosion of iron.

Paints are commonly classified according to their resin as this forms the bulk of the paint. The most common type of paint resin is the polyester resin as it is cheap and versatile; the other resins cover specialist purposes. Epoxy resins are particularly tough and intended to severe specialized applications such as the protection of underground or underwater pipelines. Automobile applications require resins that confer a very high surface finish; polyester resins do not produce a sufficiently high surface finish to satisfy consumer demand.

Resin system	Usage Decorative	Industrial	Marine	Anti-corrosion	Wood	Vehicles and aircraft
Polyester	Good	Good	Good	Good	Good	-----
Polyurethanes	-----	Good	-----	Good	Good	Good
Acrylic	Good	Good	-----	Good	Good	Good
Vinyl	Good	Good	-----	-----	Good	Good
Phenolic	----	Excellent	----	----	----	----
Cellulose derivatives	----	----	----	----	Good	Good
Oleo-resinous	----	Good	----	----	Good	----
Epoxy	----	Good	Good	Good	----	Good
Elastomer	----	----	----	Good	----	----
Amino	----	Good	----	----	Good	---

Table 6.2. Usage of organic coatings as a function of resin type

6.2.2 Methods of depositing organic coatings

The traditional methods of deposition of organic coatings (painting) are brush painting, rolling of an adsorbent roller soaked with paint and the spraying of an aerosol of paint. The technology of wet painting is not particularly complicated, the main problems being to avoid wastage of paint when spraying and prevent health problems in workers due to prolonged exposure to the organic solvents in the paint. Repetition of painting is needed to obtain a thick coating because if a thick coating is deposited in one stage, flow of the paint or organic coating can occur. Flow problems are particularly acute on vertical surfaces where it is known as *vertical sag*. The optimum surface condition for paints is a cleaned surface, free of grease and residual corrosion products with a moderate degree of surface roughness, which provides the paint layer with a mechanical interlock to the substrate. When water-based coatings are used, it is essential to electrically isolate the painted surface from any power source and a higher level of substrate cleanliness is required. For large volume painting in mass production, electrostatic spraying is used. The largest application of electrostatic spraying is the painting of car bodies. Car bodies are passed through a tunnel shaped painting booth where jets of negatively ionized paint are directed at the car bodies, which are electrically grounded. The strong electrostatic attraction between the paint particles and the car body is intended to ensure that the paint covers the entire car body without a shadowing effect. The complex shape of the car body contains numerous hidden surfaces, which would remain unpainted or 'shadowed' without electrostatic spraying.

The environmental problems associated with solvents have prompted innovations in minimizing the use of solvents. Two types of organic coatings with low solvent contents have been developed; these are *high-solid paints* and *powder coatings*. High solid paints contain a reduced fraction of liquid phase that is less than 30% by weight of the paint. High solid paints can be deposited by the methods described above but powder coatings require a multistage process where an energy source is used to convert the powder to a uniform coating. The powder coatings are deposited on the substrate by electrostatic spraying to ensure uniform coverage. Conversion of the power coating to a continuous layer of paint is achieved by exposing the substrate to a source of electromagnetic radiation such as ultra-violet light or even X-ray. This process is known as *radiation curing*. The objective of using such radiation is to melt or reconfigure the powder coating without excessive heating of the substrate. The lack of a solvent means that the organic coating compounds need not be soluble in solvent, this allows high molecular weight compounds to be included in the coating formulation. Coatings that are much tougher, harder and more wear resistant than traditional organic coatings can now be made from high molecular weight organic compounds. Most applications of powder coatings relate to the large-scale manufacture of metal items such as cars. Sophisticated coating technologies in the controlled environment of a factory become economically feasible and result in a higher product quality.

6.2.3 Corrosion problems of organic coatings

A major disadvantage of paints is that they are not completely impermeable to water and oxygen. Diffusion of water and oxygen through a thin paint film to the underlying metal surface causes eventual corrosion of the metal. It is found from measurements of the corrosion current on painted metal specimens exposed to salt spray that painting never completely prevents flow of corrosion current from anodes to cathodes present on the test specimen or connected externally. Damage to paint, known as *cathodic blistering* occurs on painted surfaces that are also protected with an imposed cathodic voltage. It appears that oxygen diffuses across the paint film to become reduced by the metal to produce alkali hydroxyl ions. A concentrated solution of hydroxyl ions accumulates at the interface between the paint and metal Blistering of the paint, i.e. the formation of voids between paint and metal, then occurs.

Despite the severe degradation of iron and other metals by electrochemical corrosion, electrochemical reactions are not necessarily destructive to metals. Many valuable metal coating processes are based on cathodic and anodic electrochemical reactions.

6.3.1 Electroplating of metals

If a metal object is placed in an aqueous solution of a salt of the same metal or a different metal and supplied with a strong cathodic electric potential, electroplating of the metal object may occur. Electroplating is caused by cathodic reactions leading to the reduction of metal from solution and evolution of hydrogen or oxygen. Typical examples of electroplating mostly involve iron or steel as the substrate and metals such as nickel, chromium and cadmium as the plating metal. Electroplating is performed in large tanks of metal salt solution with an attached power supply and immersed lead sheet that functions as the anode. The range of metals that can be electroplated is limited and electroplating of metal alloys presents considerable difficulty. The formation of a metal film on a substrate from metal ions in aqueous solution is a complex process that is still not fully understood. In most instances, this film formation process functions well without external intervention and is therefore usually overlooked. It is generally believed that electroplated coatings begin as small nodules of deposited metal, which then grow laterally to form a coating. The coating contains flaws at the boundary where the growth of one nodule joins the growth of another nodule. The principle of electroplating and structure of electroplated coatings are illustrated schematically in Figure 6.3.

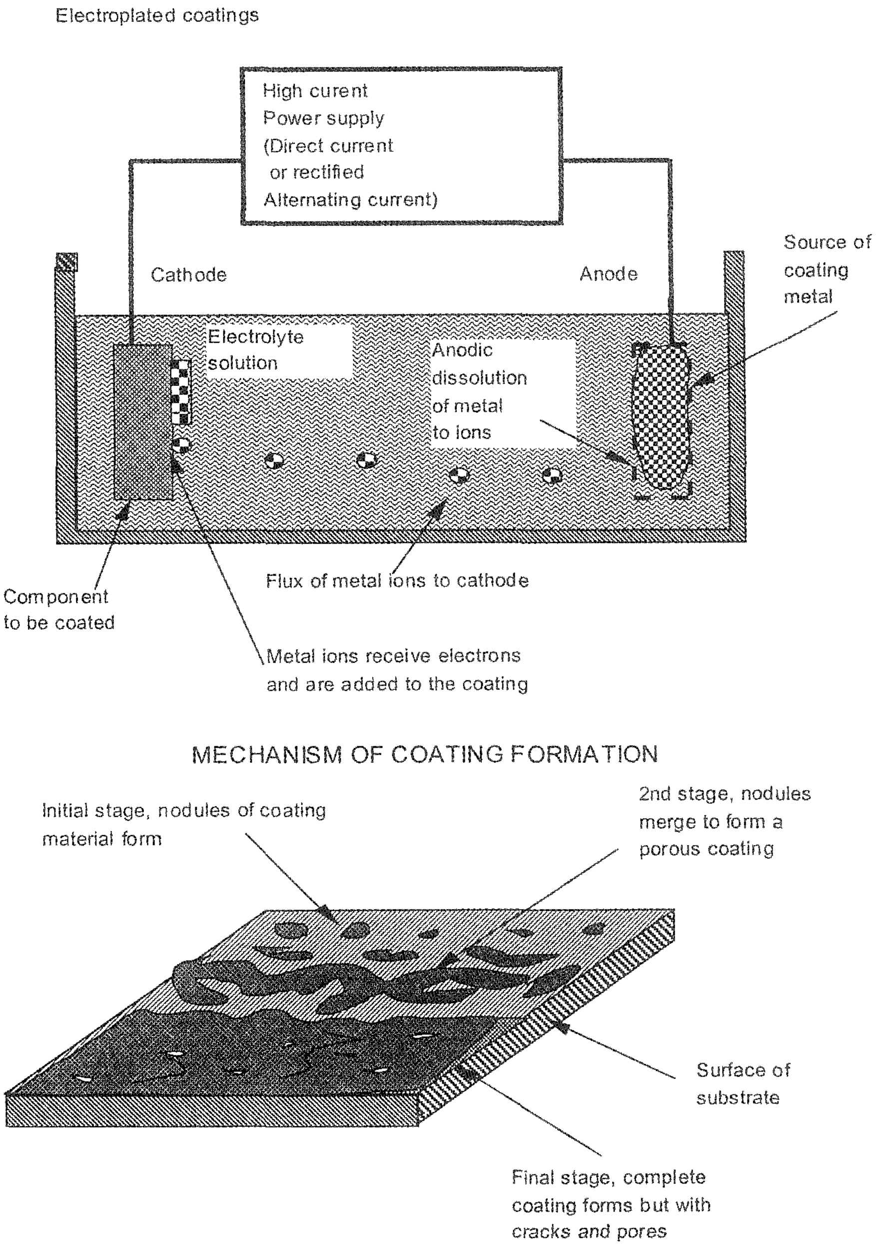


Figure 6.3. Mechanism of electroplating and structure of electroplated coatings.

Electroplated coatings tend to be rich in hydrogen from the cathodic reactions and the hydrogen can often cause microstructure problems in the electroplated coating such as brittleness. Another problem of electroplated coatings is non-uniformity of coatings. Electroplating is dependent on the local strength of electric field between anode and cathode and varies with location in an electroplating cell. Corners, edges and projections generate the most intense electric field and so are coated more rapidly than plane surfaces. Practical considerations dictate a uniform coating time for all parts of an electroplated substrate with the result that thick coatings are formed at edges, corners and projecting parts of the component. A recent development in electroplating tries to combine the advantages of oldest method of electroplating and the conventional method described above. The older method uses spontaneous redox reactions to deposit a film on both insulating surfaces and metallic surfaces and the control of deposition conditions are difficult. The recently developed electroplating technique permits coating of insulating substrates with metals having controlled grain size, thickness and growth speed. The method uses progressive outward growth of the metal from an electrode in contact with the substrate, with the cell geometry chosen so that the electron current providing the reduction passes through the growing deposit [Vincent et al.].

Chromium plating

A very common application of electroplating is chromium plating for decorative and industrial purposes. Chromium plating is performed with a chromium salt solution and a passive anode made of lead. Trivalent chromium solutions are limited to low deposition rates and are mainly used for the deposition of thin decorative coatings. Hexavalent chromium solutions provide higher deposition rates than trivalent chromium solutions and are used for the production of thick wear-resistant hard chromium coatings. Chromium solutions are used in high concentrations and at an elevated temperature of 50 to 80 °C in order to achieve rapid coating concentrations, e.g. 250 g of chromium oxide per liter of sulfuric acid solution. Current densities are of the order of 100 Amperes per square meter for fast plating of chromium. Chromium plating is prone to cracking where the plating thickness is greater than 0.5 μm and it is believed that cracking is caused by the formation of chromium hydride, which later decomposes to pure chromium. The density of pure chromium is higher than chromium hydride so that shrinkage cracks result. The physical structure of chromium plating consists of metal grains separated by small pores and fissures. This structure results from the mechanism of formation of chromium plating where granules of chromium independently grow to cover the substrate. The hardness of chromium plating can reach 1000 Hv hence its used as a wear resistant coating.

A major problem with chromium electroplating which is predicted to severely restrict its future usage is toxicity. Inorganic compounds of either trivalent chromium or hexavalent chromium are used as the source of chromium. Trivalent and hexavalent chromium ions are toxic and hexavalent chromium is carcinogenic while the lead anode is also toxic. There are health risks to electroplating personnel and problems with disposal of liquid and solid waste from electroplating. Hard chromium coatings are effective as anti-wear coatings but no more so than the more recently developed coatings produced by vacuum and plasma technology.

6.3.2 Anodising, electroless coatings , chromating and phospating

Anodizing

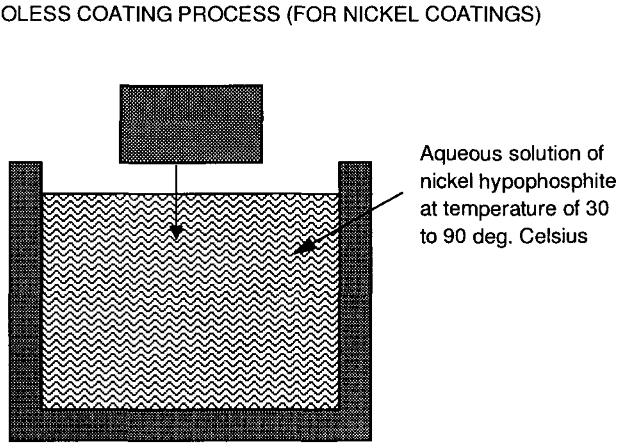
Anodizing as its name suggests, is the application of anodic electrochemical reactions to form a coating on a metal substrate. As discussed in Chapter 3, a metal may passivate when subjected to a large anodic potential. This passivation is usually caused by the formation of a film of anodic reaction productions that block further access by the electrolyte to the metal electrode. This passivation persists even when the metal object is no longer the anode but installed elsewhere as a component or structure. Anodizing is most widely used to protect aluminum alloys where a thick layer of aluminum oxide is deposited on the aluminum surface. Aluminum is immersed in an aqueous acid solution e.g. 1 molar sulfuric acid and connected to an anodic potential of approximately 5 Volts. Aluminum oxide is barely soluble in acid solutions so that anodic dissolution of the aluminum does not occur and instead a dense aluminum oxide film forms on the aluminum surface. It is not necessary to always use a strong acid, immersion in a weak organic acid, e.g., oxalic acid, allows the formation of a corrosion and wear resistant layer of amorphous aluminum oxide. Coating formation rates are approximately 1 μm per minute at a current density of 250 Amperes/square meter and coating thickness of 20 μm are typical.

As discussed in Chapter 3, aluminum oxide films confer a natural resistance to corrosion by aluminum and anodizing is essentially an enhancement of this corrosion protection mechanism. During anodizing, coloring agents are often added to the electrochemical solution in order to produce a colored decorative aluminum oxide film. Anodizing is effective at preventing slow outdoor corrosion or weathering of aluminum alloys and so is widely used in the manufacture of aluminum window frames and other building components.

Electro-less coatings

Electro-less coatings are very similar to electroplated coatings with the exceptions that a power-supply is not necessary and that the substrate need not be metal. A wide variety of applications for metal coatings on plastic and ceramic substrates are found in the electronics industry and elsewhere. The most common example of electro-less coatings contains nickel but gold, tin and copper coatings are also produced commercially. The nickel is deposited in pure or alloyed form with boron, phosphorous and tantalum for enhanced wear and corrosion resistance.

The basic principle of electro-less coatings is the spontaneous deposition of metal from a solution of hypophosphite. The mechanism of electro-less coating is illustrated schematically in Figure 6.4.



ELECTROLESS COATING PROCESS (FOR NICKEL COATINGS)

Aqueous solution of nickel hypophosphite at temperature of 30 to 90 deg. Celsius

Figure 6.4 The electroless coating process.

There are two modes of deposition, the slow mode at approximately 30°C in an alkaline hypophosphite solution and the fast mode at approximately 90°C in an acid solution of hypophosphite. The slow mode is used for deposition on polymers while the fast mode is used on metal and ceramics. This simple coating process of dipping a component into a bath of hypophosphite produces coatings with a surprisingly high wear and corrosion resistance. Coating thickness is controlled by the time of immersion; the slow mode having a coating rate of approximately 1 micrometer per hour while the fast mode has a coating rate of approximately 10 to 30 micrometers per hour. A basic problem of electro less coatings is 'edge pull-back' where the coating detaches from any underlying edge. Edge pullback is caused by stresses developed in the electro-less coating during its formation and cannot be completely prevented. Designs of components for electro-less coatings should avoid sharp edges as much as possible.

Chromating

Chromating, which is the deposition of a metal chromate protective layer, is widely used on zinc components for protection against salt water. The zinc component to be chromated is immersed in a bath of e.g. lithium chromate, which is maintained at a low pH of approximately 2. A zinc chromate layer then forms by anodic dissolution of zinc and the subsequent reaction of the zinc ions with chromate ions to form zinc chromate. Corrosion protection is obtained by the inhibition of electrochemical reactions where oxygen is reduced to the oxide anion. The release of electrons from the zinc during anodic dissolution is suppressed by the lack of conversion of oxygen to oxide anions. The conversion of oxygen to oxide anions would otherwise consume the electrons released by the ionization of zinc.

Phosphating

Phosphating is an application of passivation reactions that are found in aqueous corrosion. The substrate is placed in an aqueous bath of phosphoric acid at ambient temperature or else is irrigated with a phosphate solution. The latter method is useful for large structures. Passivation is enhanced by control of acidity to ensure extensive deposition of the passivating corrosion products. The phosphoric acid initially dissolves some of the metal to generate metal phosphates, this corrosive attack is acceptable because phosphoric acid is a much weaker acid than hydrochloric acid. The acidity of the bath is then reduced by adding alkali, this causes metal phosphate to be precipitated onto the substrate as a passivating layer. In more advanced treatments, metals other than that of the substrate are added in a 3 stage process which is listed below:

Stage 1 Initial acidic attack by phosphoric acid to generate intermediate phosphates.

Stage 2 Conversion of intermediate phosphates to protective phosphates by addition of metal salts.

Stage 3 Lowering of acidity to deposit passivating phosphate layer.

When the substrate is made of steel, the intermediate phosphates will be iron phosphate (usually ferrous phosphate) because the substrate is mostly iron. Iron does not generate a useful passivating layer because instead of iron phosphate, a mixture of mostly iron oxides, hydroxides and some iron phosphate is formed. To ensure a better quality of passivation, zinc, manganese and calcium phosphate coatings are produced. Manganese phosphate coatings have a further advantage of also being able to function as solid lubricants. The use of different metals does however add to the cost of phosphating.

The main benefit of phosphating is as an undercoating to paints where a synergistic enhancement of corrosion resistance is obtained. Moisture and oxygen always penetrate paint layers by diffusion to initiate rust beneath the paint layer, which then eventually detaches. A phosphate undercoating prevents formation of rust and ensures continuing adhesion of the paint layer to the substrate.

6.4 Plasma and thermal spraying, plasma-transferred arc, the D gun

Electroplating, chromizing, phosphating etc. are all specialized techniques that impose severe limitations on the range of coating materials and on the physical shape and size of the coated component. Plasma spraying and related techniques are much more versatile and have allowed the scope of coatings to be greatly expanded to serve a much wider range of critical applications.

6.4.1 Plasma and thermal spraying

Plasma spraying and related coating techniques represent a means of conferring the advantages of liquid state deposition to high melting point solids. Paints, which are the most common example of liquid state deposition, enable the transfer of material as discrete droplets from coating source to substrate. Once the coating material has reached the substrate, the droplets of coating remain liquid for a sufficient period of time to merge and form a continuous surface layer. With plasma spraying, the temperature of the spray is raised to several thousand degrees Celsius so that any solid, even the refractory aluminum oxide, is melted and deposited in a liquid state. The temperature of the electric arc is sufficient to at least partially convert the carrier gas and the coating material to a plasma, hence the name plasma spraying. The required high temperatures are achieved by arranging for a carrier gas and coating material to pass through an intense electric arc or an Oxy-acetylene flame. The stream of heated gas or flame of burnt gas in the case of thermal spraying is directed at a surface to produce a coating. The molten droplets contained in the spray impinge against a metal or ceramic substrate before cooling to form a solid coating on the surface of the substrate. The substrate absorbs heat from the molten coating material and must be sufficiently massive to prevent it from becoming very hot. The heat of the droplets will result in a heat-affected-zone (HAZ) at the surface of the substrate.

Plasma spraying can be used to deposit most metals and ceramics on a substrate, which is often, but not always steel. Thermal spraying, which requires cheaper equipment than plasma spraying is effective with most metals but not all ceramics, e.g. aluminum oxide. This is because the flame temperature is lower than the plasma temperature. Both these coating methods offer a fast and convenient means of depositing thick coatings of metals and ceramics on almost any shape of substrate. Deposition rates in plasma spraying typically range from 1 to 10 kilograms of coating per hour, this is approximately 1 mm coating thickness over 1 square metre of coating area. Deposition rates of metals are slightly faster than deposition rates for ceramics. Plasma spraying offers a faster coating rate than most other coating techniques.

The principal limitations are that the substrate must have a moderately high melting point. This excludes polymers and low melting point metals such as lead from becoming substrates. This restriction on substrate materials is not of much practical significance as steels or titanium alloy are the most common substrates. The range of applications of plasma and thermal sprayed coatings is very wide, wear-resistant coatings; corrosion-resistant coatings, thermal barrier coatings and reclamation of worn or corroded components are the major applications. Reclamation of damage components involves the spraying of material to restore the original dimensions of the component. Plasma spraying is often used to produce aerospace and nuclear power components. A new application is the spraying of hydroxyapatite onto the stems of orthopedic endoprostheses ('artificial joints'). The hydroxyapatite is a mineral constituent of bone and the coating provides a favorable site for bonding between bone and implant. If there is no coating, bonding is weak because true contact between bone and metal does not occur; instead a thin necrotic layer separates the two. The process of plasma spraying is illustrated schematically in Figure 6.5.

Plasma spraying operates at high energy levels, the power consumption of a typical coating unit is approximately 50 kW with an electric current of several hundred amperes flowing in the plasma arc. The impingement speed or speed at which molten particles are ejected from the plasma is typically 300 m/s with a

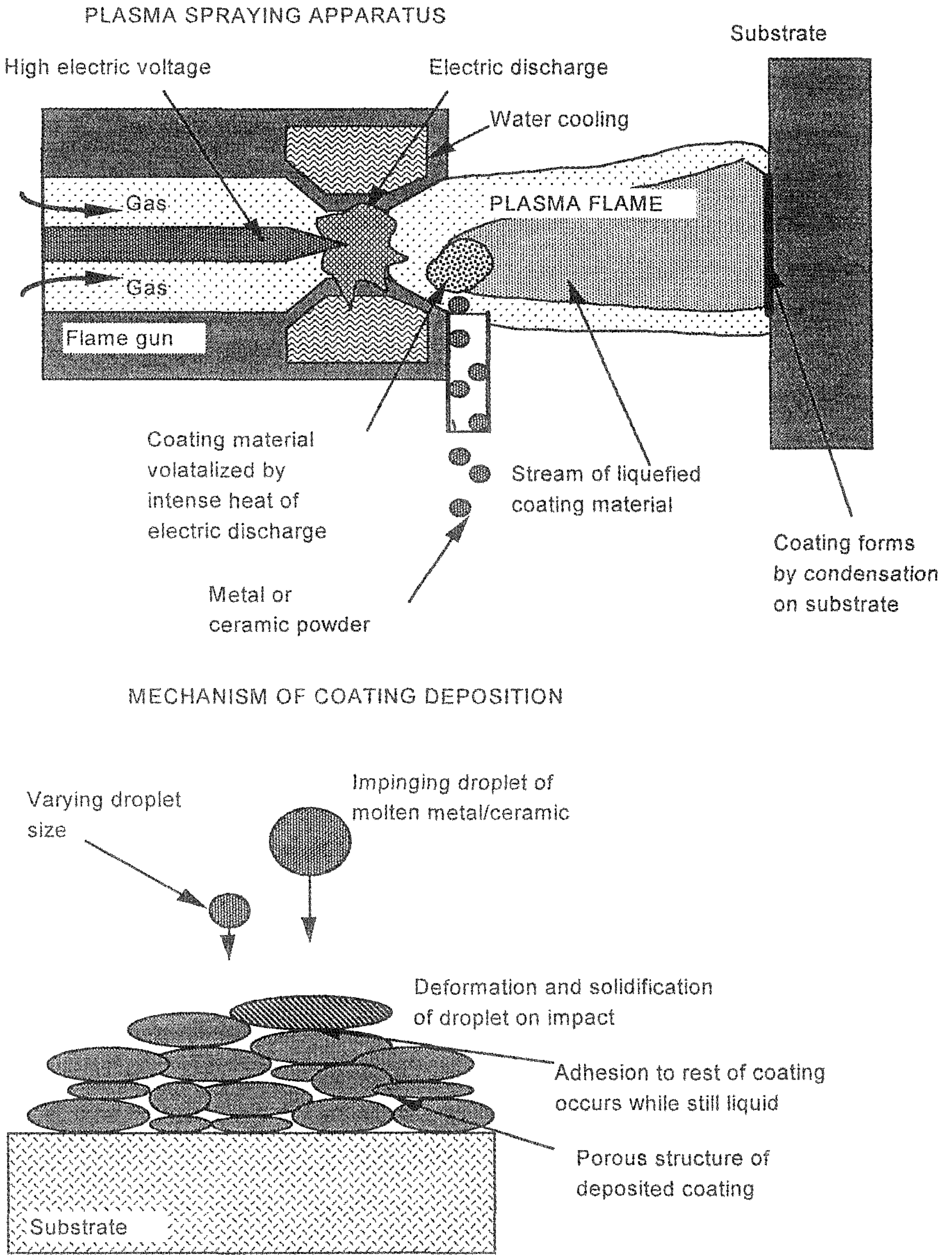


Figure 6.5. The Plasma spraying process.

maximum value of 600 m/s for advanced plasma spraying systems. These high impingement speeds are intended to ensure that the molten droplets disintegrate on contact instead of remaining as discrete droplets, this process is known as splat formation. The structure of a plasma -sprayed coating can be envisaged as a series of interlocking splats.

The distance between the plasma spraying unit and the substrate can be allowed to vary by up to 10 millimeters without significant effect on the coating quality. Components with a complex shape can be coated by plasma spraying without the need for precise tracking of the plasma sprayer around the component. A significant problem with plasma spraying is the dependence on powdered coating material. Metal powder is usually more expensive than monolithic metal and is more difficult to exclude from contamination. Optimum powder size is approximately 10 to 50 μm and there are major health and combustion risks associated with metal powders of this size. To control these hazards, plasma spraying is often performed inside a closed chamber filled with inert gas. The use of a chamber prevents the escape of any polluting or toxic debris from the plasma spraying and can be used to suppress oxidation of the metal powder during plasma spraying. The hot molten droplets are extremely susceptible to oxidation (if they are metal droplets) even if conveyed in an inert carrier gas. Oxide contamination can greatly weaken the sprayed coating as oxides act as brittle inclusions within the microstructure. The splat structure of a plasma sprayed coating and the destructive role of oxide inclusions are illustrated schematically in Figure 6.6. Adhesion mechanisms between the coating and interface are also illustrated in Figure 6.6.

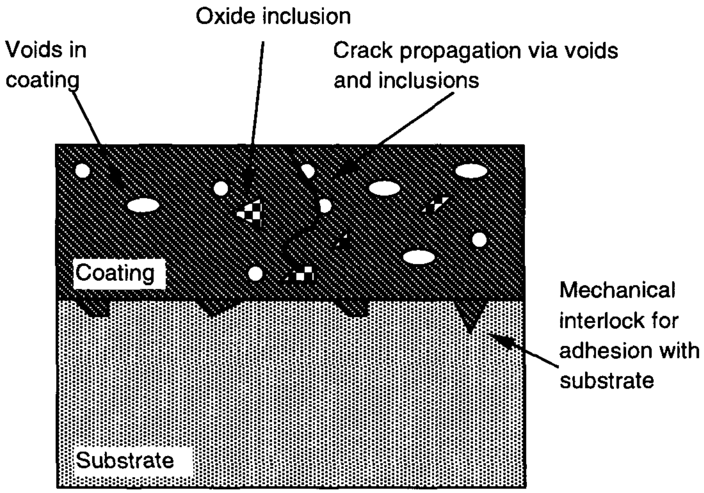


Figure 6.6. Structure of a plasma sprayed film.

The strength of adhesion or bonding between coating and substrate is critical to the performance of the coating. With plasma spraying, it is believed that the bonding is obtained from mechanical interlock and solid-state adhesion between atoms in the coating and substrate. Mechanical interlock is a physical keying between an often deliberately roughened substrate and a coating that is in very close contact with the substrate. Solid-state adhesion involves either electron transfer between the outermost atoms of the coating and the substrate or the formation of a much thicker layer of a compound material. It is modeled that the molten droplets dissolve or sweep away any oxides and contaminants that cover a metal surface. The cleaned metal substrate then reacts with the molten droplets before they solidify.

There are many coating processes related to plasma spraying. Thermal spraying as previously mentioned, is a very popular alternative to plasma spraying, as the required equipment is essentially an upgraded welding torch with an attached powder feed system. This equipment is many times cheaper than a plasma spraying system and allows coating to be performed in an unspecialized workshop. Oxygen and fuel gases such as hydrogen or propane are used to produce a flame and the flame velocity can reach approximately 800 m/s in what is termed the High Velocity Oxy-fuel systems (HVOF). Other variants are a wire-feed plasma spraying system where metal wire not powder is used as the coating material. This avoids the costs and risks of powder but is not suitable for ceramics as these cannot be formed into a wire.

Plasma and thermal spraying of corrosion resistant metals is found to be an effective, practical method of protecting steel surfaces from high temperature gaseous corrosion. A major problem with plasma and thermal spraying is the porosity of the coatings, which arises from the 'splat' mechanism of deposition. After sustained exposure to e.g. sulfurous boiler gases, it is found that gas has penetrated the coating to initiate corrosion at the surface of the underlying steel. Sustained corrosion of the steel then detaches the coating in a manner analogous to paint blistering caused by aqueous crevice corrosion

Thermal Barrier Coatings

A very important application of plasma spraying is the manufacture of Thermal Barrier coatings, which are used to protect metals from the extreme heat of flames. These coatings are widely used to protect gas turbine blades and rocket motor nozzles from burning up during use. A developing application is for a protective coating on piston heads of internal combustion engines. Demands for higher power per unit of compression volume necessitate operation of the piston heads at higher temperatures where the aluminum alloy used for pistons will fail if unprotected. A thermal barrier coating deposited on the head of the piston protects the aluminum substrate against the heat of combustion during operation of the engine.

Thermal barrier coatings are thick layers of heat resistant inorganic materials that protect the underlying metal from high temperature oxidization, sulfidization or chloridization as well as extreme temperature transients. The thickness of the coating can be as large as 1 mm. A thermal barrier coating is typically made of a refractory inorganic oxide since oxides cannot usually be corroded by oxygen. Other important properties of thermal barrier coatings are resistance to other chemical reagents, e.g. sulfur and durability against thermal shock (rapid temperature rises). The most commonly used oxides for thermal barrier coatings are blends of zirconia and yttria, which have a low thermal conductivity and a coefficient of expansivity that is similar to most metals. The yttria is present as a minor constituent of concentration 6 to 8%. Thermal barrier coatings are usually deposited on metal surfaces by plasma spraying or else by PVD (Physical Vapor Deposition). Other techniques, e.g. plasma assisted PVD have also been tested for this application. The thermal barrier coating of zirconia and yttria does not make direct contact with the underlying metal, instead an 'Oxidation Resistance Bond Coat' (ORBC) is interposed. The ORBC is required because the zirconia coating would otherwise readily conduct oxygen to permit oxidization of the substrate metal. A non-ferrous alloy of nickel, chromium, aluminum and yttrium or an aluminide are used as the ORBC and in some advanced forms of coating, a graded composition between the substrate and thermal barrier coating is developed. The thickness of the ORBC is 100 μm or more.

A basic problem of plasma sprayed coatings is porosity, which allows passage of oxygen and causes brittleness. Thermal barrier coatings usually fail by cracking and spallation after a large number of thermal cycles (heating and cooling). Cracking can occur within the coating or between the coating and ORBC due to differences in thermal expansion or oxidation of the ORBC. Post-treatment of plasma sprayed coatings with concentrated sources of heat such electron beams is currently being studied to try and eliminate the porosity inherent in plasma spraying.

6.4.2 Plasma transferred arc and the D-gun

Thermal and plasma spraying are designed not to melt the substrate: Where melting of the substrate is desired in order to achieve a higher adhesion between the substrate and coating, the Transferred Plasma-Arc process is used. Transferred Plasma-Arc (PTA) more closely resembles welding than plasma spraying since the PTA maintains closely proximity to the coating surface unlike plasma-spraying where a separation of approximately 100 millimeters is typical. The plasma-arc of PTA does not remain within the coating unit, as is the case for plasma spraying but instead extends from the coating unit to the substrate. The plasma-arc between anode and cathode inside the coating remains important for melting the coating material. Electric charge is transferred from the coating unit to the substrate by the plasma-arc and thus provides the origin of the name of the process. The metal coating powder is heated until it vaporizes and forms a plasma, which is then conveyed by the flow of gas to the substrate. A very rapid condensation of metal plasma to liquid occurs once the metal plasma reaches the substrate. A melt-pool of coating metal forms on the surface of the substrate and this melt-pool later solidifies to form a solid coating. PTA is illustrated schematically in Figure 6.7.

A pool of molten metal is formed at the region where the plasma-arc reaches the substrate; coating material is fed as molten droplets into this molten pool. PTA resembles welding more closely than plasma spraying and while bonding between coating and substrate is usually excellent with PTA some dilution of the coating by substrate is inevitable. In applications, where such dilution is intended, e.g. when producing an alloyed surface layer, the powder feed rate can be adjusted to obtain the desired alloy composition.

Advantages of PTA compared to plasma spraying include less stringent restrictions on the average and range of powder size usable and less electrical power consumption for a given volume of coating. Coatings up to 6 mm thickness can be deposited in one pass of the coating unit, the width of material covered in each pass is only limited by the diameter of the coating unit but in practice is limited to 30 mm width. Coating deposition rates are typically 10 kg/hour, which is very high, compared to any coating process. A prime disadvantage of PTA is that only an electrically conductive substrate can be coated which excludes ceramic coatings. Thermal distortion and extent of the HAZ (heat affected zone) are also greater for PTA than plasma spraying.

The D-gun or Detonation-gun is based on a cyclic detonation of combustible gas in a chamber with an exit directed at the substrate. During detonation, the substrate is blasted with a hot gas containing molten droplets of coating. The D-gun process is illustrated schematically in Figure 6.8.

The D-gun employs the same principle as plasma spraying where coating material is delivered to the substrate as molten droplets moving at high speed. The characteristics of combustion, i.e. a maximum adiabatic flame temperature and short duration of the maximum flame temperature impose more stringent limitations on the range of materials that can be deposited than is the case with plasma spraying. Hard refractory materials are unsuitable for coating with a D-gun as opposed to a plasma spray because the lower temperatures of the D-gun and higher impingement velocity

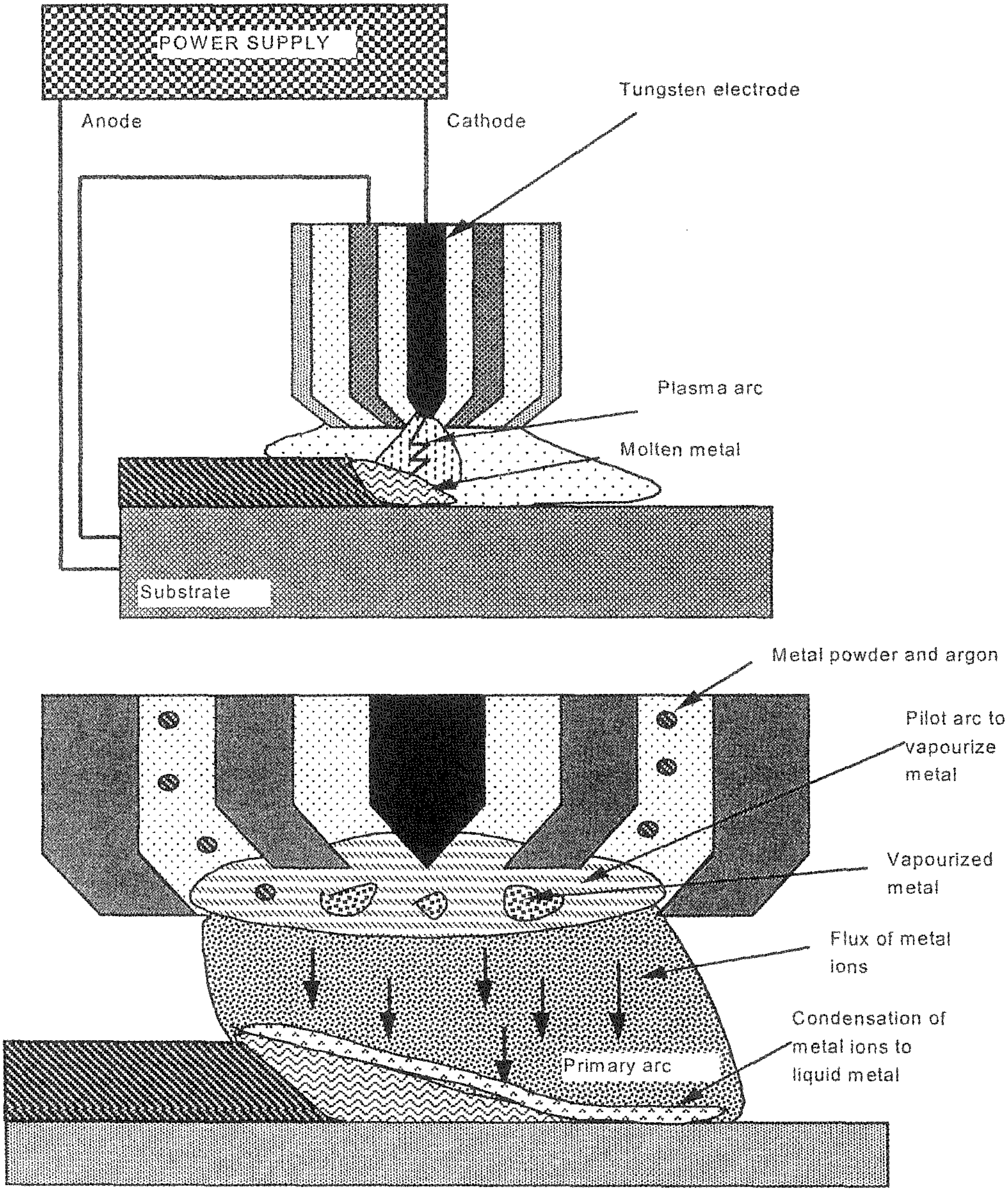


Figure 6.7 Operating system and mechanism of the Plasma Transferred Arc coating process.

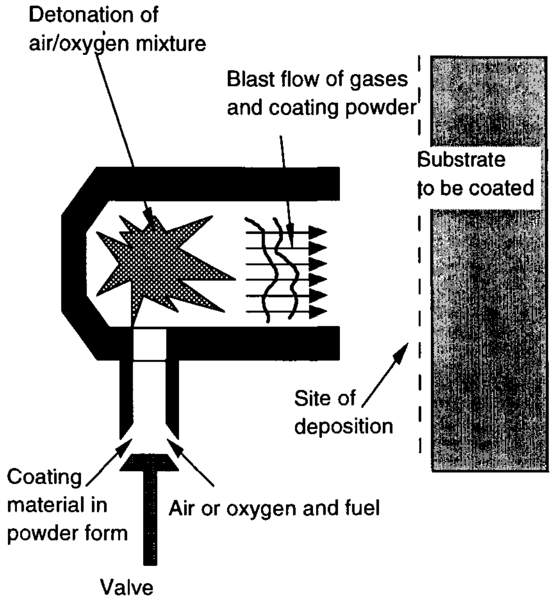


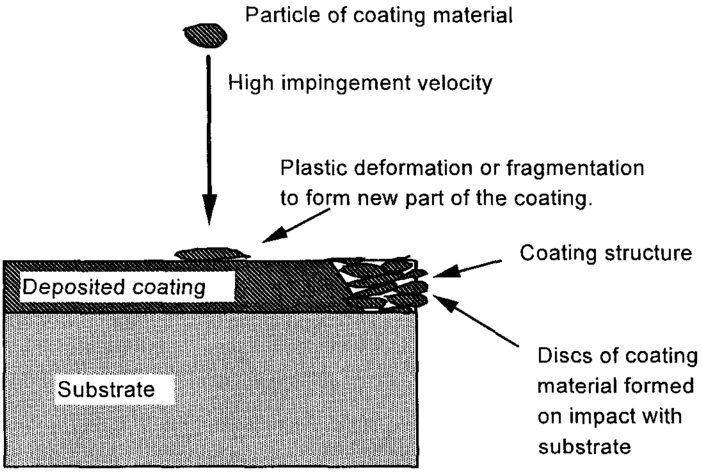
Figure 6.8. Operating mechanism of the detonation-gun coating system.

(of the coating powder against the substrate) ensure that hard particles can erode the substrate. The particles of coating must deform to a lamellar shape in order to form a coherent strong coating and this requirement effectively precludes many ceramics as coating materials. Substrate erosion when using the D-gun is illustrated schematically in Figure 6.9.

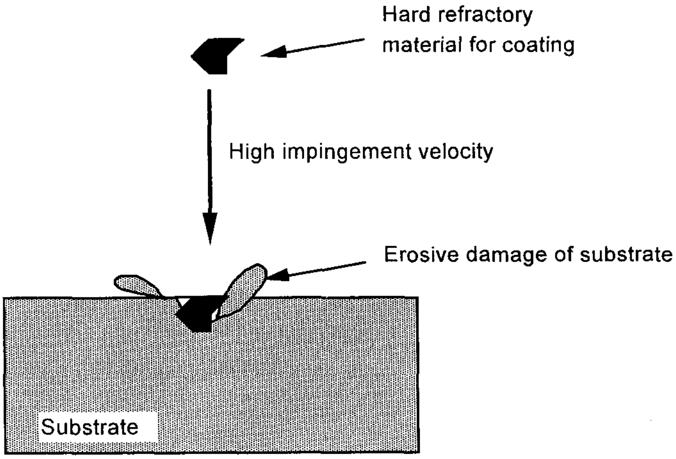
The main advantage of the D-gun is a much lower equipment cost since an expensive electrical power system is not required as is for plasma spraying.

6.5 Vacuum-based coating methods

The terrestrial atmosphere, which is so essential for human life, greatly restricts coating processes. The atmosphere contains approximately 20% oxygen and 80% nitrogen, both these gases can react with metals especially when the metals are in particulate form and at a high temperature. Most coating processes depend on the



EFFECTIVE COATING FORMATION



EROSION OF SUBSTRATE WHEN COATING MATERIAL IS HARD AND REFRACTORY

Figure 6.9. Mechanism of coating deposition from a D-gun and substrate erosion when depositing a hard material using a D-gun.

transfer of coating metal from a source to the substrate where the coating metal is transferred either as a vapor or plasma or as a spray of liquid metal particles. If these coating processes are performed in the open atmosphere, oxygen and nitrogen will react with the coating metal to form oxides and nitrides unless a perfectly inert coating metal is used. It is found that if oxygen and nitrogen are excluded from a coating site by using vacuum technology, then a wide range of high quality coatings can be produced. Developments in vacuum pumps and vacuum chambers have converted vacuum-based coating methods from laboratory curiosities to valuable industrial processes. Vacuum chambers with a volume of many cubic metres can be rapidly pumped to pressures of less than 0.1 Pa or less, which is adequate for most coating processes. A large number of components can be simultaneously coated in these large vacuum chambers thus allowing cheap mass-production of coated components.

Vacuum-based coating methods are commonly classified into two basic categories, which are Physical Vapour Deposition (PVD) and Chemical Vapour Deposition (CVD). PVD which comprises of most coating methods apart from CVD, involves the direct transfer of coating material as a vapour or plasma from the coating source to the substrate. For CVD, chemical reactions at the surface of the substrate are a necessary part of the coating process. PVD and CVD coating technologies are described below.

6.5.1 Vacuum deposition

Vacuum deposition is one of the simplest and oldest methods of coating in a vacuum. The coating metal is electrically heated in a crucible until it either melts or otherwise begins to release large quantities of vapor, i.e. sublime. This vapor then spreads to fill the entire vacuum chamber and subsequently be deposited on the component to be coated and the chamber walls. Metal coatings are deposited in this manner with no significant thermal damage to the substrate. The principle of vacuum deposition is illustrated schematically in Figure 6.10.

The quality of coating produced by vacuum deposition is inferior to other more modern coating methods. The primary limitation is that vacuum deposition does not expel contaminants before deposition, so that coating adhesion is usually poor. The energy level of the metal vapor is also too low to permit the formation of a strong microstructure in the coating.

6.5.2 Sputtering

Sputtering and related coating methods discussed below involve the direct application of electrical energy to enhance the coating quality. It was discovered in the 19th century that if a low pressure of gas is maintained in a vacuum chamber, electricity can be conducted by charged ions of gas between a cathode and anode. The ionization of gas involves the release of energy as light and other electromagnetic radiation, which means that the gas glows. Neon lights depend on this phenomenon for their function. The optimum gas pressure for electrical conduction is much smaller than atmospheric pressure to ensure that charged ions are not impeded by a larger number of uncharged gas molecules. There is an optimum gas pressure where the minimum voltage for electrical conduction is less than that required for electrical discharge both in a perfect vacuum and in open air. Conduction of electricity in the open air only occurs at much higher voltages than at lower gas pressures where a high-energy arc of ionized gas forces a path through the surrounding uncharged gas molecules. This high-energy arc is difficult to control and is not useful for vacuum-based coating methods. Mattox and other investigators later found that if an inert gas such as argon is selected, the plasma of charged argon ions can be used as a high-energy transfer medium for the coating material.

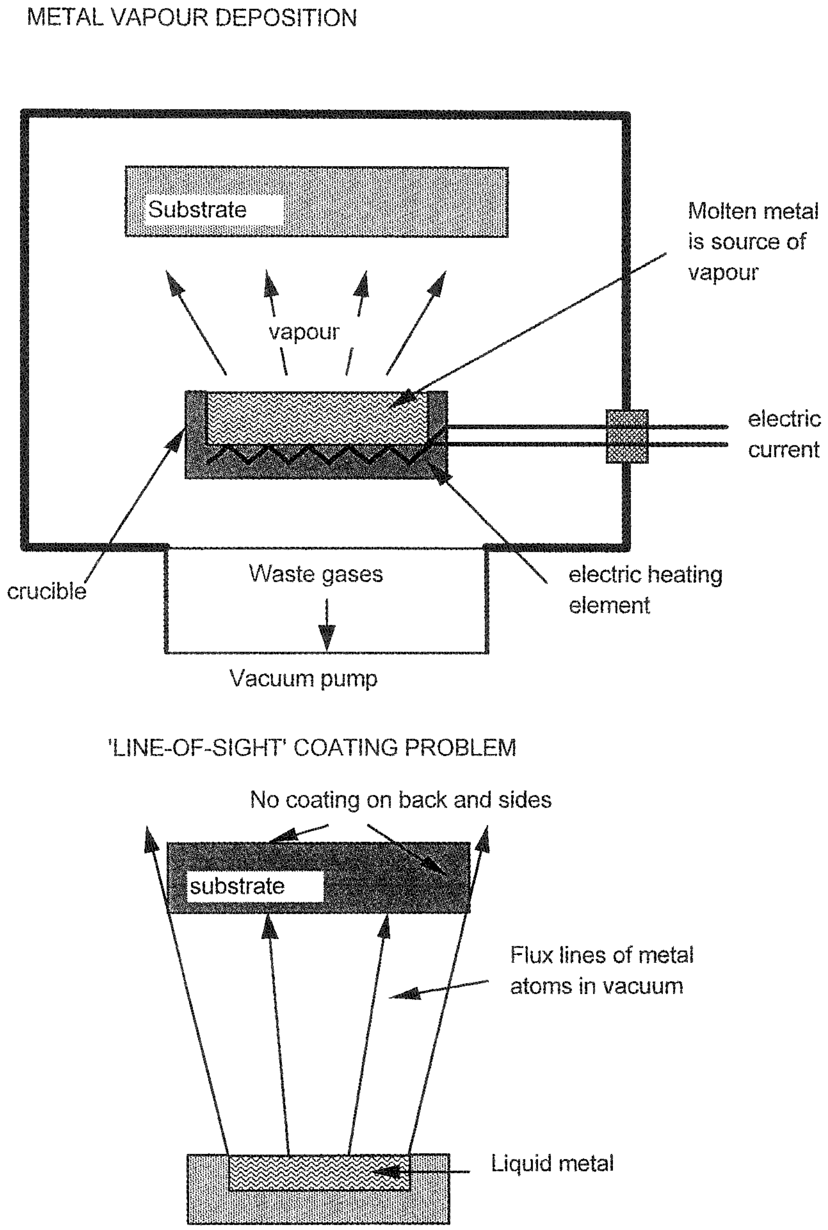


Figure 6.10. Principle of vapour deposition.

Sputtering is a term used to describe the removal of atoms from a solid by impingement of high-energy ions found in a gaseous plasma. The gaseous plasma is generated between a cathode and anode as described above and can dislodge atoms from both the substrate and a source of coating material. Dislodgement of atoms from the substrate is used to clean the substrate of contaminants such as greases and oxides. In most coating operations, sputtering of the substrate is performed before a coating is deposited to ensure cleanliness of the substrate. Once coating is initiated, atoms of coating material dislodged from the substrate diffuse across the space between substrate and coating source to form a coating on the surface of the substrate. A fundamental advantage of sputtering is that no heating of the coating source is required so that any material can be coated. The energy levels of sputtering are comparatively low which prevents excessive heating of the substrate. It is possible to coat thermolabile (heat-sensitive) polymers on to paper by sputtering.

Sputter coating or sputtering as it is known, is illustrated schematically in Figure 6.11. In Figure 6.11, an anode, a cathodic coating source and substrate are shown enclosed in vacuum chamber filled with low-pressure inert gas, which is usually argon.

Sputtering is in precise terms a form of Physical Vapor Deposition (PVD) but is usually called sputtering by practitioners. Sputtering is used to produce reflective gold coatings on polymer structures to form lightweight optical mirrors. Other applications are the deposition of corrosion and wear resistant coatings or layers of metal and insulant in an electronic 'chip'. Sputtering has become an important industrial coating process that often has no substitute.

The rate of deposition of sputtered coatings is approximately 1 μm of coating thickness per hour, which precludes coatings thicker than 10 μm for practical purposes. Most sputtered coatings range from 0.1 μm to 3μm thickness though there are exceptions to this rule. The microstructure of sputtered coatings is critical to their function and there is considerable variation in microstructure with impingement energy of coating material. If there is no extra voltage imposed between the anode and cathodic substrate, then the impingement energy of coating material is small because there is no additional electric potential available to accelerate charged coating material. When a voltage is imposed between the anode and the substrate, the impingement energy will increase. This voltage is usually referred to as 'bias'. If bias is not used, then most coatings do not form a continuous layer but are instead composed of pyramids or cones that cover the surface of the substrate. It is only when bias is used that the coating is more likely to form a layer of closely packed crystals. This distinction in coating structure is shown schematically in Figure 6.12. Heating of the substrate also prevents pyramids and cones from forming but defeats the purpose of sputtering, which is to coat at low temperatures.

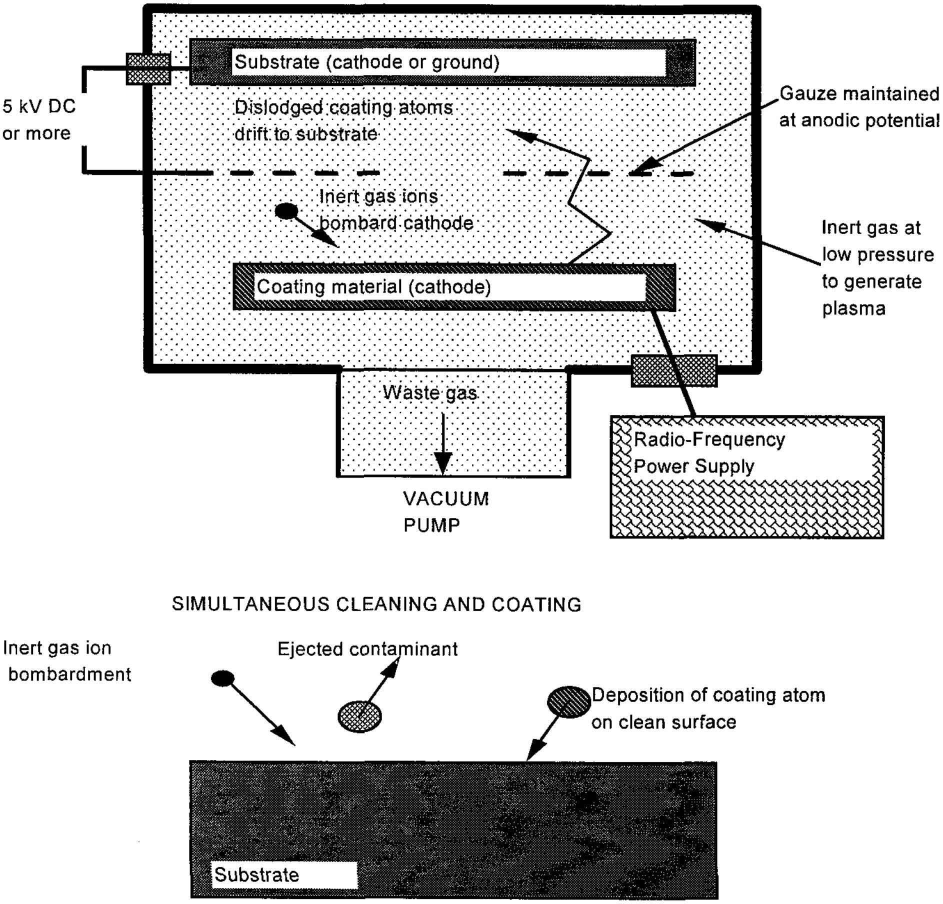


Figure 6.11 The principle of sputter coating or sputtering.

The pyramid or cone structure is formed because of the lack of atomic mobility of coating material at low impingement energies. As can be readily appreciated, the pyramid/cone structure is mechanically very weak and is rapidly removed by wear. The

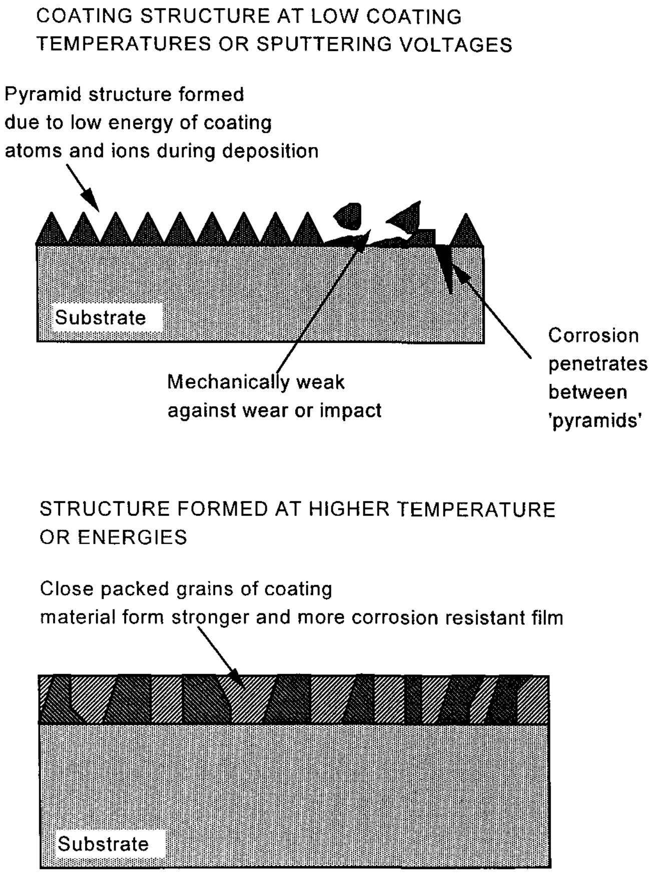


Figure 6.12 Strong and weak microstructures of sputtered coatings.

boundaries between pyramids/cones contain almost no coating; this facilitates access to the substrate by corrosive agents. Most commercial coating processes are adjusted to avoid producing coatings with a pyramid/cone structure. The disadvantage of bias or high impingement energy is greater process heating of the substrate which can be heated by 100°C or more, this temperature rise is however much less than the temperature rise during direct heating of the substrate to prevent the pyramid/cone structure.

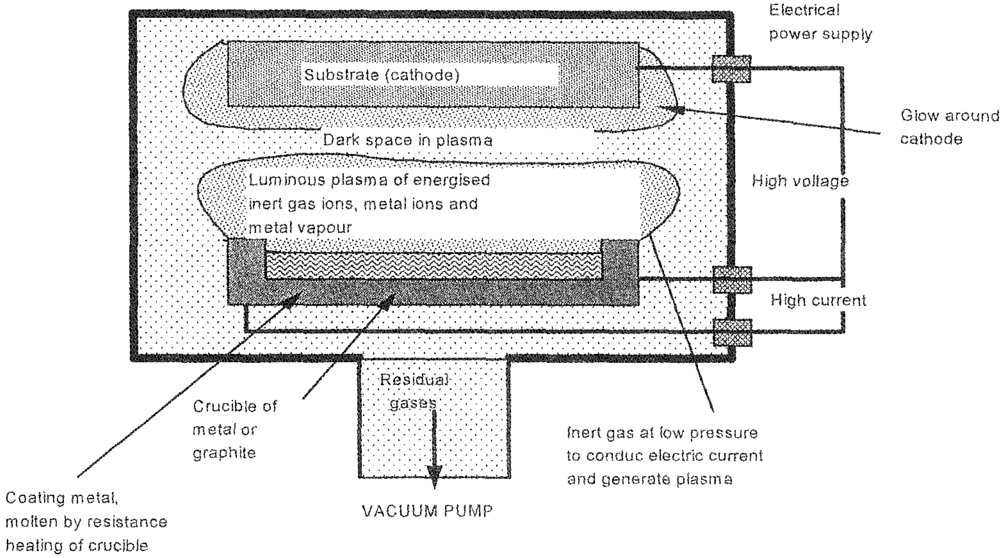
Under carefully controlled conditions of Direct Current Sputtering using a Magnetron, amorphous metal coatings can be formed. A magnetron is a circular ring-shaped magnet that is located between coating source and coating substrate. It has the function of trapping electrons which otherwise heat the substrate. Amorphous chromium-zirconium coatings 1 μm thick can be deposited on a metal surface to provide protection against very strong acids such as, 6M hydrochloric acid. Amorphous forms of a metal are typically much more resistant to corrosion than the corresponding crystalline form of the metal.

6.5.3 Ion plating and ion-beam assisted deposition

Coating quality is largely controlled by the level of electrical energy applied to the coating process. In simple terms, the more energy the coating atoms possess while in transit between the coating source and substrate, the stronger the adhesion of the coating to the substrate and the more robust its microstructure. High electrical energy also has disadvantages such as process heating of the substrate but these can be controlled to produce a higher quality of coating than otherwise attainable. In vacuum-based coating, electrical energy can either be given directly to the plasma between coating and substrate or a second source of energy can be added to the system. The direct conferral of high energy to the ions forms the basis of ion plating while ion-beam assisted deposition involves a second source of electrical energy.

Ion plating is closely related to sputtering and is often performed using a common vacuum apparatus. The fundamental differences between sputtering and ion plating are the use of electrical heating to melt the source of coating material and a much higher electric potential between the coating and substrate. Ion plating is also restricted to metals since most inorganic compounds decompose during vaporization or are too refractory. Ion plating is formally known as Plasma-enhanced Physical Vapor Deposition (PEPVD) as the operating principle is the application of energized plasma to improve the coating characteristics of vapor deposition. The operating principle of ion plating is illustrated schematically in Figure 6.13 where a vacuum system, electrically heated source of coating metal and substrate are shown.

ION-PLATING APPARATUS



ION-PLATING MECHANISM

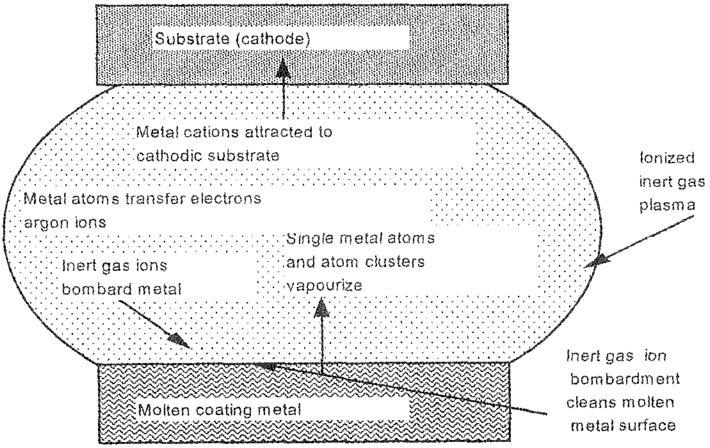


Figure 6.13 Principle of ion plating.

A major disadvantage of ion plating is the limited range of materials that can be selected as either substrate or coating since it is only possible to coat metal-on-metal by this technique.

Ion-beam assisted deposition (IBAD) represents an attempt to combine the flexibility of a low energy coating process such as sputtering with the improved coating microstructure of higher energy coating processes. IBAD involves the sputtering or vapor deposition of a coating while simultaneously bombarding the coating with high-energy ions. IBAD is illustrated schematically in Figure 6.14. The purpose of bombardment with high-energy ions of usually nitrogen at typically 100 kilo-electron-volts is the force the coating to mix with the substrate and develop a strong microstructure. The nitrogen ions may react with coating material or the substrate to form nitrides. Metal nitrides are mostly very hard substances that can increase the wear resistance of surface and so are considered desirable additions to a coating. If the coating material mixes with the substrate during the beginning of coating then a sharp interface between coating and substrate cannot form; instead a graded interface will be formed. A sharp interface between the substrate and the coating sustains a high level of stress due to epitaxial strains between the differing crystalline structures of the substrate and coating (epitaxy is discussed in Chapter 3). The sharp interface usually contains numerous defects such as voids or trapped contaminants, which form sites for crack initiation. When a coating is subjected to an external shear stress by sliding friction, fracture usually occurs at the coating-substrate interface because of its mechanical weakness. Ion plating also produces graded interface coatings though by a different means as discussed above. The distinction between a sharp interface and graded interface is illustrated schematically in Figure 6.15 for coatings produced by ion plating.

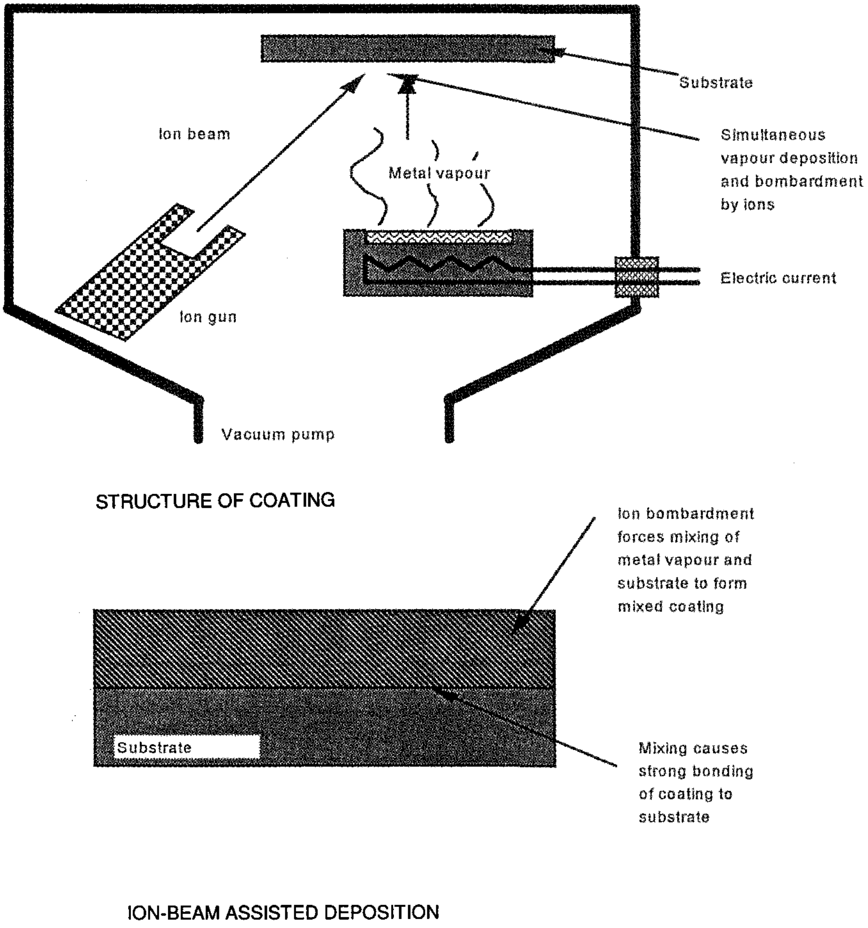


Figure 6.14 Mechanism of Ion beam assisted deposition (IBAD) .

The intermediate layer can be thought of as having a concentration ranging from 100% coating at its outer boundary and 100% substrate at its inner boundary with intermediate values inside the layer. For sputtered coatings and vapor deposited coatings the concentration of coating material declines precipitately from 100% to 0% across an interface that has a nominal thickness equal to the size of one atom. A graded interface is thought to confer substantial mechanical strength to the IBAD coating or ion-plated coating for the following reasons. Contaminants are usually removed by the high-energy bombardment of ions and if traces of contaminants remain, these do not cause a major loss of mechanical strength. Epitaxial strains are accommodated over a finite depth of material instead of being localized at an interface. The inter-metallic compound formed between coating and substrate is often very strong and can therefore withstand the interfacial stresses between coating and substrate.

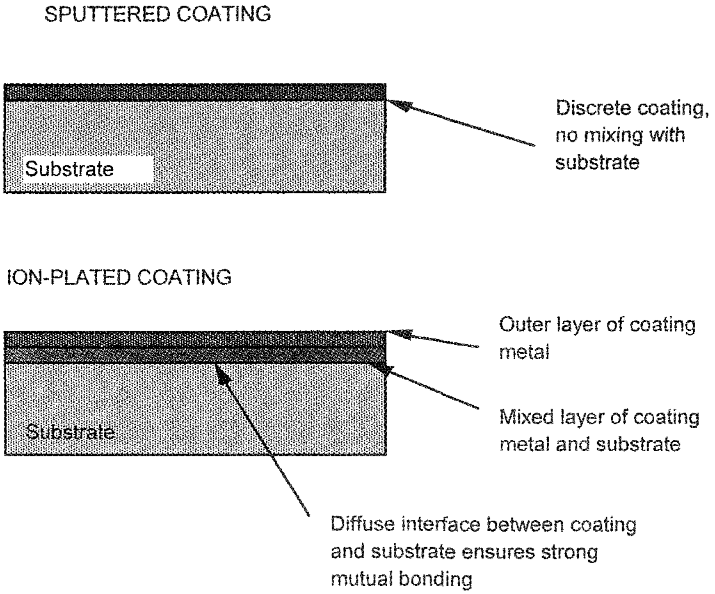


Figure 6.15 Formation of intermediate layer between substrate and coating by ion plating.

It was found during wear tests on steel wires, that ion-plated layers of chromium and other metals survived long periods of wear whereas sputtered layers of the same metals were quickly removed by interfacial fracture of the coating.

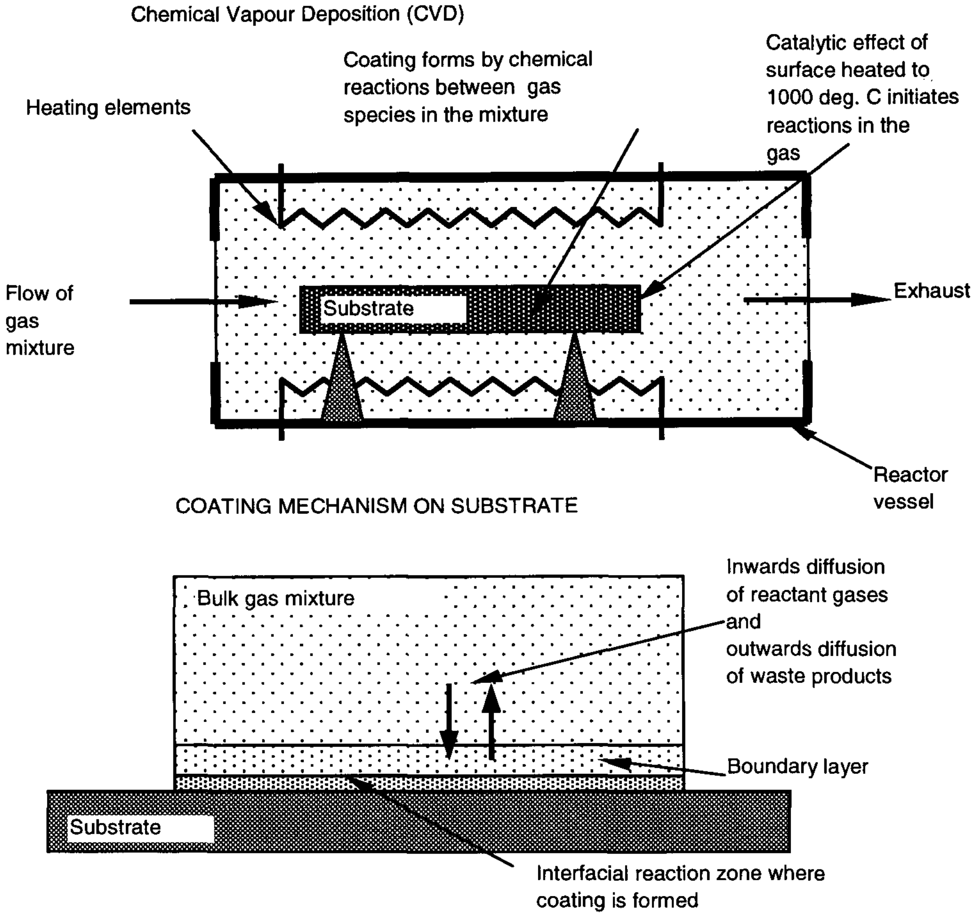
The level of process heating in ion plating can be high enough to completely alter the microstructure of the coated component. For example, during attempts to ion-plate steel wires with chromium and zirconium for wear resistance, the wires were observed to glow red-hot. The hardness of the wires declined from an initial value of 490 Vickers to 230 Vickers after ion plating. The original work hardened microstructure of the wire had been destroyed by process heating. Similar problems may occur with IBAD.

In most applications of ion plating an inert gas such as argon is used to generate a plasma, but it is also possible to use a reactive gas such as nitrogen and so deposit nitrides instead of a metal coating. This process is known as reactive ion plating and provides a useful extension to the scope of ion plating.

6.5.4 Chemical Vapour Deposition (CVD) and Plasma Enhanced CVD

Chemical Vapor Deposition (CVD) and its derivative process, Plasma Enhanced Chemical Vapor Deposition (PECVD) utilize the catalytic effect of a material during chemical reactions in a gas. The surface of a solid provides a fixed site for several molecules to be adsorbed and then react whereas in the gas, only momentary collisions between two molecules are statistically probable. If the reaction product is a solid then a coating can be formed on the surface of the material. CVD and PECVD involve the preparation of a chemically unstable mixture of gases, which are heated to high temperature and then passed over the component to be coated. A coating is then formed by reaction between the various components of the gas mixture; the substrate is usually never consumed or converted during these chemical reactions. A disadvantage of the CVD process is that it functions at temperatures ranging from 800°C to 2000°C which causes many materials to either melt or sustain microstructural damage, e.g. grain growth or phase decomposition. The critical advantage of PECVD is the ability to coat at much lower temperatures of less than 500°C. PECVD coatings, but not CVD coatings can be deposited on steel components with an acceptable level of microstructure damage by process heating.

The CVD process is illustrated schematically in Figure 6.16, which shows a vacuum chamber containing the coating gas mixture, which is passed over the heated substrate. Coating formation occurs inside a boundary layer of reacting gases, which covers the substrate.



EXAMPLES OF FILM FORMATION REACTIONS

(i) Thermal decomposition (pyrolysis)

$CH_3SiCl_3 \dashrightarrow SiC + 3\,HCl$

(ii) Reduction

$WF_6 + 3H_2 \dashrightarrow W + 6\,HF$

(iii) Oxidation

$SiH_4 + O_2 \dashrightarrow SiO_2 + 2H_2$

(iv) Hydolysis

$2AlCl_3 + 3H_2O \dashrightarrow Al_2O_3 + 6\,HCl$

(v) Co Reduction

$TiCl_4 + 2BCl_3 + 5H_2 \dashrightarrow TiB_2 + 10\,HCl$

Figure 6.16. Operating principle of Chemical Vapor Deposition (CVD).

PECVD can be performed in a modified CVD apparatus, which includes a power supply, cathode and anode to generate plasma. Plasma is generated between two parallel cathodes and anodes, coating occurs inside this plasma where the substrate is located. Heaters are necessary for the PECVD process but the required temperature is much lower so the technical sophistication of the heating system and chamber insulation is lesser than for CVD. The PECVD process is illustrated schematically in Figure 6.17.

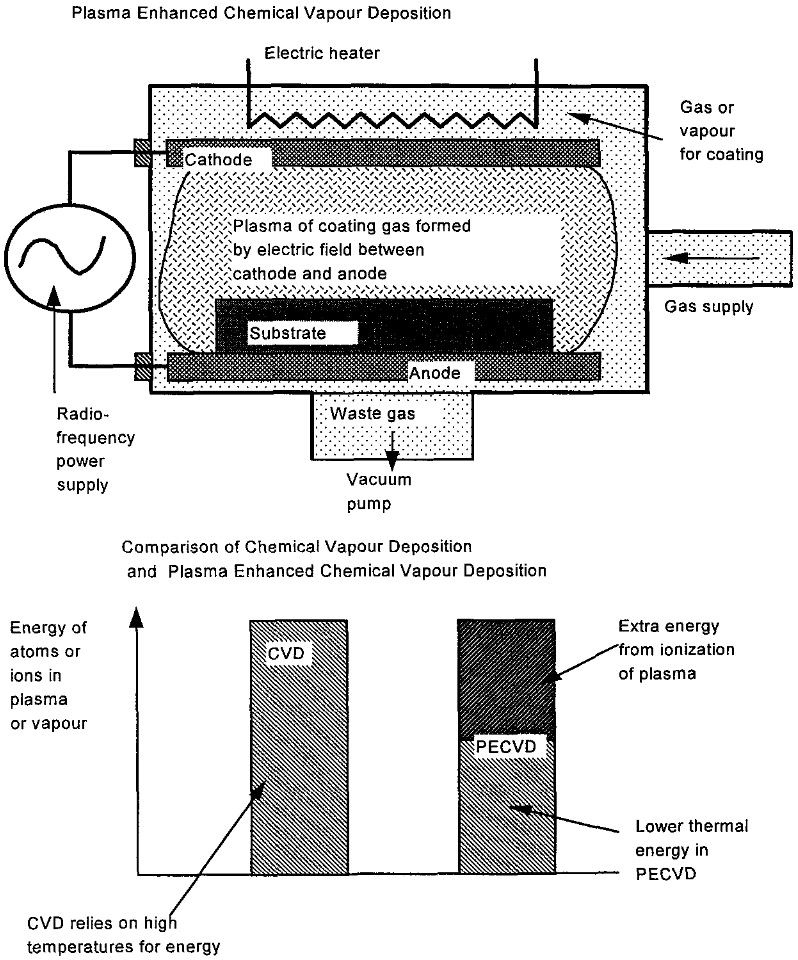


Figure 6.17 The operating principle of Plasma Enhanced Chemical Vapor Deposition (PECVD).

A significant advantage of CVD and PECVD over PVD is a much higher deposition rate, which allows thick coatings to be developed in a reasonable time. As discussed above, PVD coatings are effectively limited to a few micrometers which is unacceptably thin for many applications, e.g. abrasive wear and high temperature corrosion. It is also possible to generate components by CVD where a material is not amenable to conventional techniques such as casting or rolling. For example, tungsten tubes can be formed by deposition from tungsten hexafluoride on to a cylindrical former, which is later removed. The tube thickness is approximately 1 mm, which is 100 times larger than the maximum thickness that can be reasonably deposited by PVD, i.e. 10 μm. PVD film thickness greater than 10 μm require impracticably long coating times of more than 1 working day

PVD and CVD can be used on the same substrate to produce a multi-layered coating where CVD is often used to produce the thicker coating or coatings. Multi-layered coatings are becoming more widely adopted with improvements in the handling technology to transfer substrates from different coating units, from e.g. a CVD unit to a PVD unit inside the same vacuum system. Multi-layered coatings based entirely on PVD coatings are also possible.

6.5.5 Arc Evaporation

Arc Evaporation shares a common principle with Vacuum Evaporation that coating metal is transferred from source to substrate as a vapor cloud. The difference between Arc Evaporation and Vacuum Evaporation is that electrical discharge, i.e. an arc, is used to evaporate the metal instead of using heat. A major advantage of Arc Evaporation is that the electrical discharge not only heats the coating metal but also ionizes the metal vapor produced. When the coating metal is vapor is ionized and therefore converted to plasma, it is found that the quality of coating is greatly improved. A dense uniform coating is produced without the undesirable columnar structure that is a problem with sputtering (discussed above). It is believed that the higher energy of metal ions as opposed to vaporized atoms momentarily confers sufficient mobility to coating atoms so that the crystal structure of the coating can grow laterally to cover the whole substrate. A lack of lateral crystal growth predisposes a coating material to produce a columnar structure. An arc evaporation system is shown schematically in Figure 6.18.

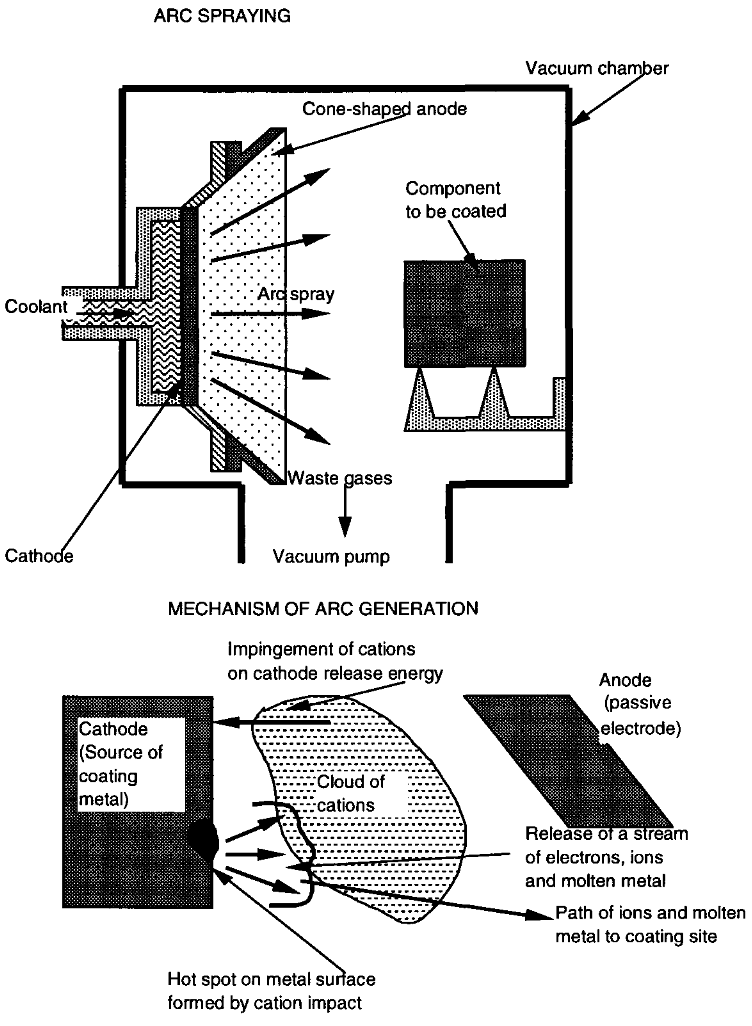


Figure 6.18 Principle of operation of an arc evaporation system.

An electrical discharge is obtained by imposing a cathodic polarity on the coating material, which is in the form of a disc while an anode is built into the surrounding apparatus. Careful cooling by water jackets of both the anode and cathode (coating material) is necessary to prevent bulk melting of either. The substrate is presented in front of the cathode to receive the ionized flux of metal ions. A major problem with arc evaporation that has still not been effectively controlled is the release of droplets of liquid coating metal. These droplets may only be a few micrometers in diameter but they are sufficiently large to form large protuberances on the substrate surface when they cool and solidify. A coating should ideally be smooth and of uniform thickness so that these protuberances degrade the coating. This problem is illustrated schematically in Figure 6.19.

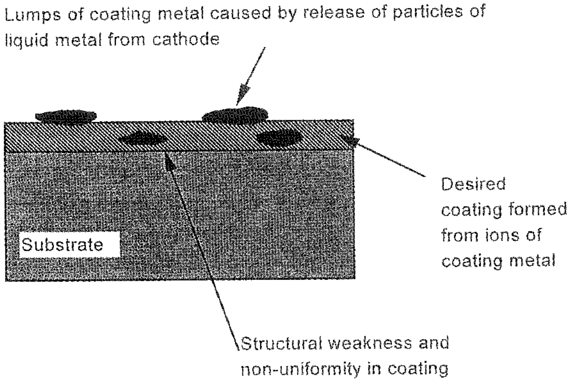


Figure 6.19 Degradation of arc evaporation coatings by uncontrolled formation of lumps inside the coating.

It is believed that non-uniform electric discharge causes what are termed 'active spots' to form on the coating source. These 'active spots' are areas of abnormally hot molten metal. It is believed that recoil or bouncing of electrical arc from one part of the coating source to another can create regions of abnormally intense electron bombardment on the surface of the coating source. When the intensity of electron bombardment exceeds a critical level, particles of molten metal are ejected from the coating source as well as individual atoms and ions.

The critical advantage of Arc Evaporation over Physical Vapor Deposition (PVD, i.e. sputtering or ion-plating) is a much higher coating rate combined with a tendency to produce dense coatings. Many engineering applications of coatings, such as wear or corrosion resistance require a coating thickness of several micrometers. Many hours are required by PVD to reach such a thickness and so PVD is not normally used to produce coatings thicker than approximately 1 µm. Arc Evaporation has the economic advantage of faster component processing since coatings of more than 1 μm thickness can be obtained in less than 10 minutes of processing. Coating quality is lower than PVD coatings for the reasons described above and process equipment requires careful shielding to prevent leakage of electromagnetic radiation to the external environment. Emission of intense light and ultra-violet rays (which are a form of electromagnetic radiation) and disruption to radio transmission by electrical noise from the arc are two major problems that have to be addressed in the design or installation of Arc Evaporation systems.

6.6 Friction surfacing, weld overlays and explosive bonding

Friction surfacing, weld overlays and explosive bonding share the characteristic of being suitable for the rapid deposition of thick metallic layers on metals. In theory it may also be possible to apply these techniques to other hard refractory substances, e.g. ceramics but this does not appear to have been investigated yet. The processes involved in these three coating methods are however distinct and are suitable for different applications.

Friction surfacing

Friction surfacing is an adaptation of friction welding where the consumable does not stop rotating nor remain in the same location on the substrate. Instead after the initial heating period, the consumable continues to rotate while being simultaneously drawn across the substrate to deposit a thick layer of metal on the substrate. Friction surfacing can be considered as deliberately engineered severe wear in which relative deposition between the two contacting bodies is only possible by the continuous deposition of a massive transfer layer. The process of friction surfacing is illustrated schematically in Figure 6.20.

During coating, the bulk of the consumable is separated from the substrate by an intermediate layer of hot and continuously deforming consumable. This intermediate layer appears to undergo a cyclic process of formation and sudden release to the substrate to create a coating, which resembles a stack of cards that is tipped over on side. Adhesion between these lamella or 'cards' is however high thus allowing the formation of a coherent strong

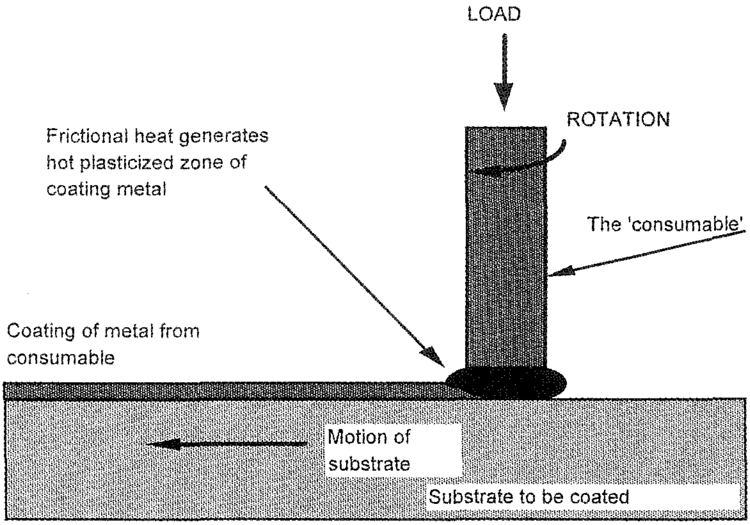


Figure 6.20 Mechanism of friction surfacing.

coating. The microstructure of the coating is characteristic of hot forging above the recrystallization temperature of the material and is therefore considered superior to weld overlays with their as-cast microstructure and solidification cracks. Friction surfacing occurs at a lower temperature than fusion welding; because of this lower temperature, there is virtually no Heat Affected Zone beneath the coating. Friction surfacing produces coatings that are typically 1 or 2 mm thick and thicker coatings can be produced by multi-layer surfacing, i.e. repeated friction surfacing over the same area.

Since friction surfacing is effectively a solid-state welding process, metals, which are non-wetting or insoluble in the liquid state, can be joined together. Aluminum coatings on steel or vice versa, steel coatings on aluminum are a basic example of this aspect of the process. Not all metal combinations are however possible as high thermal conductivity and resistance to adhesive wear suppress friction surfacing. For this reason, brass is difficult to deposit by friction surfacing and aluminum can only be deposited by applying a very high stress on the consumable. This results in a rapid downwards-vertical movement on the consumable, thus preventing the upward movement of heat in spite of the high thermal conductivity of aluminum. The thermal conductivity of brass is also high and it has good resistance to adhesive wear so that it is unsuitable for friction surfacing. For reasons not perfectly understood, some metals do not form a substantial intermediate layer between the consumable and substrate. The contacting end of the consumable therefore becomes plastically deformed to form a sleeve surrounding the unused consumable. A further problem that is common to all material combinations is the porosity in the boundaries between strips of coating formed by successive passes of the consumable. Porosity can be suppressed but not eliminated by ensuring overlap between adjacent passes of the consumable on the substrate. These problems with friction surfacing are illustrated schematically in Figure 6.21.

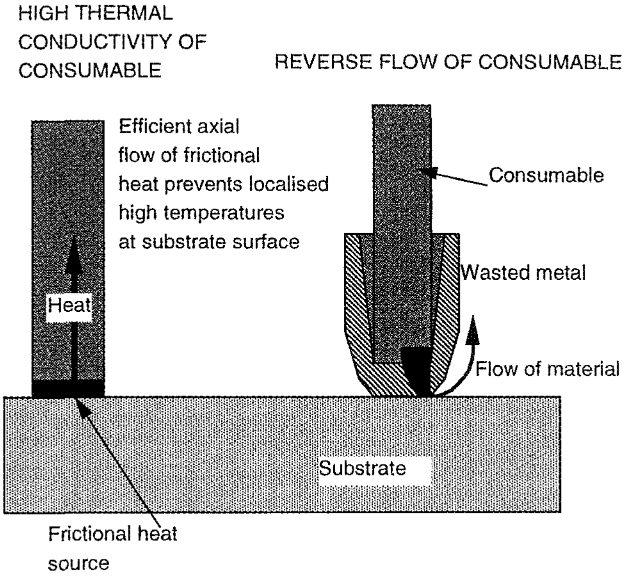


Figure 6.21 Problems associated with friction surfacing: thermal conductivity, wear resistance and consumable return.

Friction surfacing is also limited to planar surfaces or surfaces with uniform curvature, e.g. cylinders as more complex surface geometry hinders physical contact between consumable and substrate. It is however possible to perform friction surfacing on an inclined surface, provided that the consumable moves across the slope (not up and down).

It is probable that many of these problems with friction surfacing will remain for the foreseeable future as most research appears to be directed to optimizing and commercializing proven material combinations such as the deposition of hard metals on steel.

Weld Overlays

Weld overlays represent a simple extension from joining by welds as the same technology can be used. A coating is formed by depositing weld material onto a surface in a series of adjacent contacting strips. Thick coatings can be formed by successive overlays of welds. The coating process can be performed manually or by automatically by a robotic welding system for mass production. Very thick coatings can be produced but with the limitations of a possibly brittle cast microstructure and a deep Heat Affected Zone. Operator discomfort for manually produced weld overlays is also considerable as long periods of exposure to welding vapors and molten metal splatter are entailed.

Explosive Cladding

Explosive cladding is possibly the most direct form of deposition in that the barriers to adhesion between coating and substrate are overwhelmed by the application of a large amount of mechanical energy. A flat sheet of cladding material is placed on top of the slab of material to be clad. A wedge is placed between the slabs and cladding material at one end of the slab while a quantity of explosive is located at the other end of the slab. Cladding is performed by igniting the explosive to abruptly force the sheet into adhesive contact with the slab. High-speed camera recordings reveal that during the blast of the explosive, a localized linear zone of flexion moves from the site of the explosive towards the wedge. Adhesive contact between cladding sheet and substrate is developed by movement of a front situated at the line where the cladding sheet is forced to flex. For the brief moment of the explosion, there is sufficient energy for metal and oxides to be molten at the front where cladding and substrate first make contact. It is believed that a gaseous plasma forms at this front to exclude atmospheric oxygen and allow adhesive bonding between the cladding and substrate. Extensive plastic deformation often causes the interface between bonded cladding and substrate to be rippled or assume a saw-tooth profile. Abnormal phases of materials have been observed at the coating-substrate interface due to the severe plastic deformation and sudden heating that occurs during cladding.

The method of explosive cladding and the mechanism of adhesion between cladding and sheet are illustrated below in Figure 6.22.

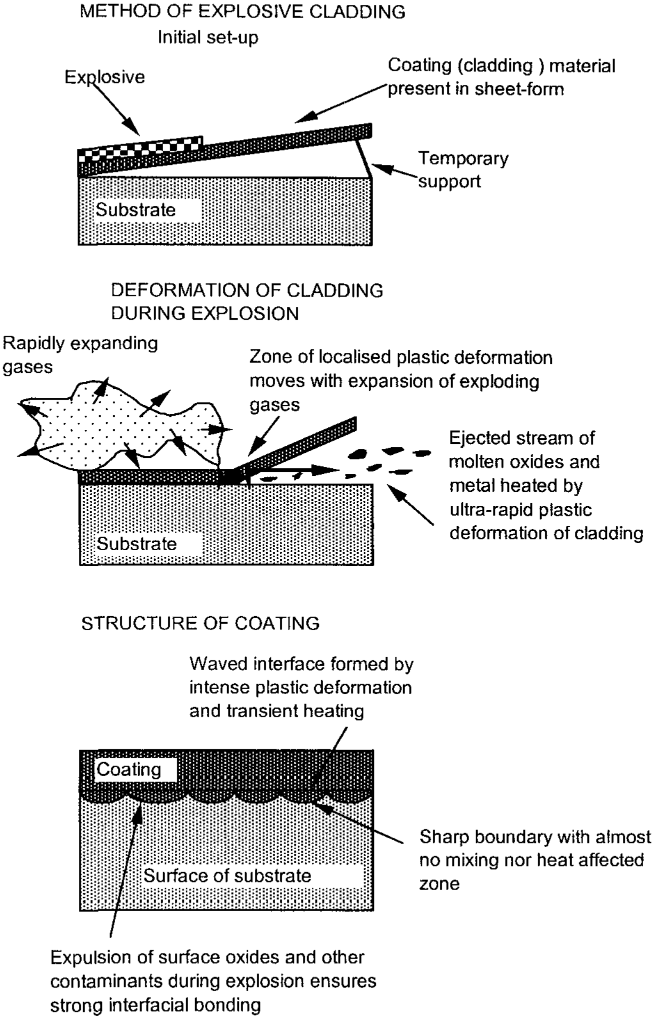


Figure 6.22. The method of explosive cladding and mechanism of bonding between coating and substrate.

Similar metals, e.g. aluminum and aluminum alloy or dissimilar metals can be coated by this method which is most suitable for the rapid deposition of large areas, e.g., 1 square metre, with several mm thickness of coating. It is possible to deposit metallic coatings where plasma or thermal spraying only form porous weak layers and where weld overlays entail the formation of brittle intermetallic compounds between coating material and substrate. A recent example is the deposition of 1 mm thick sheets of NiTi on structural steel for the manufacture of turbine impellers with resistance to cavitational wear. In this application, the lack of fusion during coating ensured that the required stoichiometry of NiTi was maintained and the lack of porosity in the coating conferred enhanced resistance to cavitational wear. Disadvantages include, hazards to factory personnel and neighborhood if the quantity of explosive used is underestimated, difficulty in coating substrates with complex form, limitation of cladding materials to ductile metals, minimum thickness or strength for the substrate.

6.7 Advanced coating techniques

6.7.1 Diamond like carbon

Diamond Like Carbon is a term used to describe hard carbon films, which are mostly metastable amorphous materials but can include a microcrystalline phase. DLC films can be obtained by using a variety of methods such as RF/DC plasma-assisted chemical vapour deposition (CVD), Arc evaporation, and ion beam deposition from a variety of solid or gaseous source materials. DLC films possess very high hardness generally in the range of 2000-5000kg/mm^2 and a low friction coefficient. The applications of DLC coatings extend from simple mechanical part to optical, or electronic components. The use of DLC is especially attractive in applications where it is required that the thickness of the protective film be less than 50nm, as in the case of magnetic recording media. One such application is in computer fixed disks, where the trend towards higher storage density and miniaturization of size has led to the requirement of very low flying heights between a disk and a recording head. The stringent requirements of ultra thin films with wear-resistant properties can be met by the deposition of hard amorphous carbon at low substrate temperature and by increasing high adhesion to head materials. The properties of the films primarily depend on the source gas and processing conditions.

6.7.2 *Electron beam assisted physical vapor deposition*

Electron beam assisted physical vapor deposition is a materials coatings technology whereby a coating (metal, alloy, or ceramic) is melted, vaporized in a vacuum, and then deposited on a component or part requiring the surface properties inherent in the coating. The principle is similar to physical vapor deposition except that an electron gun is used to cause localized melting and vaporization of the target material. The substrate to be coated is held horizontally on a fixture and the coating is obtained by vaporizing the metal/ ceramic to be coated while the fixture is rotated at a constant speed. A strong mechanical bond, uniform microstructure, and relatively high deposition rates make it an attractive and versatile coating process. Since it is performed in a vacuum, it is an environmentally friendly technology, suitable as a replacement for other coating processes in many applications. This process can easily obtain coating thickness in the order of few microns and materials, which are difficult to coat using PVD and that, require CVD techniques can be coated with this technique.

6.7.3 *Electrostatically assisted combustion chemical vapor deposition*

Another new method recently developed is Electrostatic assisted Combustion Chemical Vapor Deposition (CCVD). The process involves spraying atomized precursor droplets across an electric field where the droplets undergo combustion and chemical reaction in the vapor phase near the vicinity of the heated substrate. This produces a stable solid film with excellent adhesion onto a substrate in a single production run. This process is capable of producing thin or thick strongly adherent coating with controllable stoichiometry, crystallinity and texture. This technique has already been applied to the deposition of simple {e.g Al2O3, SiO2 , SnO2 }, multi-component oxides {e.g La(Sr)MnO3 , BaTiO3 , PbTiO3, PZT} and doped oxides {e.g Y2O3 - ZrO2 (YSZ), SnO2 - In2O3}, as well as polymer films (e.g polyvinylidene fluoride).

6.8 Summary

Discrete coatings can be made of almost any material and can be deposited in a variety of ways. By quantity, the most important method is painting which generates corrosion resistant coatings. Advanced coating methods, which permit coatings for special applications, e.g. extreme wear resistance or very high temperatures, are usually but not always performed in a vacuum. The use of a vacuum enables metal deposition to proceed without oxidation, which would otherwise disrupt the coating process. Coating under a vacuum or inert gas atmosphere can be slow or rapid, depending on the coating process and coating material. High-energy deposition using a plasma is also a very important means of deposition. Oxidation during coating deposition is prevented by excluding the air from the substrate with a flow of non-oxidizing plasma and vapor. Coating methods based on direct physical contact are also gaining popularity. An important method is friction surfacing where frictional heating and extreme wear generate clean metal surfaces for coating. Explosive bonding functions by a similar principle of direct metal-to-metal contact.

RECOMMENDED READING.

Paints and organic coatings.

John Howard, 'Versatile coatings', New Scientist, 17 February 1996, 'Inside Science' series number 88.

W.M. Morgans, Outlines of paint technology, Vol. 1 1982, Vol. 2, 1984, publ. Charles Griffin & Co. Ltd.

Oil and Colour Chemists' Association, Australia, 'Surface coatings, Vol. 1 - Raw materials and their usage', 2nd edition, Chapman and Hall Ltd., London, 1983.

Vacuum based coating methods.

D.H. Buckley, Surface effects in adhesion, friction, wear and lubrication, publ. Elsevier, Amsterdam, 1981.

International Conference on Surface Modification Technologies, the conference series.

Friction surfacing.

W.M. Thomas, The deposition of hardfacing, corrosion resistant and non-ferrous material by Friction Surfacing, The Welding Institute Research Report, (Abington, United Kingdom), No. 303, June 1986.

Explosion bonding.

R.H. Richman, A.S. Rao and D. Kung, 'Cavitation erosion of NiTi explosively welded to steel', Wear, 1995, Vol. 181-183, pp. 80-85.

Electroplating

Vincent Fleury, Wesley A. Watters, Levy Allam & Thierry Devers 'Rapid electroplating of insulators', Nature, Vol 416, 18th April 2002, pp. 716-718.

CHAPTER 7

INTEGRAL COATINGS AND MODIFIED SURFACE LAYERS

7.1 Introduction

Integral coatings are coatings, which have no distinguishable boundary with the substrate. Instead there is a continuous gradient in concentration from purely substrate material to pure coating material. This continuous variation as opposed to a sharp boundary is associated with strong bonding of the coating to the substrate and the coating is considered to be completely integrated with the substrate. Modified surface layers contain substrate material which has either had its composition, microstructure or residual stress profile altered to improve the properties of the surface, e.g. hardness and corrosion resistance. There are several methods of manufacturing integral coatings and modified surface layers. These are listed schematically in Figure 7.1.

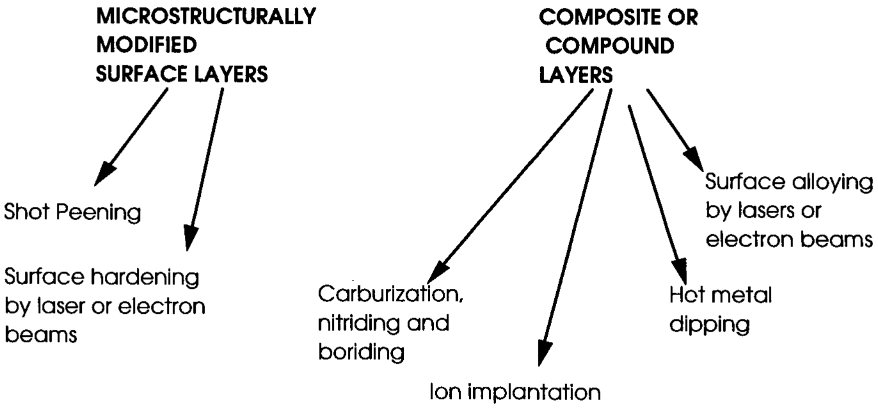


Figure 7.1 The basic categories of integral coatings and modified surface layers.

Disparate techniques involving localized and concentrated sources of heat, mechanical bombardment, controlled inter-metallic corrosion (galvanization), and solid state diffusion are used to form a wide variety of integral coatings and modified surface layers.

The boundary between a coating and its substrate is often the greatest source of weakness for a coating. Coating failure mechanisms such as mechanical de-bonding and flaking are usually located at this boundary. One means of preventing such failure of the interface between coating and substrate is to create a diffuse interface where a distinct boundary is absent. This is the common characteristic of integral coatings and modified surface layers. The diffuse boundary not only implies a graded variation in composition from coating to substrate but also a smooth transition between the hardness of the coating and the hardness of the substrate. The hardnesses of coatings and substrates are usually different and high mechanical stresses can be generated if the hardness transition is abrupt. Integral coatings are often stronger than the corresponding discrete coating of identical chemical composition because of the lack of acute stress gradient at the substrate-coating boundary. The strength of an integral coating is also less sensitive to voids and inclusions than the equivalent discrete coating because voids or inclusions cannot initiate the peeling of an integral coating. The differences in hardness profile and failure mechanisms between integral and discrete coatings are illustrated schematically in Figure 7.2.

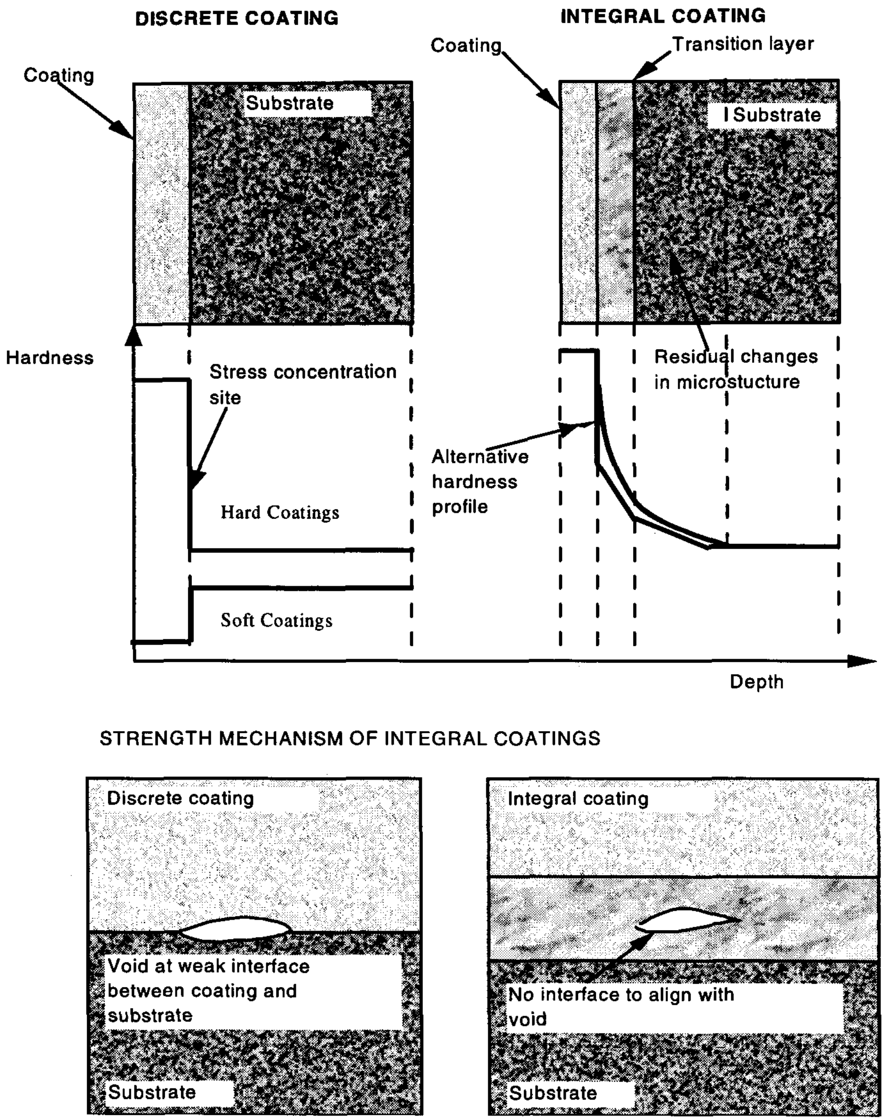


Figure 7.2. Comparison of hardness profiles and failure mechanisms for integral and discrete coatings.

It is not always necessary to deposit a completely different material on a substrate in order to improve the properties of the substrate. For example, wear resistance can be improved simply by altering the microstructure of a surface layer. It is now possible, using high-energy ionic bombardment, to create a surface alloyed layer on the substrate by what is termed implantation of ions into the crystal lattice of the substrate. Modified surface layers can be engineered to retain the advantageous material properties of the substrate while suppressing its undesirable properties. Questions, such as: What are integral coatings and modified surface layers? What is their range of applications? When are they superior to discrete coatings? These are all relevant to surface engineering.

7.2 Thermally or mechanically modified surface layers

One of the simplest means of producing a modified surface layer is to heat the substrate for a short period of time. Thermal changes to the microstructure will almost certainly result and some of these may be beneficial. Other changes such as shrinkage cracking and quench cracking are of course undesirable and ideally must be excluded from the coating/surface hardening process. Mechanical energy such as impact or wear can also be applied to change the characteristics of the surface.

Flame hardening is used in workshops to raise the surface hardness of components. A hot flame from an oxy-acetylene torch or substitute is directed at the component to be heat-treated. After a short period of exposure to the flame, the surface of the component becomes hot. Surface hardening is achieved by immediately quenching in water or oil emulsion.

The micro-structural mechanism of hardening in steels depends on the critical cooling rate for transformation of austenite to martensite (instead of pearlite and ferrite). Flame hardening is an established technology that is widely used in industry where its comparative cheapness is exploited. The entire surface of a component or selected parts of a component can be hardened by appropriate application of a flame. The prime limitation of flame hardening is the heat intensity that can be achieved with a flame. It is found that microstructures with superior corrosion or wear resistance are obtained only when higher heating intensities are used. Indeed, this is the main reason for the development of more advanced heat treatment technologies.

7.2.1 Induction hardening

When a conducting body is placed in the vicinity of a magnetic field, a transient induced current flows in the conducting body. If the magnetic field is generated by an oscillating current and is itself therefore, oscillating in field strength, a continuous induced current will flow in the conducting body. The resistive heating caused by induced current flows can be exploited to produce a clean non-contact method of heating any conductive body. Induction hardening is widely used in industry to generate 'hard-facing' which are non-ferrous wear resistant coatings. A major disadvantage of the process is the need for a magnetizing coil that is shaped to closely fit the component.

Application of induction hardening to steels

When the heated body is magnetic as is the case for most steels, magnetic hysteresis becomes an important additional source of heating. Magnetic hysteresis causes additional heating by energy dissipation from high frequency reversals of magnetic field induced inside the steel. In graphical terms, the heating effect from induction can be stated as the sum of electrical and magnetic heating, i.e.:

Inductive heating effect = Ohmic resistive heating from induced current + Magnetic hysteresis loss from reversal of magnetic field.

When steel reaches the Curie temperature, such induced magnetism ceases and it becomes necessary to compensate for the loss of heating from magnetic hysteresis by increasing the electrical power available to the inductor coil. For steels, the Curie temperature is 768°C when the carbon content is less than 0.45%C. At higher concentrations of carbon, the Curie temperature will vary and there is a progressive decline in magnetism from 200°C below the Curie temperature. Other magnetic metals such as cobalt and nickel have Curie temperatures of 1120°C and 335°C .

High frequency alternating electrical current tends to flow through the outer layers of any conductor so that if the induced current is generated by a high frequency oscillating magnetic field, surface heating of a conductor is possible. This is the principle of surface hardening by induction, which is illustrated schematically in Figure 7.3.

Cooling of the heated material is achieved by quenching with water or other fluids. In common with other surface hardening methods, there is only a very short holding period at high temperature. There is no 'thermal soaking' as otherwise the core of the treated component would also become hot thereby defeating the purpose of induction hardening which is to harden the surface of the material while leaving its core unaffected. To compensate for the short holding period, it is an accepted practice to reach a temperature about 50°C higher than the equivalent temperature for conventional hardening. A heating rate of the order of 5000 °K/s is possible at the surface of a metallic substrate. The depth of heating is typically a few mm. There is always some heating in the interior of the substrate as it is not possible to prevent any induced current from flowing at the core of a conductor.

The effect of induction hardening on a steel component is evaluated in terms of the level of hardness increase and the depth of hardening. The hardness increase is determined primarily by the carbon content of the steel, with some influence by induction heating and quenching rates. Carbon levels from 0.30% to 0.50% C give a good hardening effect without quench cracking, giving typical hardness values ranging from 50 to 60 HRC. At higher carbon levels, trial runs of induction hardening at various process settings should be performed to check for quench cracking prior to mass production. Figure 7.4 shows a platen shaft being progressively hardened followed by rapid quenching while Figure 7.5 shows the microstructure of the platen shaft after induction hardening.

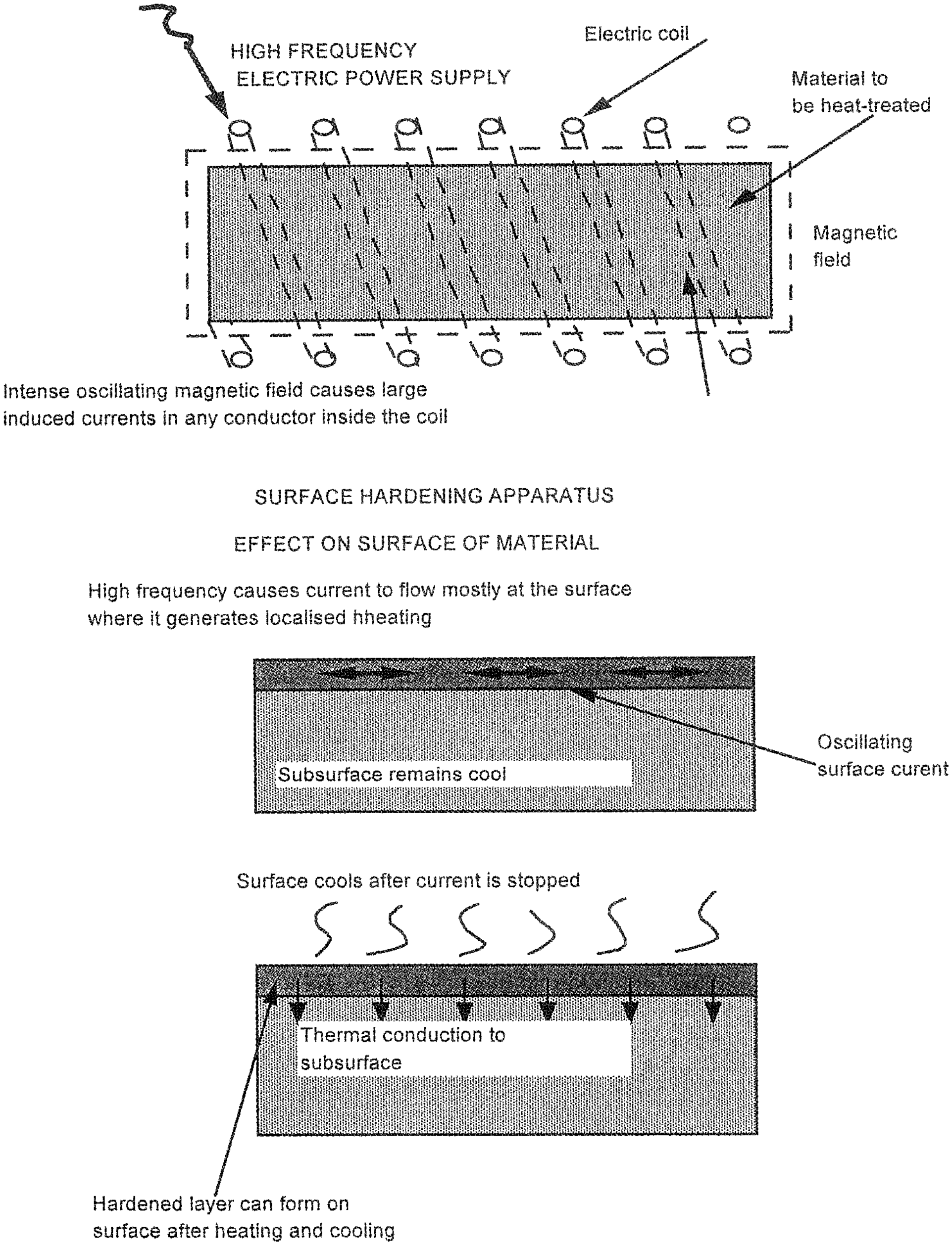


Figure 7.3. Operating principle of surface hardening by magnetically induced current.

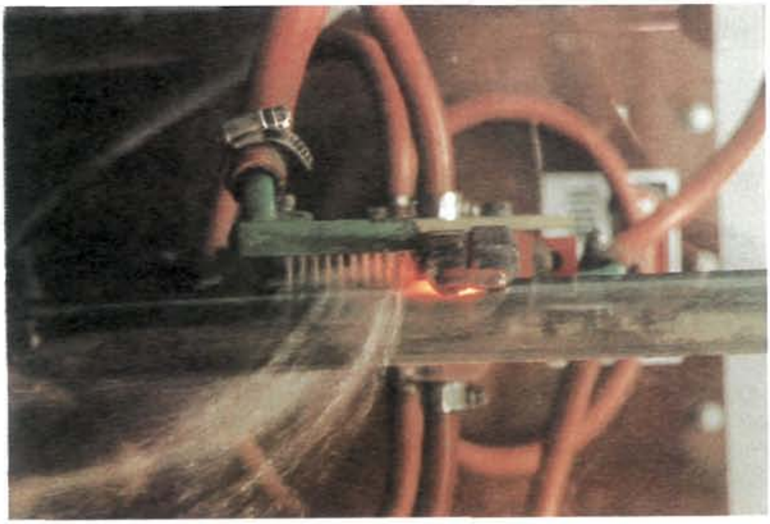


Figure 7.4 Induction hardening of platen shaft. (Courtesy of Doxon Engineering Pte. Ltd, Singapore.)

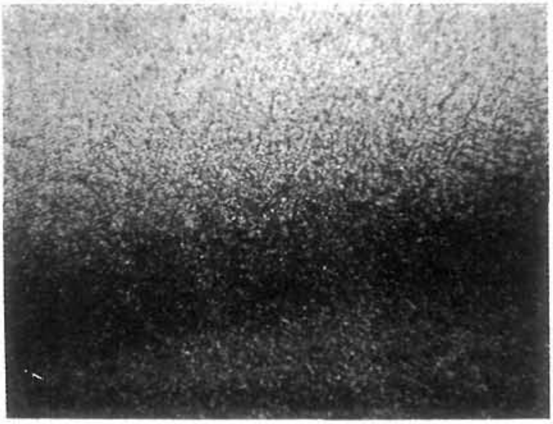


Figure 7.5. Microstructure of steel, induction hardened to a surface hardness of 800 VPN (Courtesy of Doxon Engineering Pte. Ltd, Singapore.)

A wide range of steels can be induction hardened - both the plain carbon steels and low chromium steels. The chemical composition of some steels commonly used for this purpose are shown below in Table 7.1

Table 7.1: Composition of steels suitable for induction or flame hardening.

C%	Mn%	Cr%	Mo%	Ni%
0.33- 0.39	0.50 - 0.80	0	0	0
0.38 - 0.44	0.50 - 0.80	0	0	0
0.43 - 0.49	0.50 - 0.80	0	0	0
0.48 - 0.55	0.60 - 0.90	0	0	0
0.50 - 0.57	0.40 - 0.70	0	0	0
0.42 - 0.48	0.50 - 0.80	0.40 - 0.60	0	0
0.34 - 0.40	0.60 - 0.90	0.90 - 1.20	0	0
0.38 - 0.44	0.60 - 0.90	0.90 - 1.20	0	0
0.38 - 0.44	0.50 - 0.80	0.90 - 1.20	0.15 - 0.30	0
0.38 - 0.44	0.70 - 1.00	0.40 - 0.60	0.15 - 0.30	0.40-0.70
0.37 - 0.43	0.50 - 0.80	0.60 - 0.90	0.15 - 0.30	0.70 - 1.00

Si: 0.15 - 0.40%, **P**: 0.035% max, **S**: 0.035% max, for all steels.

For optimal hardening effect, the microstructure of carbon steels should consist either of tempered or normalized martensite. With annealed steels, the dissolution of carbides at austenitic temperatures requires a longer soaking time than the duration of the heating-up phase of induction hardening. Annealed steels therefore, require longer heating periods to dissolve carbides at austenitic temperatures and hence not suitable for induction hardening.

Induction hardening is generally followed by tempering at 150 to 200°C to achieve lower surface hardness. Reduced hardness is crucial if the component needs to be subsequently ground or where the component must be resilient against impact load. Figure 7.6 shows the typical effect of tempering on the hardness profile of alloy steel. Where localized tempering on some part of the component is required, the same induction coil can be used for tempering following the initial induction hardening.

The hardening depth is basically defined by the depth of martensitic transformation, which in turn is governed by the depth where the temperature is above the austenizing temperature of the steel. This depth is effectively controlled by the power and the

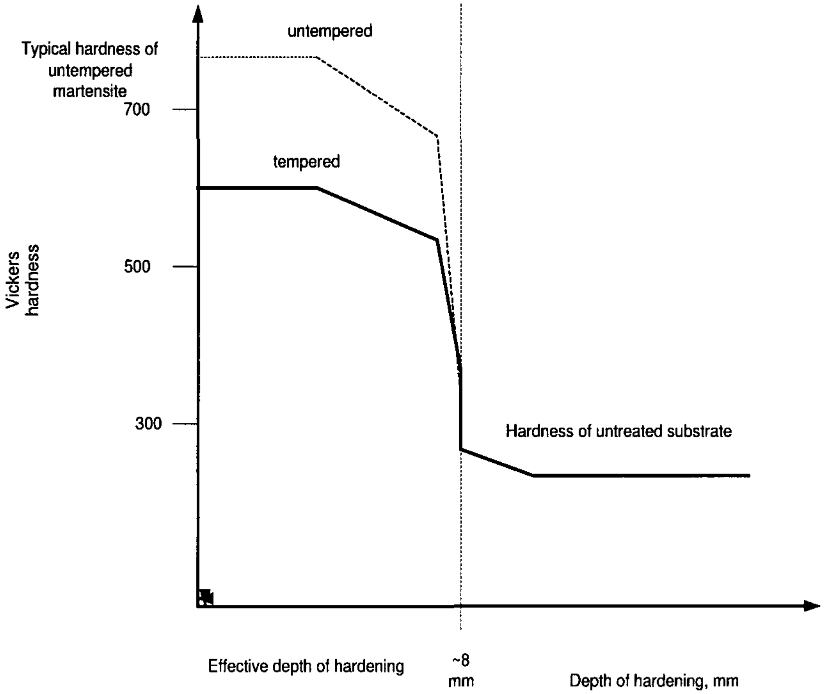


Figure 7.6 Schematic diagram depicting the effect of tempering on the hardeness profile of alloy steel (Adapted from *Thelning*).

frequency of the alternating current that is used. Typical power densities range from 1 MW/m^2 to 20 MW/m^2 of the surface to be treated. For a given power, the depth - frequency relationship can be expressed by the equation:

$$d = k/f^{0.5} \qquad \{7.1\}$$

where d is the depth of hardening [m],f is the frequency of the alternating magnetic field of the inductor [s^{-1}] and k is a constant [m*s$^{0.5}$]. The value of k varies from 0.02 at 20°C to 0.5 at 800°C.

Equation {7.1} is at best a first approximation for predicting the depth of hardening.

The depth of hardening is closely related to the depth to which induction currents flow below the surface and this depth can be more precisely calculated than the hardening depth. The equation for the depth where the electrical current is 1/e or 37% of the surface value is defined as:

$$d' = k_1 * (\rho / (\mu * f))^{0.5} \qquad \{7.2\}$$

Where d' is the limiting depth [m], k_1 is a constant [$(m*s)^{0.5}$], ρ is the resistivity [Ohm-m], μ is a dimensionless relative magnetic permeability of the component and f is the frequency as defined in {7.1}. The value of k_1 is 500 $(m/(Ohm*s)^{0.5}$ and a typical value of μ is 150 for iron with less than 0.05% carbon but can increase to 1000 for high silicon iron. The resistivity of carbon steel is approximately 1.7E-7 ohm-m. It is usually assumed that the depth of hardening, d, is approximately 50% of the d'.

According to {7.2}, a 10 mm depth of current flow in a carbon steel requires a frequency of approximately 4 Hz. The depth of current flow is inversely proportional to the square of the frequency so that a 1 mm depth of current flow necessitates a 400 Hz frequency of current. By comparison, the public electricity supply has a frequency of 50 Hz, which would enable a hardening depth of approximately 3 mm to be obtained. The wide range of electrical alternating current frequency required by induction hardening imposes the extra cost of specialized electrical supply systems for a manufacturing enterprise.

7.2.2 Laser and electron beam surface hardening

When subjected to a sudden short pulse of thermal energy, the surface layers of most metallic materials cool down as fast as they heat up. This is due to the good thermal conductivity of metallic materials. Frictional surface heating manifests this characteristic of efficient thermal conduction, where the worn surface is subjected to transient high frictional temperatures. Similarly, precipitate changes in surface temperature can be achieved by directing an external source of thermal energy to the surface. Both laser and electron beams have the characteristic of acting as a localized but intense source of thermal energy on any surface that they impinge against. When a laser or electron beam scans a surface, any point on the surface lying in the path of the scan will experience a precipitate rise in temperature when the beam passes over that point. If the scanning speed is sufficiently high, the period of heating will be very short so that any adsorbed thermal energy remains in a surface layer. As soon as the beam is directed elsewhere, thermal conduction ensures extremely rapid cooling of the surface. Cooling rates of 100,000°K/s have been deduced from microstructural examination of laser scanned specimens while for electron beams, a cooling rate of 1000 to 10 million °K per second have been estimated. This very rapid cooling produces a surface layer of quenched material about 0.1 to 1.0 mm thick with microstructures very different from the bulk microstructure. Thus, it is possible to produce a martensitic surface layer on top of a ferritic steel using laser or electron beam. A hardened layer with extremely fine microstructure is typically generated for steels. Laser treatment of cast iron produces a hard surface, which displays higher wear resistance than the untreated cast iron.

When any point on the treated surface is irradiated by the laser or electron beam for a short period, localized melting is possible around the impingement area of the beam provided that the laser or electron beam has sufficient energy. For most engineering materials, localized melting can be achieved with a power of a few kW and a beam diameter of approximately 1 mm. This quantity and density of power is within the reach of commercially available lasers and electron beam systems. The energy flux within the molten material is so high that strong convection currents homogenize the liquid material. If the surface to be treated is pre-coated with another material, this material will be mixed or alloyed with the substrate by laser/electron beam scanning. It is therefore possible to produce an alloyed surface layer, a popular example being the pre-coating of chromium onto steel to create a coating of stainless steel for corrosion or wear resistance. This process is known as *Laser Surface Alloying* (LSA) or *Electron Beam Surface*

Alloying (EBSA) depending on the energy source. The processes of heat conduction, localised melting and homogenization, surface hardened layers and surface alloyed layers are illustrated schematically in Figure 7.7.

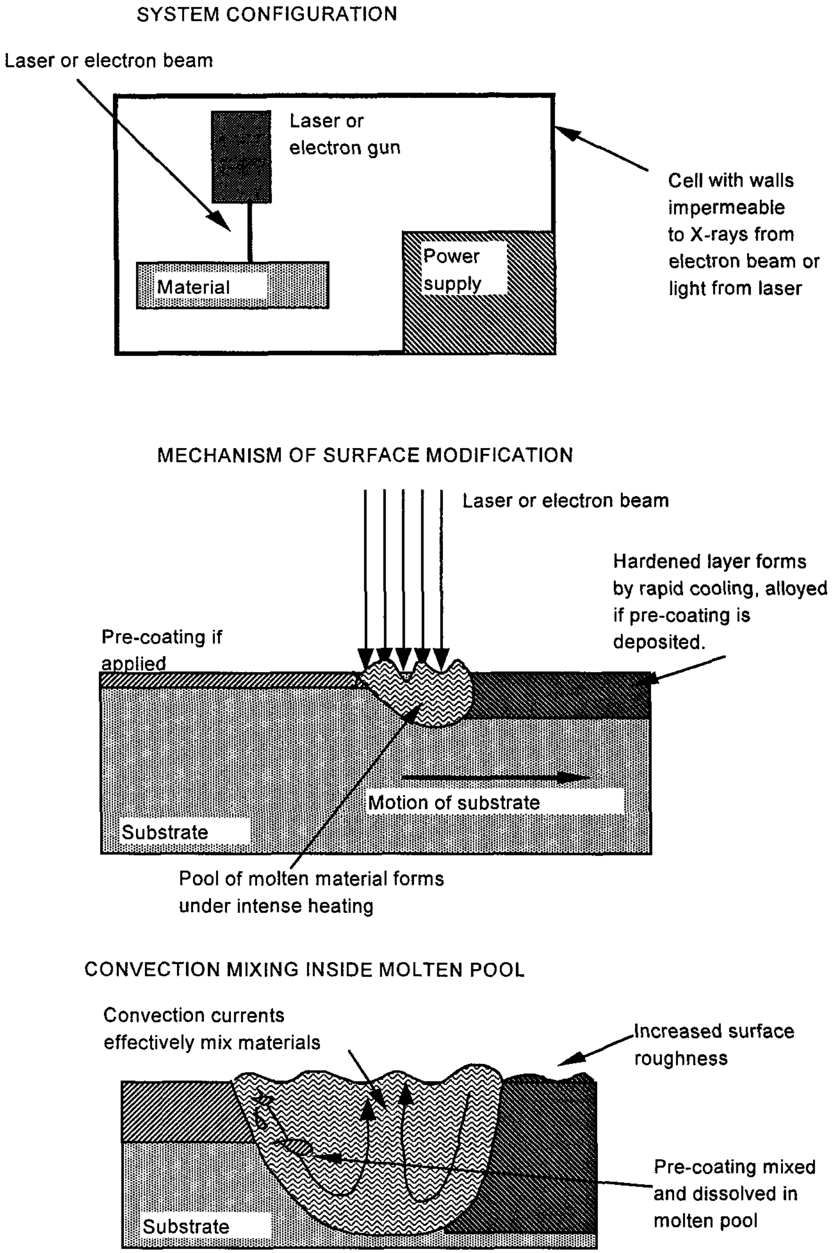
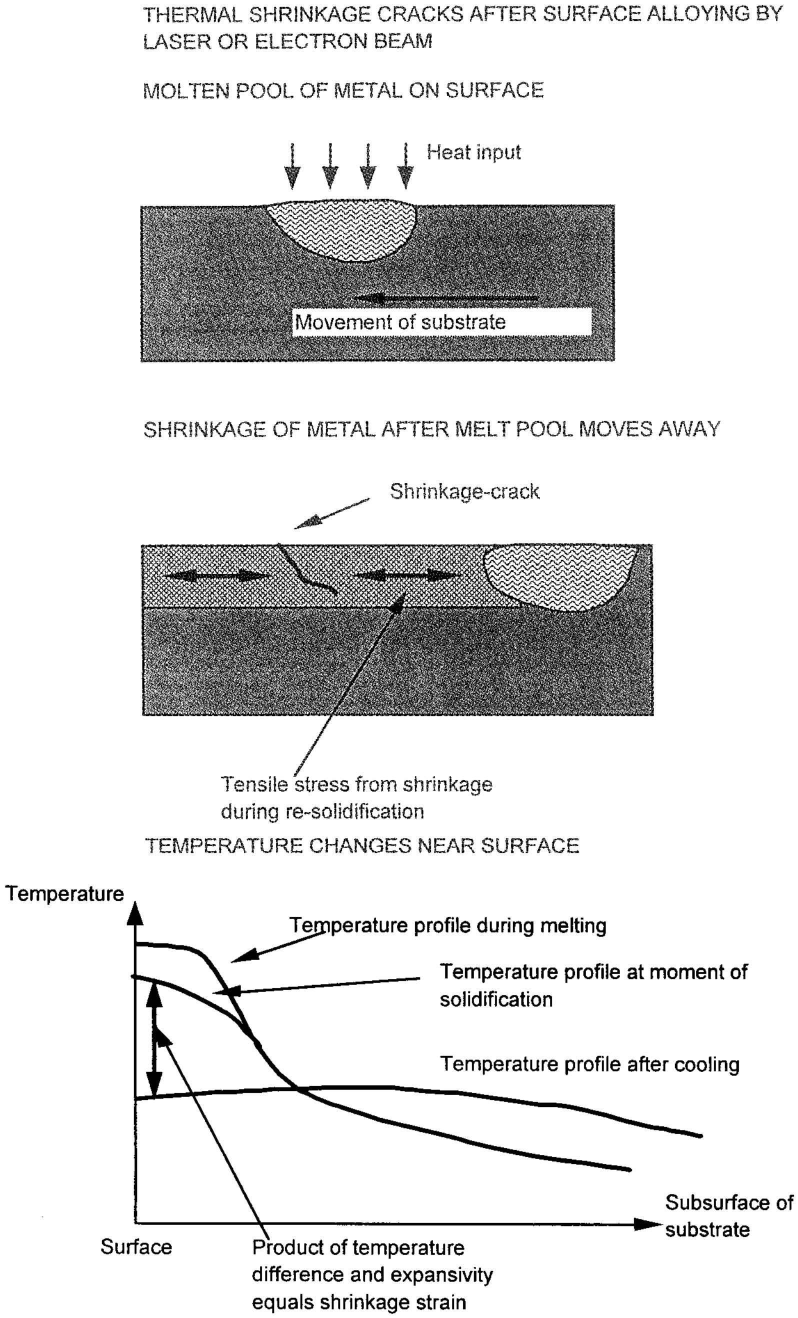


Figure 7.7. Mechanism of surface hardening and surface alloyed layer formation by scanning with high power laser or electrom beams.

LSA and EBSA are currently actively being investigated and developed as a means of producing high quality wear or corrosion resistant coatings with the minimum consumption of precious non-ferrous metals. Not-withstanding the distinct advantages of EBSA and LSA, their applications are limited to electrically conductive materials and optically non-reflective surface respectively. In addition to the above, the high equipment cost of EBSA and LSA, especially that of EBSA which requires a vacuum system, poses a serious limitation on the economic viability of such techniques in medium and small-scale industries. Steel components must be demagnetized prior to EBSA and an inert gas jet should be directed at the site of laser beam impingement to prevent excessive oxidation of the substrate during treatment. Steels are prone to excessive oxidation damage without an inert gas atmosphere. The coatings produced by EBSA and LSA are not perfectly homogeneous as the surface tends to have more alloying element than the subsurface and solidification cracks are likely to occur. Treated surface is rougher than most machined surfaces and may require secondary finishing operation before the component can be put to use. The mechanism of thermal shrinkage cracking in LSA and EBSA coatings is illustrated schematically in Figure 7.8.

Cracking is the result of two principal factors, (i) different cooling rates of the various parts of a metal surface after LSA and EBSA, (ii) the contrast between adjustment by plastic flow to expansion of metal at high temperatures and brittle fracture upon contraction of metal at low temperatures.



THERMAL SHRINKAGE CRACKS AFTER SURFACE ALLOYING BY LASER OR ELECTRON BEAM

MOLTEN POOL OF METAL ON SURFACE

SHRINKAGE OF METAL AFTER MELT POOL MOVES AWAY

TEMPERATURE CHANGES NEAR SURFACE

Figure 7.8. Mechanism of thermal shrinkage cracking in Laser Surface Alloyed and Electron Beam Surface Alloyed coatings.

7.2.3 Shot-peening

Shot peening is a process where a metal specimen is impacted by a larger number of small steel balls with diameter ranging from 0.18 to 2.00 mm. With sufficient kinetic energy, the impacts of the steel balls cause compressive stresses aligned to the surface to be formed just below the impacted surface. Surface compressive stresses help to suppress fatigue wear and may be used to counteract the weakening effect of surface cracks. Impact of the balls also causes plastic deformation at the surface of the specimen and this causes work hardening.

The application of shot peening

Shot peening is widely used in the manufacture of helical springs, transmission shafts and torsion bars for automotive or aerospace applications. The purpose of shot peening is to introduce residual compressive stresses in the surface of the component. This residual compressive stress delays fatigue crack initiation thereby increasing their fatigue life. Compressive stress formation is only possible with ductile substrates, which are usually metals. With ceramics, e.g. alumina, shot peening is more likely to damage the surface. A major limitation of shot peening is the difficulty in predicting the protective effect, which depends on the complex mechanics of hard metal balls bouncing against each other and a softer metal surface. Critical parameters of shot peening such as ball impact velocity and impingement angle are difficult to measure and control. The formation of sharp edges on fractured balls after sustained use adds further complication to prediction or control. The depth of work hardening of the surface layers is inversely proportional to the hardness of the material undergoing shot peening. The depth of hardening is also proportional to the size of balls until a critical ball diameter is reached where little further hardening is obtained. Improvements in fatigue life of approximately 10 times and increases in limiting stress under fatigue of 10% have been claimed. For design purposes, 50 to 100% increase in fatigue life would be a more reasonable estimate.

Evaluation of shot peening is approximate and can be performed by metallurgical analysis or by non-destructive methods. Metallurgical analysis typically reveals plastically deformed layers. A metallurgical cross section of a shot peened steel is shown below in Figure 7.9.

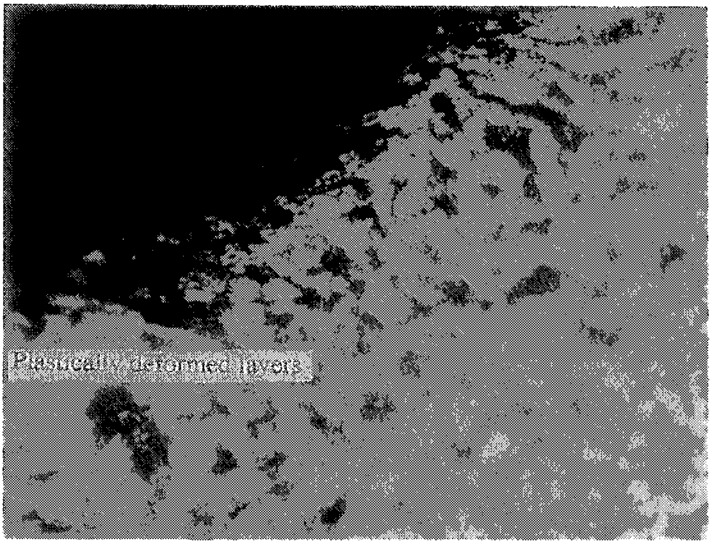


Figure 7.9. Subsurface microstucture of steel after shot peening.

For steels, the effect of shot peening is very sensitive to the initial microstructure and consequently the heat treatment, which controls microstructure. For a 0.5% medium carbon steel subjected to various heat treatments, it was found that shot peening after oil-quenching and tempering produced the largest hardness increase of 1100 Hv close to the surface as compared to a core hardness of 200 Hv. For annealed and normalized steels, the maximum hardness attained was limited to 300 Hv only. Regardless of heat treatments, the hardening effect was limited to within 500 μm of the surface. Rotating -bending fatigue test data revealed that shot peening caused the greatest extension in fatigue life for the oil-quenched and tempered specimens. Evaluation of shot peening is based on the use of Almen strip as detailed in standards such as SAE J442 and J443. The principle of measurement is the curvature of a strip when exposed on one side to shot peening. Strips of different thicknesses of 0.79, 1.29 and 2.39 mm with a tolerance of +/- 0.02 mm are used for evaluation of shot peening. The intensity of shot-peening is then assessed from the depth of bending of the strip from its mid-point to either of its ends. This test is considered to be qualitative as it provides no information on microstructural changes induced by shot peening nor any fatigue life data. The test is very convenient and easy to use as compared to microstructural analysis.

Erosion processes related to shot peening

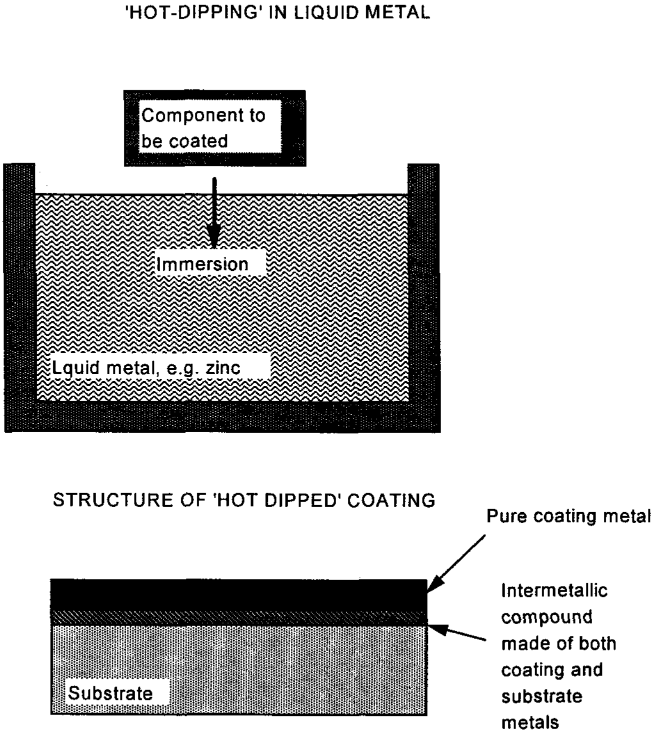
A surface preparation technique that is similar to shot peening, but differing in application is the erosion. Erosion is used for the removal of surface layers such as paints, oxides and corrosion products to prepare for the final coating. Finer particles such as sand, are used for the removal of coatings as opposed to shot peening by hard steel balls. Sand-blasting as the technique is known, is used for the cleaning of ship-hulls of rust, marine organisms and old paint. In a recent development, modified wheat starch particles are used as the eroding medium to remove paint from aircraft fuselage without damaging the aluminum panels. Damaging the aluminum panels by particle impact will result in stress concentration sites, which are highly undesirable. The wheat-starch is non-toxic unlike the earlier paint stripping methods based on paint solvent (methylene chloride).

7.3 Thermochemical methods of coating

The combination of surface heating and diffusion of hardening agents forms the basis of many surface treatments that are vital to manufacturing technology. Most but not all of these methods have been in use for a considerable period of time and are applied to a wide variety of metal components.

7.3.1 Galvanization and hot-dipping

It is not always necessary to use the sophistication of a laser to achieve an alloyed coating on a metal substrate. Galvanization, which is a specialized form of hot-dipping, develops alloyed and intermetallic compound layers on metal surfaces by dipping the entire metal component into a bath of molten metal. The molten metal not only wets the surface of the component in the same way as water wets an immersed object but also reacts with the metal substrate to form intermetallic compounds. This process is an example of corrosion by liquid metals discussed in Chapter 3 and hot-dipping has to be carefully optimized to ensure the correct balance of metal coating to intermetallic compound layer. Hot dipping is most effective when there is a large difference between the melting point of the coating metal and the melting point of the substrate metal. Typical applications of hot-dipping are zinc on steel and tin on steel. Hot dipping of metals other than steel is comparatively rare. Hot-dipping is illustrated schematically in Figure 7.10.


'HOT-DIPPING' IN LIQUID METAL

STRUCTURE OF 'HOT DIPPED' COATING

Figure 7.10. The process of Hot-Dipping and structure of hot-dipped coatings.

Galvanization

Galvanization is an effective, cheap means of preventing aqueous corrosion of steel components and structures. In the galvanization process which is mostly applied to steel sheet, a steel sheet is passed through a bath of liquid zinc at about 450°C at speeds of ≈ 1 m/s for a period of time ranging from 10 seconds to over a minute depending on the thickness of coating required. Coating thickness is usually specified in terms of weight of coating material deposited per unit area and is of the order of 0.05 to 0.1 kg/square metre. This is equivalent to a coating of approximately 10 μm in thickness. A small amount of aluminum, about 0.1 wt/%, is added to the molten zinc to modify the microstructure of coating. The coated component is rapidly cooled after exit from the zinc bath to suppress further formation of inter-metallic compounds. The galvanized coating has a multi-phase laminar microstructure that ranges from pure zinc directly beneath the exterior surface to zinc-iron compounds in the interior of the coating with pure iron or steel below. The galvanized coating possesses very strong bonding to the substrate because of the reaction between iron and zinc and so almost never peels off the substrate. Galvanized coatings are however vulnerable to mechanical wear and have poor corrosion resistance under acidic conditions. For practical purposes, the minimum pH of any acid in contact with galvanized coatings should be greater than 4 or 5. Sliding or fretting contact with hard materials, e.g., steel rapidly wears out galvanized coating to expose the substrate, which is plain steel.

A recent development is the manufacture of 55% aluminum-zinc coatings by a continuous hot-dipping process. This alloy coating is found to exhibit significantly improved life over conventional zinc coatings. The microstructure of the solidified alloy coating contains dendrites rich in aluminum and a zinc rich inter-dendritic region phase. The variation in zinc concentration causes preferential corrosion of zinc in the inter-dendritic region and sealing of the inter-dendritic space. Corrosion resistance of the coating is therefore controlled by the lifetime of the aluminum rich phase (dendrites).

Hot dipped pure aluminium coatings are a possible substitute for galvanization where the aluminum coating shows superior corrosion resistance compared to both zinc and aluminum-zinc coatings. As noted above, galvanized coatings have poor wear resistance and the use of hard metal rollers for continuous feeding of the steel sheets results in loss of coating efficiency. Recent research has been directed to the removal of the bottom rolls in continuous galvanization. Another currently researched project is on the use of electrochemical galvanization for sintered porous parts. Conventional methods of galvanization do not provide a very good corrosion resistant coating on porous and permeable surfaces. The new process uses a weakly acidic galvanization electrolyte and operation is performed either with low current densities or high current densities for low metal content.

7.3.2 Carburizing, carbonitriding, nitriding , nitrocarburizing & Boronizing

These coating techniques share a common purpose of improving the mechanical properties and wear resistance of a component. They are rarely used to improve corrosion resistance. The substrate is mostly steel or iron-based alloys. The process is performed at high temperatures where some modification of the bulk substrate microstructure is inevitable. Hardening effect is achieved by the direct formation of hard phases in the ferrous substrate without or followed by a subsequent quenching of the entire component. The tendency of the surface to cool more rapidly than the interior of a component during quenching is exploited to obtain hard surface layers.

From the viewpoint of ferrous metallurgy, carburizing and carbonitriding are classified as austenitic thermochemical treatments. The underlying principle of this classification is that diffusion of carbon and nitrogen into the iron/steel is performed in the austenitic temperature range followed by quenching to room temperature. In contrast, nitrocarburizing and nitriding (discussed in Chapter 6) are classified as ferritic since the impregnation of nitrogen and carbon into the substrate is performed below the eutectoid temperature of steel. A major practical distinction is that austenitic treatments cause much more distortion of the treated component than ferritic treatments because of the transition from austenite to martensite, which results in a larger volumetric change of unit cell.

Carburization

Carburizing is the enrichment of the surface of a ferrous component with carbon for subsequent quench hardening. In order to achieve practicable treatment times, this process is normally performed between 825 and 925°C. Carburization is based on close contact with either solid media (lumps of carbon or coke) or a carbon-rich salt bath or in carbonaceous gases. When solid media are used, the process is called pack carburizing. Pack carburizing is inconvenient because it necessitates elaborate and cumbersome handling of components. Environmental constraints are progressively restricting the scope for salt-bath carburizing since cyanides are used as the carburizing agents. Cyanides are salts of hydrogen cyanide, which has the chemical formula HCN, both the acid and its salts are lethally toxic. Gas carburizing and a recently introduced technique, plasma carburization, are now the most commonly used techniques. Pack carburization, molten-salt and gas carburization are illustrated schematically in Figure 7.11.

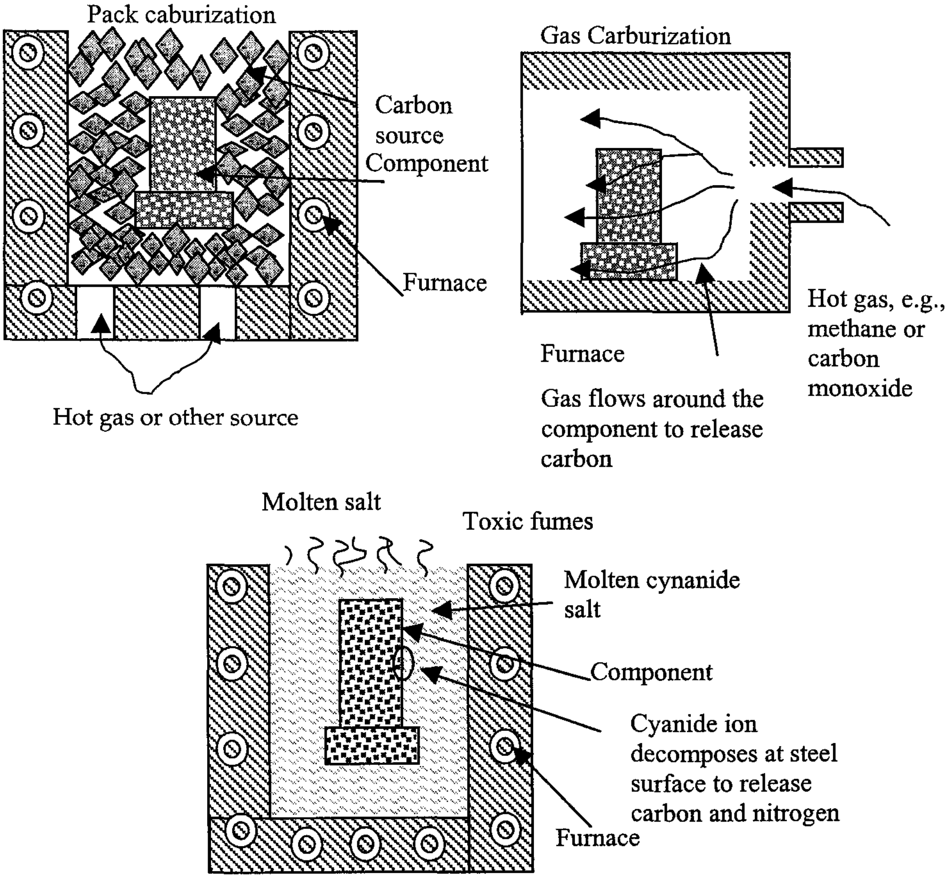


Figure 7.11 Schematic illustration of the carburizing methods

Whichever method is employed, carburization depends principally on the presence of carbon monoxide at the surface of the component to liberate atomic carbon for diffusion into the substrate. With gas carburizing, generally three chemical reactions take place in the furnace where carburization is performed. These reactions, which release the atomic carbon can be expressed as follows,

The Methane reaction CH_4 <----> C + 2 H_2 {7.3}

The Water-gas reaction CO + H_2 <-----> C + H_2O {7.4}

Oxidation of carbon 2CO <--------> C + CO_2 {7.5}

Carburization is normally performed with a surface carbon concentration ranging from 0.7 to 0.80%. If the surface carbon concentration remains constant during carburization, then the carbon content as a function of depth below the surface and time can be directly found from Fick's 2nd law of diffusion to give the equation:

$$C_x - C_c = (C_s - C_c)[1 - e^{\left(\frac{x}{2\sqrt{Dt}}\right)}] \qquad \{7.6\}$$

Where C(x) is the concentration of carbon in the metal (steel) in wt% at depth 'x' [m] below the surface. C(s) is the surface concentration of carbon, i.e. at x = 0, and C(c) is the initial concentration of carbon prior to carburization. D is the diffusivity coefficient [m^2/s] and 't' is time since the onset of carburization [s]. Equation 7.6 predicts that at infinite time, the carbon concentration reaches the surface value for all depths below the surface. At finite times, the carbon concentration is predicted to decline asymptotically to the initial value, C(c), with increasing depth. An example of the subsurface distribution of carbon concentration is illustrated in Figure 7.12.

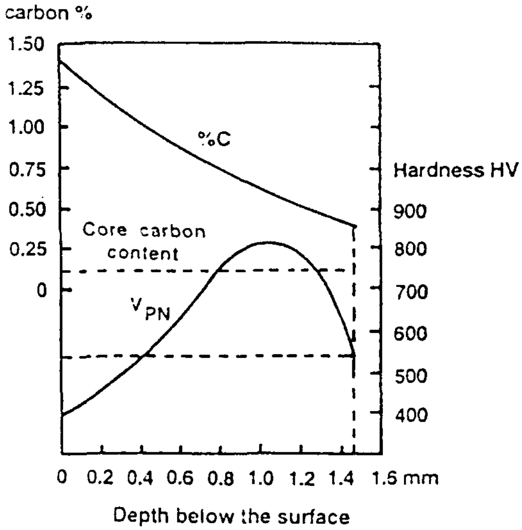


Figure 7.12. Theoretical subsurface distribution of carbon during carburization (Courtesy of *Thelning* and Bofors Ltd) .

It can be seen from Figure 7.12 that the carbon distribution declines gradually with depth so that a truly integral coating is formed. Once the process of carbon accumulation is completed, the furnace temperature can be lowered to 830°C for hardening. The mode of hardening controls the final characteristics of the 'case' or carbon enriched layer with high hardness. In the simplest instance, the component can be quenched to produce a mostly martensitic surface layer. The martensitic layer is usually too brittle for use even though it has potentially good wear resistance. For this reason, it is a common practice to reheat the quenched component and temper it followed by slow cooling. The quench may also be modified by performing rapid cooling to an intermediate temperature followed by slower cooling rate. The heat treatment processes of carburized components employ the same metallurgical principles as for uncarburized components.

The depth of carburization, which is the depth of penetration of carbon from the surface into the substrate, is not easy to determine precisely since the substrate has its own original carbon content. As a convenient approximation, the depth of carburization is often specified as the depth below the surface where the carbon concentration reaches a specified level, e.g. 0.4%C. The depth of carburization will depend on the temperature, the treatment time, the carbon potential of the carburizing medium and the composition of the steel.

The carbon potential may be defined as the carbon content, which a steel foil can acquire when equilibrium is established between the carburizing medium and the specimen. In practice, this means the carbon content in the specimen after a very long time has elapsed. The carbon concentration at the surface of the component and the depth of carburization are both proportional to the carbon potential. The effect of carbon potential and time on the subsurface carbon distribution in steel is illustrated graphically in Figure 7.13.

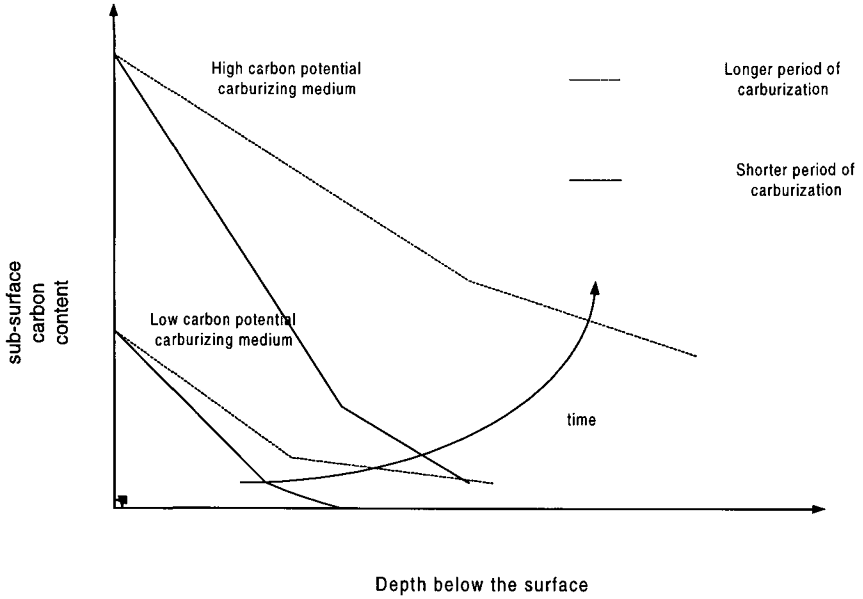


Figure 7.13. Effect of carbon potential on carbon distribution in a steel component (Adapted from *Thelning*).

Once the distribution of carbon in the steel component is established, the next question to resolve is the phase or compound of carbon that were formed. Austenite is the stable phase at temperatures where carburization is performed, the carbon solubility in austenite can be determined using the iron-carbon binary phase diagram. When carbon exceeds its solubility limit in austenite, metal carbides will form from iron or the alloying metals. Common alloying metals such as Cr (chromium) , W (tungsten), Mo (molybdenum) , V(vanadium) and Ti (titanium) are called 'carbide formers' and will readily form Cr_7C_3, WC, Mo_2C, VC, TiC or complex carbides such as Fe_4W_2C and Fe_4Mo_2C. These carbides can confer useful increase of wear resistance of the steel. Carbides can also be deleterious because of the loss of dissolved alloying metal from the steel.

In most engineering applications, less attention is given to the carbon distribution than to the depth to which hardness exceeds a specified minimum. A parameter, known as the 'case depth', is commonly measured by hardness tests alone. The case depth is widely used for quality control of the carburization process. It should be noted here that carbon is not an easy element to analyze or assay since it is a non-metal of low atomic weight, which for reasons discussed in Chapter 8 impede metallurgical analysis. The case depth is often defined as the depth below the surface where the hardness equals say 550 VPN. Hardness measurements are performed by micro- indentation on a sectioned metallurgical specimen where the section is in a plane normal to the 'case'. Representative hardness profiles of alloy steels are shown in Figure 7.14.

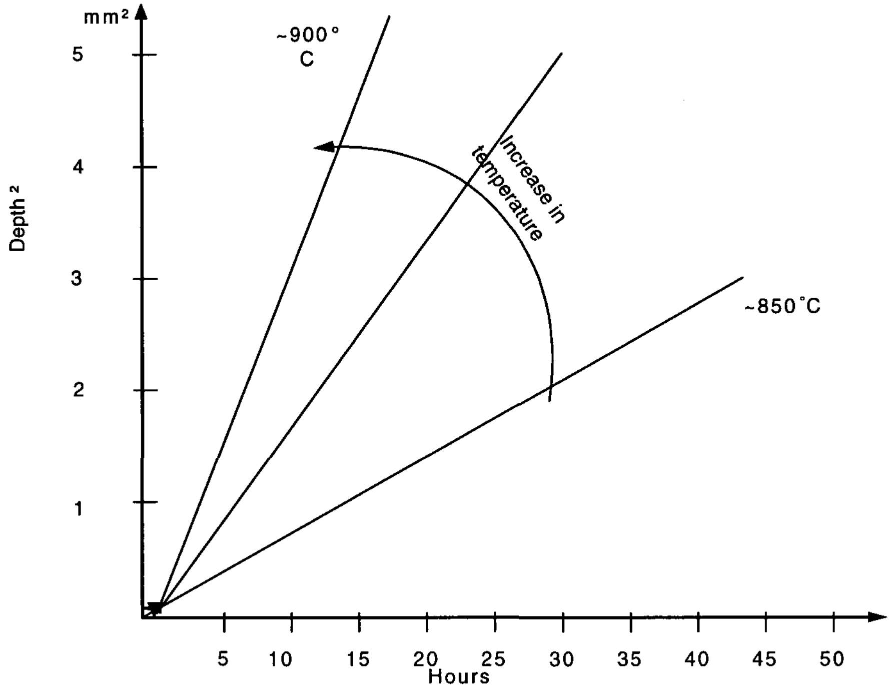


Figure 7.14. Typical hardness profiles of carburized alloy steels (Adapted from *Thelning*).

The depth of hardening does not necessarily coincide with the depth of carburization because the hardening is a result of the combined effect of carbon impregnation and rapid cooling of the substrate. If the cooling rate is rapid and promotes martensitic transformation, then the case depth will increase independently of the thickness of the carburized layer. The dimensions of the treated component as well as the quenching medium control the cooling rate. Smaller components cool faster than large components; which means that component size and case depth are inversely related. The hardness distributions of alloy steel specimens for varying size are shown schematically in Figure 7.15. For a case hardness level of e.g. 700 VHN, there is a 40 % increase in the case depth with decrease in specimen diameter from 100 mm to 10 mm.

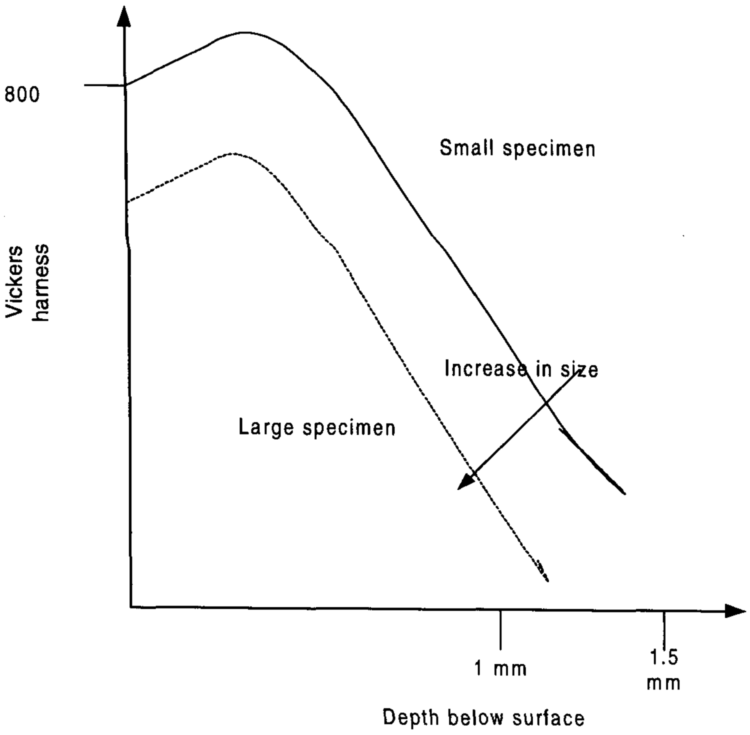


Figure 7.15. Schematic illustration of influence of specimen diameter on hardness distribution (Adapted from *Thelning*).

The effect of case hardening is more pronounced in alloy steels, which have a higher hardenability than the plain carbon steels. Alloy steels primarily contain chromium, molybdenum and some of the other trace elements such as silicon, phosphorus and sulfur. A high hardenability implies that martensite will form from austenite even at a moderate cooling rate typical of large components. If a significant depth of hardening is desired in a large component, then it is expedient to choose alloy steels instead of plain carbon steel.

Predicted case depth as defined by a specific hardness value for varying carburizing temperature and time is shown schematically in Figure 7.16. It can be seen that maximum hardness is obtained at the surface of the component for shorter periods of carburizing, whereas, longer periods of carburizing increase the case depth and shift the peak hardness to subsurface.

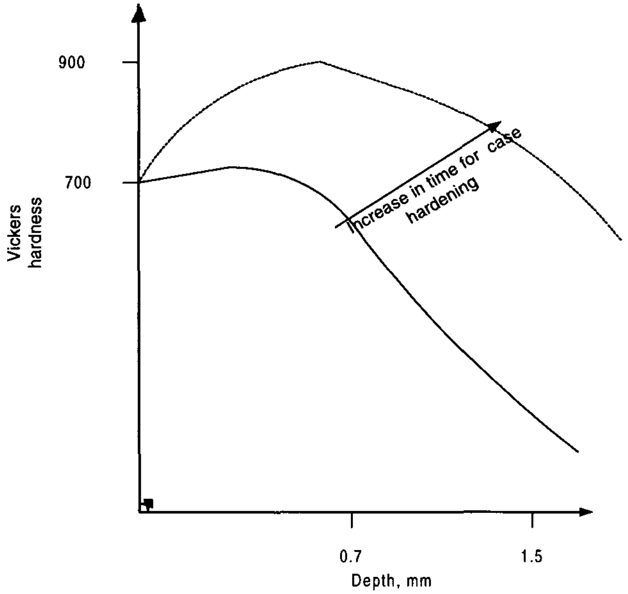


Figure 7.16. Schematic illustration of effect of carburizing time on case depth (Adapted from *Thelning*).

It is also possible to predict the case depth by applying the general law in diffusion that thickness is proportional to the square root of time. If the effect of rapid cooling is ignored and it is assumed that a hardness of 550 VPN is obtained by the martensitic transformation of surface material with a minimum 0.4%C, then the case depth can be determined by the parabolic relationship:

$$x = k^{*}t^{0.5} \tag{7.7}$$

Where x is the case depth [m], k is a constant [$m^{*}s^{-0.5}$] and t is time [s]. It is found that in general, this law applies and the values of k as a function of temperature [Thelning] are :

Temperature [°C]	Constant, k [$m^{*}s^{0.5}$]
875	3.4 E 4
900	4.1 E 4
925	5.2 E 4

While hardness shows a general proportionality to carbon content in case hardening, there can be significant deviation from this rule. For carbon contents above 0.7 % e.g., the martensitic start (Ms) temperature declines sharply with carbon content. A decline in the martensitic transformation temperature promotes the retention of austenite during quenching. Since austenite is softer than martensite, retention of austenite causes a softening of the hardened surface. At the atomic level, some carbon would have been dissolved in retained austenite instead of contributing to the supersaturated martensitic structure. The hardness profile of a carburized steel tends to exhibit a maximum hardness below the surface with a low surface hardness which is usually indicative of an excess of retained austenite. As an example of austenite formation and its consequences, the hardness and carbon distributions of nickel steel are shown in Figure 7.17.

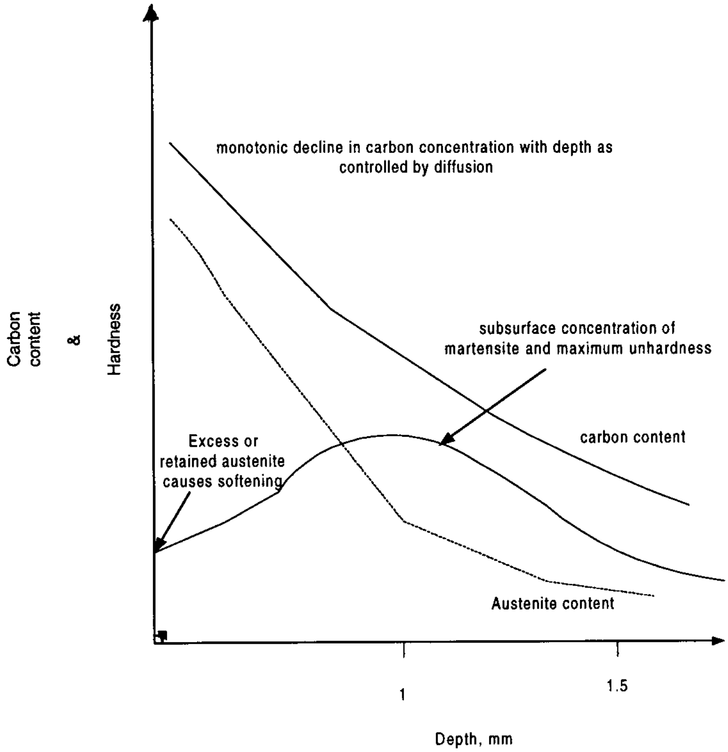


Figure 7.17. Schematic of hardness and carbon content profiles of carburized Ni steel (Adapted from *Thelning*).

It can be seen that the surface hardness is much lesser than the peak hardness, which is located approximately 1.0 mm below the surface. A schematic diagram of the metallurgical cross section of steel is shown in Figure 7.18, the surface concentration of austenite can be seen as a white phase.

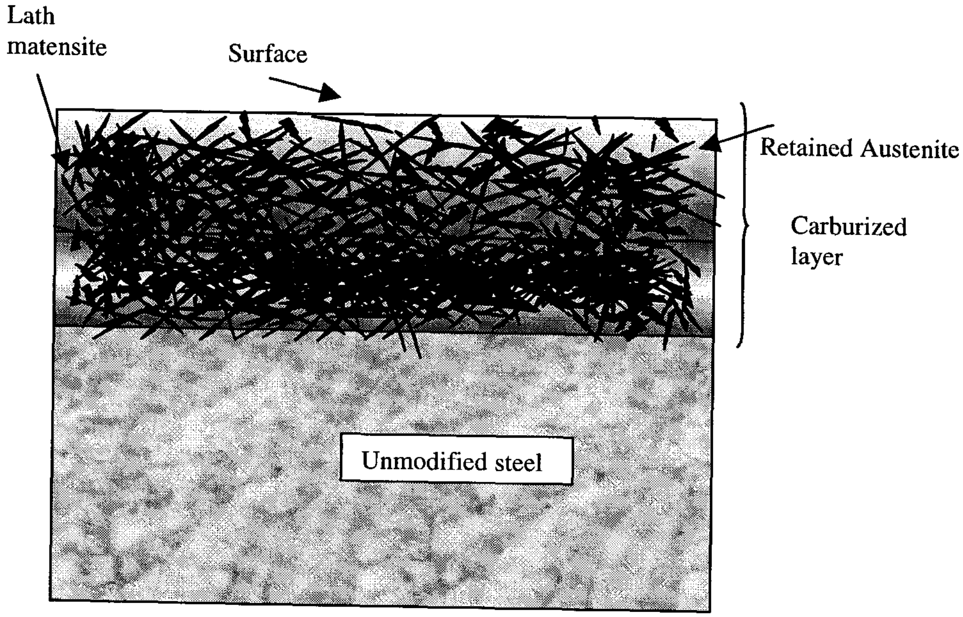


Figure 7.18. Schematic microstructure of a carburized steel showing an excess of retained austenite close to the surface.

Reduced hardness close to the surface is undesirable for many applications and remedies to this problem include: (i) quenching to subzero temperatures to ensure complete martensitic transformation (ii) quenching from a lower temperature than the carburization temperature to encourage the formation of carbides which consume the excess carbon.

Problems of austenite retention ensure that alloy steels are rarely directly quenched to room temperature. Instead, these steels are first quenched to a temperature slightly lower than the carburization temperature before a second quench to room temperature. Retained austenite is not entirely destructive as it is tougher than martensite. A basic question for any carburization application is to decide how much retained austenite can be tolerated. It is found from experience that the maximum hardness can be located very close to the surface if the carbon potential of the medium is close to the optimum and not too excessive to cause a drastic fall in the martensite start temperature. The optimal carbon content in the case (carburized layer) varies with the type of steel. For a nickel alloyed steel (0.5 - 3% Ni), 0.65 - 0.7% C is appropriate, while for a Cr-Mn steel, 0.7% C is preferable. A higher carbon content of 1.0% C is however, needed for a Cr-Mo steel. The optimum carbon content is strongly influenced by the presence of strong carbide formers such as chromium.

The brittleness of marteniste precludes it from most applications except for wear resistance. Tempering of case-hardened steels is routinely practiced as a prudent measure. Tempering is usually performed at 200-350°C except when a hardness of 60 HRC is required which necessitate a tempering temperature of less than 180°C. Tempering below 160°C should be avoided since the material becomes susceptible to grinding cracks if the component needs to be surface finished.

Carbonitriding

Carbonitriding is usually understood to be a modified carburizing process with the addition of ammonia gas into the furnace chamber. The process is normally performed at 800- 900°C with 3 - 8% ammonia content, which leads to a nitrogen content of approximately 0.4% at the surface of the component. Nitrogen atoms and ions diffuse through the steel when ammonia dissociates at the surface of the steel substrate. Dissociation is the decomposition of any chemical species into its elements or smaller molecules without reaction with any other chemical species. Dissociation of ammonia releases both hydrogen and nitrogen but it is found that there is insufficient hydrogen penetration of the steel to cause any embrittlement. The mechanism of ammonia decomposition and nitrogen absorption is illustrated schematically in Figure 7.19.

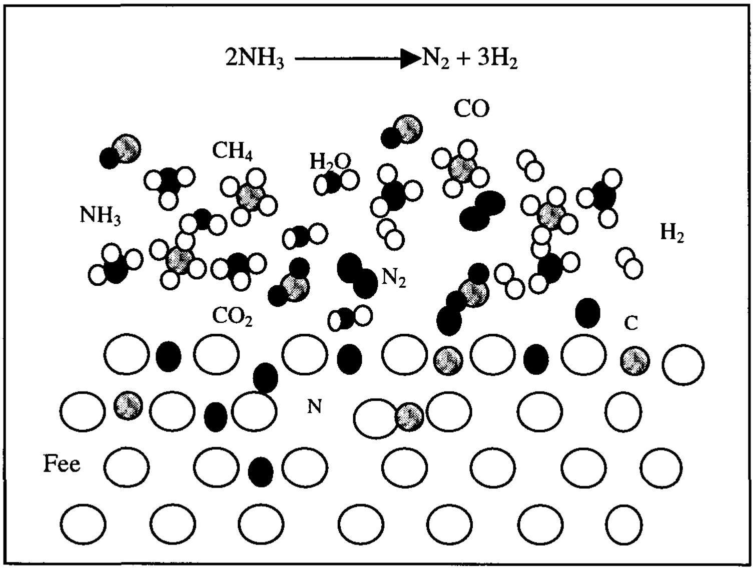


Figure 7.19. Dissociation of ammonia and penetration of steel by nitrogen in the carbonitriding process.

Carbonitriding enables both carbon and nitrogen to be simultaneously absorbed into the substrate to form complex carbonitrides, which enhance the hardness of the carburized layer. The combined enrichment of steel by carbon and nitrogen can produce a greater increase in hardening than carbon alone.

The mechanics of carbonitriding can be modelled in terms of the diffusion of nitrogen and carbon atoms into the steel substrate where the metallurgical transformation from ferrite to austenite influences the rate of diffusion. Ferrite has a b.c.c. crystal structure while austenite has a f.c.c. structure. The f.c.c. structure contains a large central void in the unit cell which facilitates diffusion of solute atoms. The diffusion constants of nitrogen and carbon in ferrite and austenite are listed in Table 7.2. Diffusivity is governed by the Arrhennius equation of reaction rates where a pre-exponential constant or frequency factor and activation energy are the independent variables. The precise form of the Arrhennius equation for diffusion is:

$$D = Ff*\exp(-AA/(R*T)) \qquad \{7.8\}$$

where D is the diffusion coefficient [m^2/s] , Ff is the frequency factor [m^2/s], AA is the activation energy [J/mole], R is the universal gas constant [J/(mole*K)] and T is the absolute temperature [K]. A 'mole' is the defined as the mass of a substance where the number of kilograms is equal to the nominal atomic or molecular weight.

Table 7.2. Diffusion constants of carbon and nitrogen in ferrite and austenite.

Phases	Frequency factor	Activation
energy		
	[m^2/s *E-6]	[MJ/mole]
Carbon/ferrite	0.39	80.3
Nitrogen/ferrite	0.49	76.8
Carbon/austenite	67	156.8
Nitrogen/austenite	91	168.4

Application of {7.8} reveals the relative speed of diffusion of nitrogen and carbon in ferrite and austenite. Calculations based on {7.8} and the data from Table 7.2 give the following values for diffusion coefficients of carbon at 800°C and 900°C: 1.548896E-12 m^2/s and 6.933044E-12 m^2/s. For nitrogen, the corresponding diffusion coefficients are: 5.728974E-13 m^2/s and 2.865048E-12 m^2/s. The ratio of carbon diffusion coefficient to nitrogen diffusion coefficient at 800°C is 2.703618 and at 900°C, the ratio is 2.41987.

Nitrogen diffuses slower than carbon in austenite and ferrite at the high temperatures typical of carbonitriding but nitrogen diffuses faster than carbon at lower temperatures. The change in relative speeds of diffusion affects the related process of nitrocarburizing. According to the theory of diffusion, there should be less penetration by nitrogen into the substrate than for carbon. In practice, it is found that nitrogen and carbon show approximately equal penetration without the formation of an exclusively carbide layer below a carbonitride layer.

An example of carbon and nitrogen distribution in a mild steel carbonitrided at 860°C is shown in Figure 7.20.

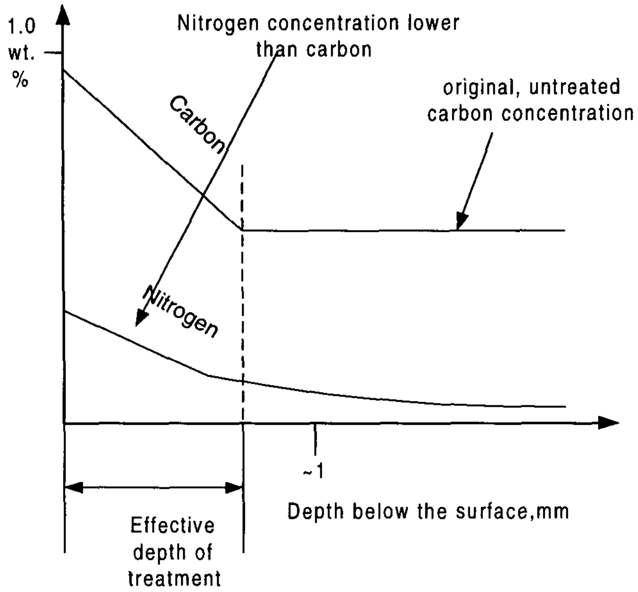


Figure 7.20 Carbon and nitrogen distribution versus depth in mild steel carbonitrided at 860°C (Adapted from *Thelning*).

The concentrations of carbon and nitrogen decline uniformly with depth to reach the residual concentration in steel. This type of carbon and nitrogen distribution is characteristic of diffusion. The carbon is present in greater concentration than nitrogen; this feature also confirms the controlling role of diffusion.

The relative concentrations of carbonaceous and nitrogeneous gases (i.e. ammonia) have a strong influence on the composition of the carbonitrided layer. Carbon contents in the steel in the range of 0.4 to 1.0% have little effect on nitride formation. Nitrogen concentration in the steel increases with the content of ammonia in the furnace chamber but declines with temperature. Experimental evidence for the effect of temperature and ammonia content is presented in Figure 7.21 where a schematic graph of the nitrogen content of a thin foil after carbonitriding versus temperature and ammonia content is shown. A thin foil is used for testing because near-uniform nitrogen concentration in the foil is ensured by efficient diffusion of nitrogen into the steel foil.

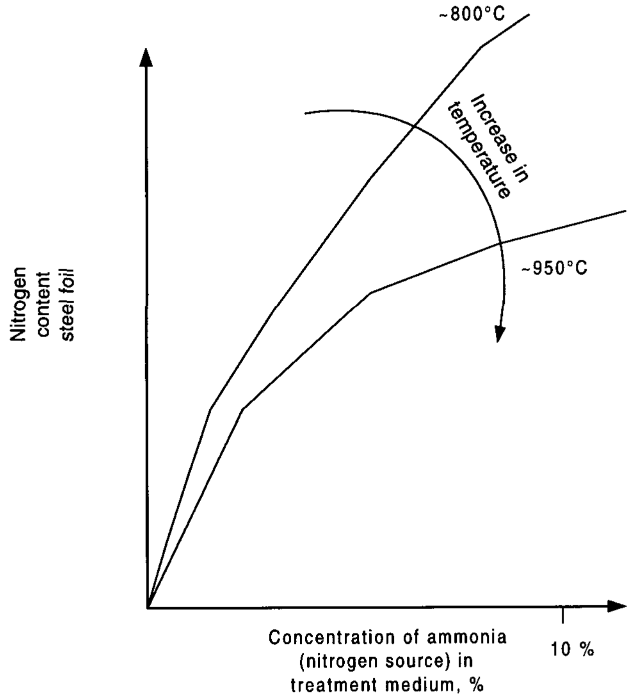


Figure 7.21 Nitrogen content of steel foil as a function of temperature and ammonia content in the furnace gas after carbonitriding (Adapted from *Thelning*).

A nitrogen content in treated steel of more than 0.2 wt %. increases the hardenability of the carburized layer. The minimum temperature for austenite survival and the start temperature for martensite formation are also reduced when nitrogen content is above 0.2 wt %. The declines in austenite survival temperature and martensite start temperature are notable for alloy steels. A major effect of nitrogen is to generate a larger amount of retained austenite in carbonitrided steels as compared to carburized steels. The influence of nitrogen content on the austenite survival temperature (often referred to as A1) and the martensite start temperature (Ms) of a 0.8% C eutectoid steel is displayed in Table 7.3.

Table 7.3. Effect of nitrogen content on Austenite survival temperature (A1) and Martensite start temperature (Ms) of a eutectoid steel (Adapted from Thelning Karl Erik).

Nitrogen [wt. %]	A1 [°C]	Ms [°C]
0.00	720	205
0.39	682	154
0.66	670	108

Nitriding

Nitriding is the enrichment of the surface layers of steel by nitrogen alone as opposed to carbon and nitrogen. Nitriding is practiced at lower temperatures than carbonnitriding, typically below 590°C, which is the eutectoid temperature of the Fe-N phase system. Typical treatment temperatures range from 510 to 570°C and at this level of temperature, the solid solubility of nitrogen in steel is only 0.1 wt. %. Despite the lower level of nitrogen content than carbonitriding, a substantial hardening effect can still be achieved. In industry, nitriding is performed by immersion in a salt-bath, heating in ammonia gas or using the more advanced technique of plasma nitriding. The salt-bath technique relies on molten cyanide salts, which release toxic fumes and are hazardous to dispose of. Gas nitriding uses ammonia in the same manner as for carbonitriding. Plasma nitriding is gaining popularity as an advanced clean technique that applies vacuum technology in a similar manner to PVD and CVD coatings desribed in Chapter 6. Plasma nitriding is described in section 7.3. There is also an obsolescent process, powder nitriding, which involves the embedding of a component in heated nitride salts. Solid-state diffusion at high temperatures is then exploited to form a nitride enriched surface layer in the component. This method is slow and not suited for the formation of nitride layers with precisely specified properties. There is usually some variation in the properties of the nitrided layer because the transfer of nitride by solid-state diffusion is critically dependent on the quality of contact between the nitride salts and the component. Since powder nitriding is now rarely if ever used, it shall not be discussed further.

The application of nitriding to form a hardened layer on steels can be interpreted in terms of the Fe-N phase system, which is shown in Figure 7.22.

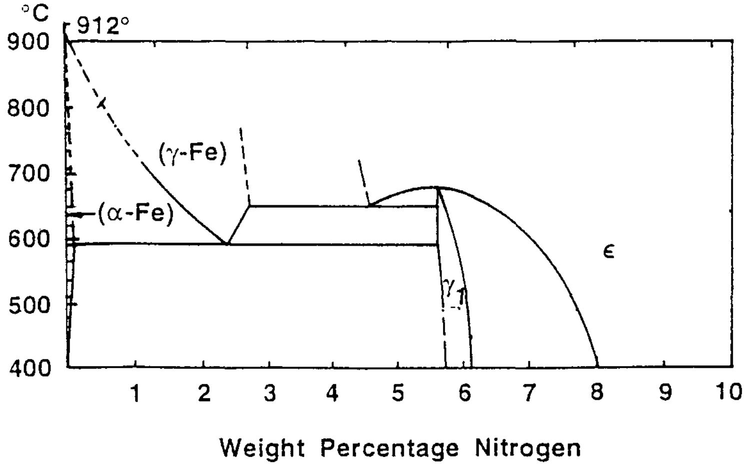


Figure 7.22. The Fe-N binary phase diagram (Adapted from *M. Hansen*).

The Fe-N phase system contains 4 major phases, which have distinct crystalline structures and metallurgical properties. Listed in rising concentration of nitrogen, the phases are (i) α, (ii)γ, (iii) ε, (iv) ζ which correspond to (i) nitrogen dissolved in metallic iron, (ii) the nitride Fe_4N, (iii) the non-stoichiometric nitride Fe_2N_{1-x} and (iv) the nitride Fe_2N. Non-stoichiometric means that the material does not possess an exact chemical formula but has instead variable proportions of chemical elements. All these compounds have different crystal structures, which influence the characteristics of nitride films. The crystalline structure and nitrogen content of each phase is listed below in Table 7.4

Table 7.4 Crystalline phase and nitrogen contents of nitrides.

Phase	Crystal structure	Weight%N	Atom% N	Ratio N to Fe
ζ-Fe2N	Orthorhombic	11.14	33.3	1 : 2
ε-Fe2N1-x ~ 4.54	Close packed hexagonal	4.5 - 11.0	18.0 - 32.0	1: 2.03
γ-Fe4N	Face-centred cubic	5.9	22.0	1 : 4

The ε-phase is considered to be resistant to scuffing (galling) while the γ-phase is thought to resist delamination or other forms of fatigue-based sliding wear. The ζ-phase is not known to serve any useful purpose.

During the nitriding process, nitrogen not only penetrates the steel but also forms a superficial layer of nitride by gaseous corrosion as described in Chapter 3. This superficial layer is usually too brittle for mechanical applications. Hence, it is typically removed by grinding. This outer layer is often referred to as the 'white layer' because it cannot be etched in standard nital acids and appears white when viewed under the optical microscope. The white layer is typically 25 μm thick and is composed of a poly-phase of γ and ε nitrides. As shown in Table 7.4, these phases have different crystalline structures, which induce intense localized strain fields in the white layer. These strains contribute to severe brittleness of the white layer, which is prone to micro-cracking. The white layer is also called the 'compound layer' because it is basically composed of compounds of iron and nitrogen.

The beneficial effect of nitriding for steels is generated by the formation of small needle-shaped inclusions of Fe_4N in the ferritic (iron) substrate below the white layer. The region of the substrate below the white layer is termed the 'diffusion layer' as this is where the nitrogen diffuses through a metal rather than forming a separate nitride layer. A schematic diagram of the structure of the white and diffusion layers are shown in Figure 7.23.

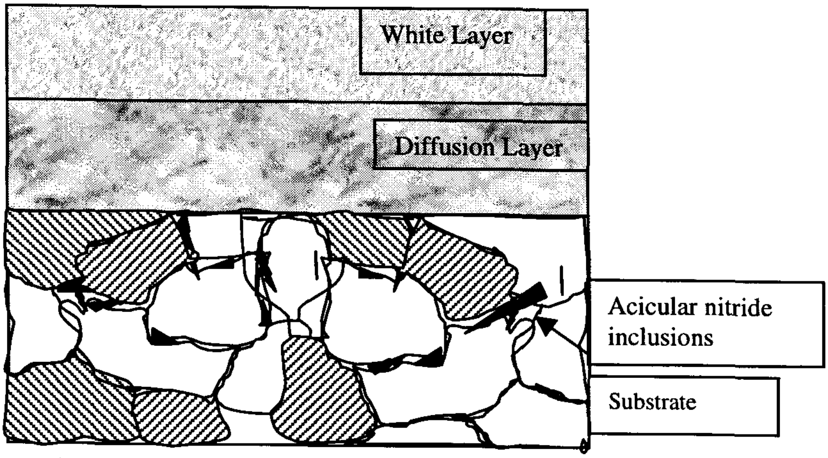


Figure 7.23. The structure and formation of the nitride coating on steel.

The alloy content of a steel strongly influences the effect of nitrogen in the diffusion layer since most alloying elements are more chemically active than iron. Where 'nitride formers' such as Al, Ti, Cr, Mo and V are present, the formation of finely dispersed alloy nitride precipitates will be promoted. These nitride precipitates distort the ferrite crystal lattice of the steel substrate and serve as effective barriers against dislocation movement. The steel substrate is thereby strengthened against plastic deformation. The diffusion zone is strengthened not only by the solid solution of nitrogen in ferrite but also by the precipitation of finely dispersed alloy nitrides. The effect of nitride formers (Al, Ti, etc.) on the hardness of 0.35% C, 0.3%Si, 0.7% Mn steel is shown in Figure 7.24.

For similar reasons to carburization, the effect of nitriding on a steel is usually evaluated in an industrial context by the depth of hardening rather than the depth of nitrogen penetration. The efficiency of nitriding is usually quantified in terms of a specified hardness below the surface. Hardness reaches a maximum value at the surface to decline to the non-nitrided value within about 1 mm depth. After very long periods of nitriding, the maximum hardness may be found at a small depth below the surface but the surface hardness will still be very high. The distribution of hardness versus depth with nitriding time is shown schematically in Figure 7.25 for an alloy steel.

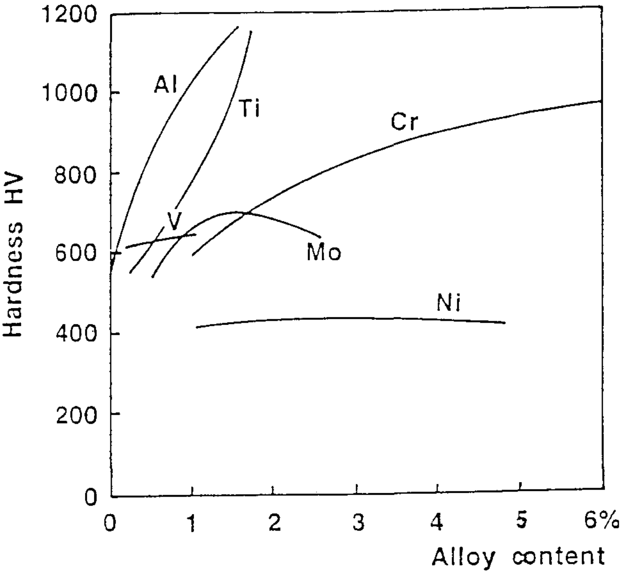


Figure 7.24. Effect of alloying elements on the hardness of a nitrided steel (Courtesy of *Thelning* and Bofors Ltd.).

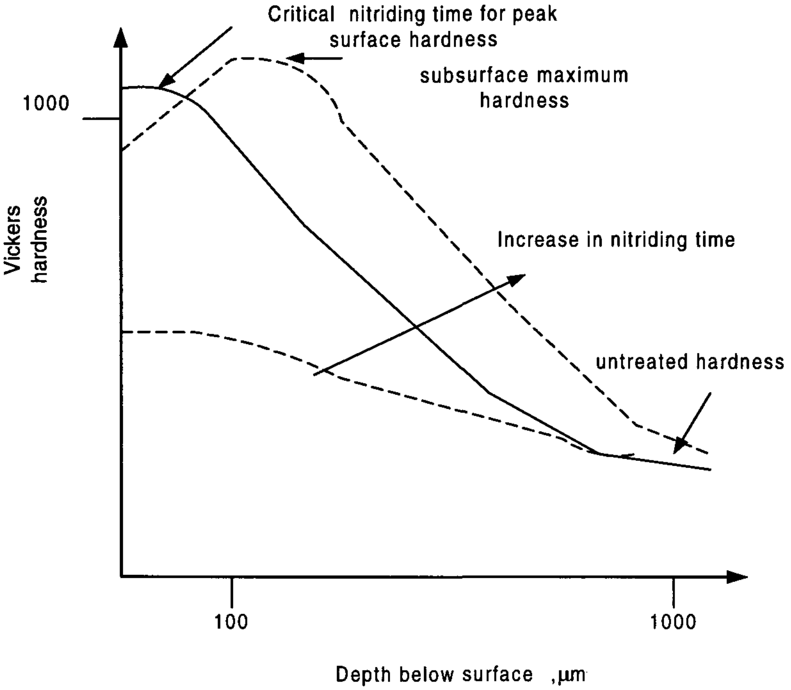


Figure 7.25. Schematic illustration of hardness distribution of nitrided alloy steel for different nitriding periods (Adapted from *Thelning*).

A very high hardness of approximately 1000 Hv can be achieved which is beyond the capability of carburization.

An increase in treatment time or temperature results in coarsening (excessive size) of the alloy nitride precipitates causing a drop in the surface hardness. If the extra cost of prolonged treatment times is acceptable, it is advantageous to nitride at lower temperatures. Gas nitriding is seldom performed at temperatures below 500°C. The advantage of plasma nitriding is that far lower temperatures are practicable.

Nitrocarburization

Nitrocarburization is based on gas or plasma nitriding and involves the addition of small amounts of carbonaceous gases to the process. These gases may be hydrocarbons or inorganic gases such as carbon monoxide. The surface layer produced remains predominantly nitrides. Nitrocarburization can be seen as complimentary to carbonitriding where carbon is more significant to the process than nitrogen. Nitrocarburization is performed at higher temperatures than nitriding, 570°C is commonly selected for industrial applications. The prime effect of carbon addition is to bias nitriding to the formation of the ε-phase

Nitrogen diffusion in iron is more rapid than carbon diffusion at low (or comparatively lower) temperatures. Diffusion coefficients for nitrogen and carbon in ferrite for the temperature range 300 to 700°C are shown in Table 7.5.

Table 7.5. Diffusion coefficients of nitrogen and carbon in ferrite (Adapted from Thelning Karl Erik).

Temperature [°C] [m^2/s].	Diffusion coefficient of nitrogen, [m^2/s]	Diffusion coefficient of carbon,
300	5.3 E-14	4.3 E-14
500	3.6 E-12	4.1 E-12
700	4.4 E-11	6.1 E-11

At 300°C, the diffusion coefficient of nitrogen is greater than that of carbon. The diffusion coefficients are equal at a temperature slightly less than 500°C. Thus the diffusion coefficients control the penetration of carbon and nitrogen into the steel substrate. The subsurface concentration of nitrogen after diffusion at a temperature close to 500°C will be approximately equal to the concentration of carbon if there are similar concentrations of nitrogen and carbon at the surface of the substrate. At higher temperatures typical of carburization, the subsurface concentration of carbon will be higher.

Boronizing

Boronizing is usually performed by the pack cementation method where the metal component is held in contact with a solid boron-rich material for a long period of time at high temperature. Transfer of boron from its source to the metal substrate is primarily by solid state diffusion at high temperatures. The source of boron is typically boron carbide, which is one of the more readily available forms of boron. Boron is rare and more expensive than either carbon or nitrogen and so the choice of boron compound significantly affects the cost of the process. An activator, e.g. ammonium difluoride is mixed with the boron carbide to ensure rapid boronization. An inert packing substance then dilutes the boron carbide, ammonium difluoride mixture, which can be silicon carbide. The proportions of boron carbide, ammonium difluoride and silicon carbide are 5%, 2% and 93%. Treatment temperature is close to 900°C and this temperature is maintained for about 6 hours to produce a coating of approximately 50 μm thickness. This surface layer produced by boronization typically has a hardness of 700 to 750 Hv after tempering and the hardening phases are FeB and Fe_2B; with the latter exhibiting a strong hardening effect without excessive embrittlment. After boronizing, the metal component must be carefully tempered to suppress the brittleness of the boride phases. The carbon present in the boron carbide and silicon carbide does not usually cause excessive carburization of the metal component. Boronizing is observed to improve wear resistance of components provided that brittleness is carefully controlled and that the wear process does not involve severe impact.

7.4 Advanced surface modification technologies (ASMT)

Surface modification technologies can be defined as recently developed methods of producing integral coatings where a scientific design of the coating is attempted. Older techniques such as carburization follow traditional practices and the resulting coatings cannot be easily modified for specific applications. In Surface Modification Technology, the objective is to refine coating techniques to allow the production of coatings, which closely follow the requirements of the application. Vacuum-based techniques such as plasma nitriding or ion implantation are used for SMT. The laser and electron beam are also used for ASMT because of their comparative ease of control.

7.4.1 Plasma nitriding and plasma carburization

Plasma nitriding and the related process of plasma carburization are established coating techniques that are used to generate moderately thin coatings of hard compound layers on metal components. With plasma nitriding, the thickness of the compound layer is typically in the range of 1 to 5 µm and the coating contains also the original substrate material with alloy nitrides or carbides. These nitrides and carbides are present as a fine dispersion of precipitates, which strengthen the metallic substrate considerably.

Plasma nitriding involves an electric discharge between an anodic vacuum chamber and a cathodic component in the presence of gaseous nitrogen. The vacuum technology is the same as described in Chapter 6. Plasma carburization is very similar to nitriding except that a carbonaceous gas, e.g. methane or propane is substituted for the nitrogen. The component is deliberately given a cathodic voltage (often referred to as 'bias') by connecting it to a power supply. After an initial pumping out of atmospheric gases, the chamber is filled with nitrogenous gas to a pressure of approximately 1 - 10 Torr (which is equivalent to 0.1 to 1 kPa). The pressure of the nitrogenous gas is critical to the process as it determines the nature of the electrical discharge between the component and the vacuum chamber. The plasma nitriding apparatus is illustrated schematically in Figure 7.26.

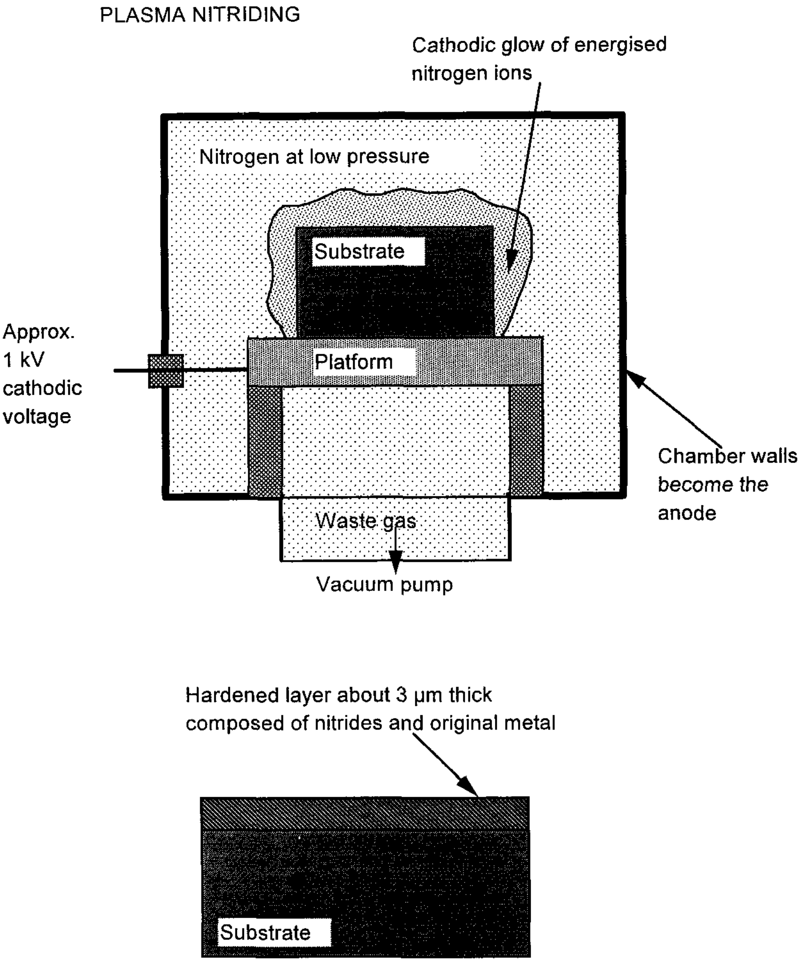


Figure 7.26. Principle of operation of plasma nitriding apparatus and structure of coating produced.

The conductivity of electricity by any gas is controlled by its pressure. When the pressure is very low then it is evident that there is insufficient gas to conduct electricity. Conversely when the pressure is high, approximately atmospheric pressure or higher, the high frequency of collisions between gas molecules disrupts the flow of electricity unless an arc forms. An arc is a localized flow of ionized gas or plasma, which conducts electricity and is distinct from the bulk of the surrounding gas. At intermediate pressures, it is found that the gas is more conductive and a stable flow of current by the bulk of the gas is possible. The flow of current generates a glow of ionized gas around the component, this phenomenon of light generation is termed ' glow discharge'.

Current flow characteristics of nitrogen versus pressure and the nature of arcing are illustrated schematically in Figure 7.27.

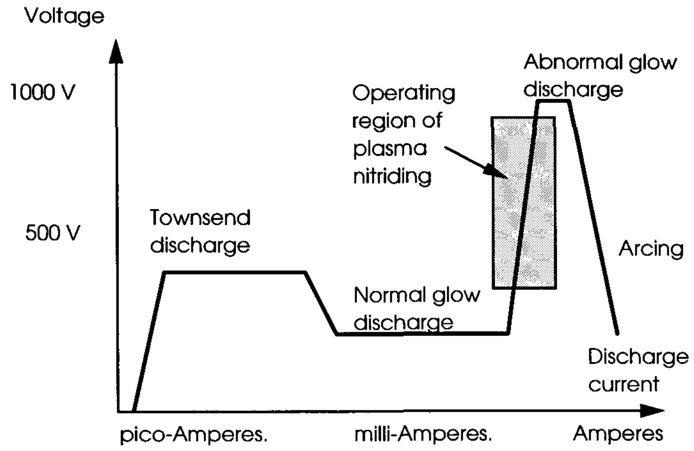


Figure 7.27. The electrical discharge characteristic of nitrogen.

Plasma nitriding depends on what is called 'abnormal glow discharge' which is the highest level of current flow and gas pressure that still permits uniform current flow without arcing.

The glow discharge is associated with a high-energy state of ionized nitrogen, which is able to penetrate the metal substrate rapidly at temperatures much lower than by simple chemical reaction between nitrogen and solid metal. Once the nitrogen has entered the surface of the metal it progresses to the subsurface by diffusion and as a result the depth of coating attained is proportional to the square root of coating time. The coating process is performed at temperatures in the range of 500°C to 800°C depending on the metal to be coated. Irons and steels are preferably coated at the lower range of temperatures to prevent excessive degradation of microstructure. The required duration of coating to achieve a useful coating thickness is normally several hours.

A high hardness of coating is obtained, e.g., 1000 Hv, which enhances wear resistance. The critical advantage of plasma nitriding compared to gas nitriding is that thin layers of nitrides with a very fine grain size can be formed on the surface of the component. The thickness of a plasma nitrided coating is typically 3 micrometres with a graded interface with the substrate that is approximately 100 mm thick. The nitride coatings formed by plasma nitriding possess usefully high wear resistance and are not limited by the brittleness of coatings formed by gas nitriding. Both steels and non-ferrous metals such as Ni-based and Ti-based alloys have been plasma nitrided industrially.

The coatings induce surface compressive stresses by the insertion of nitrogen and thus improve resistance to mechanical fatigue. With stainless steels, the coatings do not give protection against corrosion. This is due to the formation of chromium nitrides. When dissolved chromium in steel is converted to chromium nitride, the formation of a protective chromium oxide layer on the steel surface is suppressed and rapid aqueous corrosion becomes possible.

Plasma carburization

Plasma carburization uses the same apparatus and operating principle as plasma nitriding. The key difference between plasma nitriding and carburization is that methane or propane is used instead of nitrogen. Methane and propane are hydrocarbons and act as the carbon source for carburization. The hydrogen that is also released upon decomposition of methane and propane is dispersed as a waste product. Plasma carburization generates a supply of ionized carbon species at the substrate surface. The high energy and ionization of the plasma accelerates the transfer of carbon from the gas phase to the steel surface. This prevents the slow reaction kinetics of hydrocarbon decomposition in the gas phase from limiting the rate of carbon supply.

Plasma carburization is typically performed at voltages ranging from 300 to 500 V and a gas pressure between 260 to 800 Pa. Operating temperatures range from 750 to 1050 °C, which is significantly lower than the conventional carburizing temperatures of 900°C to 1095°C. The main advantages of plasma carburization compared to conventional carburization are speed of treatment as carbon saturation is attained in a few minutes and lower temperatures. Alloy content does not affect the time to achieve carburization and uniform case depth can be obtained on even complex component shapes. Excessive carbide formation on the edges of thin sections can also be suppressed by control of the rate of carburization.

7.4.2 Surface alloying by laser and electron beam

The technology used for surface hardening by lasers or electron beams can be greatly extended in application by the pre-coating of the substrate with a layer of alloying metal. A basic example of this practice is the prior deposition of chromium on steel. A surface layer of stainless steel can then be formed on the steel substrate to protect it from corrosion. The term 'surface alloying' means to form a surface layer that is an alloy of the substrate metal and a layer of metal or non-metal deposited on the substrate. Other examples of alloying are carbon on steel to form fully martensitic high-carbon steel or nickel on aluminum to form nickel aluminide for wear resistance. If a coating of the alloying metal is difficult to generate, the alloying metal can be used in powder form. The powder is placed in the path of the laser or electron beam. The principle of surface alloying is illustrated schematically in Figure 7.28.

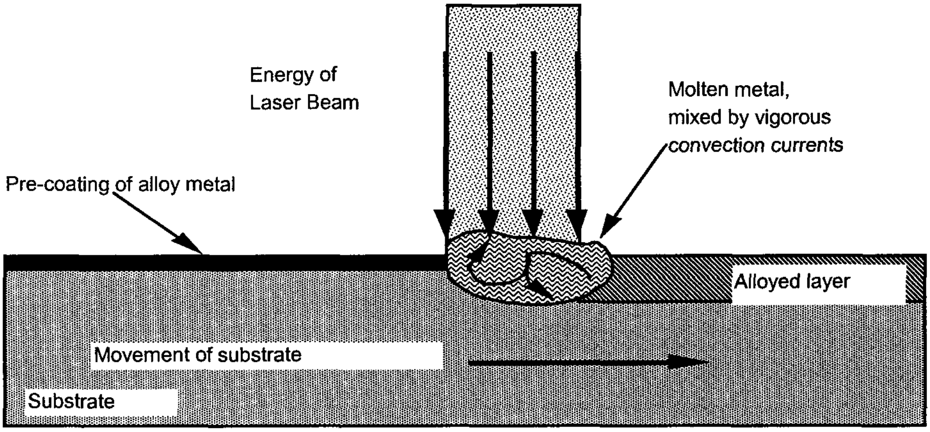


Figure 7.28. Principle of surface alloying of a metal by laser or electron beam.

Effective mixing depends on intense convection currents generated in the molten pool of metal that is formed by the laser or electron beam. These currents effectively mix the alloying element and substrate. In general though, the concentration of alloying element will be marginally higher at the surface of the alloyed layer than at the bottom of the alloyed layer.

The depth of alloyed layer corresponds to the depth of the melt pool and typically ranges from 100 to 500 μm. As with surface hardening, the surface roughness of the alloyed layer is considerable and usually requires grinding. A major limitation of the alloyed layer is the presence of shrinkage cracks particularly in brittle alloyed layers. It is essential that the coating material is not excessively reflective for laser surface alloying or else the laser beam will not adequately heat the substrate due to reflection.

7.4.3 Ion implantation

Ion implantation is related to ion-plating and Ion Beam Assisted Deposition (IBAD) but involves much higher energies of ionic bombardment. The objective of Ion implantation is not to directly deposit a coating as is the case with ion-plating or IBAD but to produce a modified surface layer.

When high-energy ions bombard upon a material, its crystalline lattice sustains damage in a manner similar to neutron bombardment described earlier. If the number of ions impinging at a substrate is sufficiently large, the crystalline structure of the outer surface layer eventually disintegrates. At a critical level of dislocation density inside the damaged crystal, the crystalline order ceases and an amorphous structure is formed. Ion-implantation is so called because ions of a different material from the substrate are forcibly inserted at some depth inside the substrate. The objective of ion-implantation is to form an amorphous layer on the surface of the substrate, which has superior, wear and corrosion resistance compared to untreated surfaces. The reasons for the improvement in such properties by ion implantation are not fully known. Various workers attributed this to the increase in hardness, amorphous structure and change in chemical composition. The number of ions implanted is much less than the number of substrate atoms in the amorphous layer. Hence, ion-implantation produces a modified surface layer and not a discrete coating. The number of ions implanted is typically in the range of 1.0 E18/m^2 and the number of atoms in the corresponding amorphous layer is approximately 1.0 E20/m^2 which gives a concentration of about 1% for the implanted material inside the surface layer. This concentration of 1% is applicable only when the implanted ions remain in the outermost layer of substrate atoms. Since the ions penetrate to a depth of 1 µm or more, which is equivalent to 5,000 atom layers (at an average spacing of 0.2 nm), the actual concentration is far less than 1%. In some applications, a higher dosage is required than 1.0 E18/m^2. It is found that with nitrogen ion implantation, a dosage of nitrogen ions of 4.0 E21/m^2 is required to give a significant effect on wear. This dosage of nitrogen ions is equivalent to a 1% concentration to a depth of 1 µm. The change in microstructure produced by ion implantation is shown schematically in Figure 7.29.

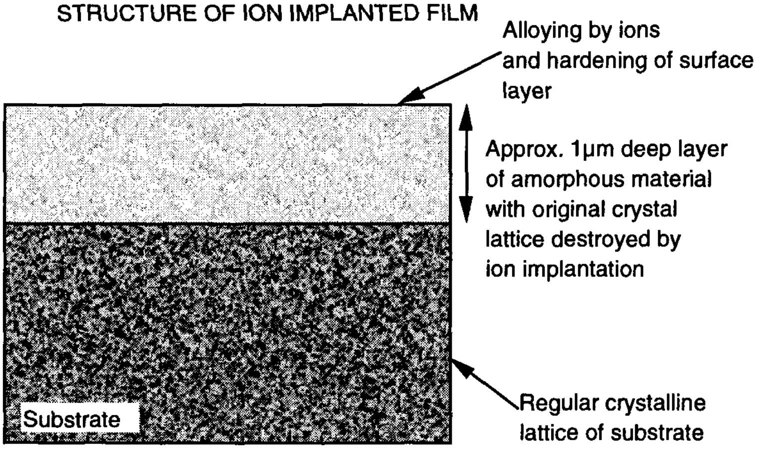


Figure 7.29. Schematic illustration of the effect of implantation.

An ion-implantation apparatus is costly and highly specialized. Such facility is usually held by specialist establishments, which operate the system for companies wishing to produce components with ion-implanted surfaces. For ion-implantation by a gas, e.g. nitrogen, exposing the gas to thermionic emission of electrons from a heated filament generates the nitrogen ions. Nitrogen ions are then filtered from the bulk of the nitrogen gas by electrostatic attraction through a small hole. Ions of the appropriate energy level are then extracted from the rest of the ions by passing the ion beam through a curved track inside a magnetic field. Only ions of the appropriate energy (speed) follow the curved path accurately to reach the substrate outside of the system. Electrostatic fields prior to impingement on the substrate further energize the ions. The peak energy of ions before impingement is typically 100 keV. The ion-implantation apparatus is illustrated schematically in Figure 7.30.

Small components can be held stationary under an ion beam that covers the entire component while larger components can be covered by a TV-raster system for the ion beam.

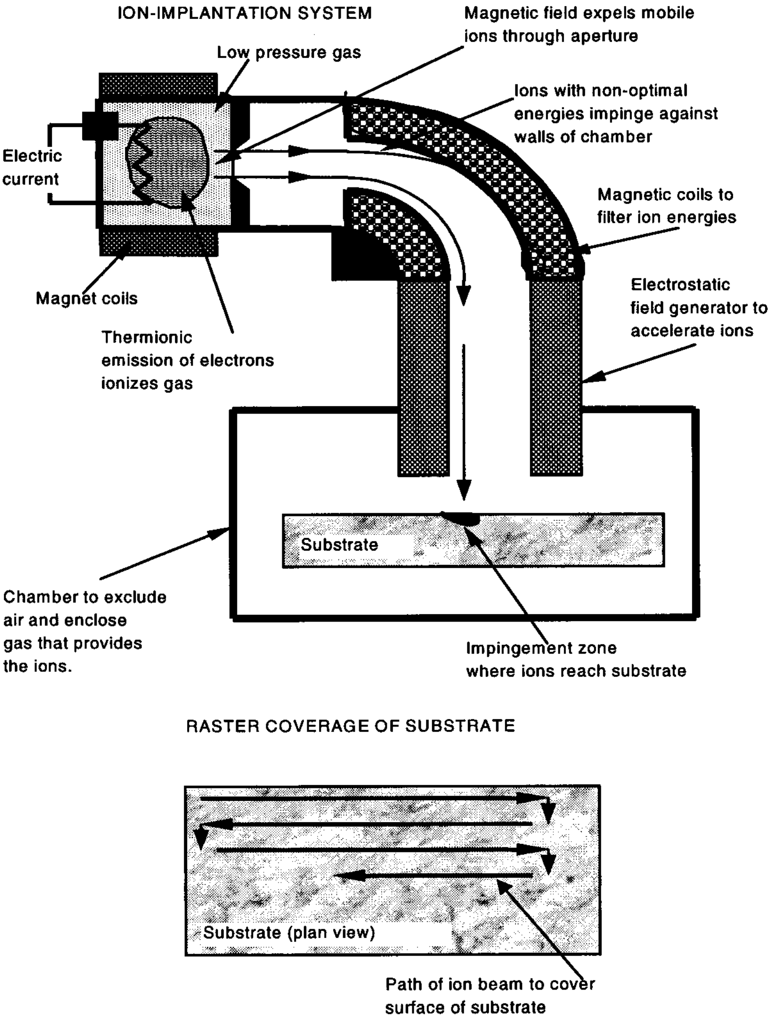


Figure 7.30. Ion-implantation apparatus and coating method.

The principal advantage of ion-implantation is that it is a clean and precise means of conferring superior properties on components with negligible distortion or other forms of degradation. A component can be made wear resistant with almost no change in dimensions and surface roughness and no toxicity problems with coating materials. A cheap material such as nitrogen is suitable for use in implantation. The main disadvantage of ion-implantation apart from the apparatus cost is that the depth of ion-implantation is limited to 1 μm approximately, which precludes its use from many severe wear and corrosion applications where larger wear and corrosion depths are inevitable. Multiple ion implantations with several implanted elements can increase the thickness of the modified surface layer, but only by a limited amount. Treatment time is also longer than plasma nitriding so that productivity of the ion-implantation system is limited. Ion-implantation is suitable for conferring wear resistance on electrical switch gear and corrosive-wear resistance to metal orthopaedic implants where only a minute amount of wear can be tolerated.

7.5 Summary

Integral coatings find many important applications in the coating of metals. Applications on other materials have so far been limited. In the future, it is foreseeable that integral coatings may be developed on non-metals such as ceramics. A wide range of technologies, ranging from the traditional and widely used to the new, which are just being adopted for general industrial application, forms integral coatings. The prime purpose of integral coatings is for surface hardening and improved wear resistance. There are also some applications for corrosion and fatigue resistance.

Irrespective of the coating technology used, all integral coatings share a common feature of a diffuse boundary with the substrate, which ensures strong bonding to the substrate. An integral coating almost always contains a mixture of substrate metal and coating material. This means that the affinity between substrate and coating for chemical reaction or formation of a hybrid crystallographic structure is very important. By contrast, discrete coatings can be deposited on a substrate with little regard to its composition.

RECOMMENDED READING

M.Hansen, "Constitution of Binary Alloys", McGraw-Hill, 1958.

Karl Erik Thelning, "Steel and its Heat Treatment-Bofors Handbook" London, Butterworths, 1984.

B. Bhushan and B.K. Gupta, Handbook of Tribology, Materials, Coatings and Surface Treatments, McGraw-Hill Inc., 1991.

C.J. McHargue, Ion Implantation in Metals and Ceramics, International Metals Reviews, Vol. 31, 1986, pp. 49-76.

CHAPTER 8

CHARACTERIZATION OF SURFACE COATINGS

8.1 Introduction

There are numerous techniques to characterize the bulk solids but the characterisation of surface coatings occupies a major portion of the available techniques. Since coatings are important for the control of degradation of materials most of the techniques dealt in this chapter refers to evaluation of surface coatings for specific applications. Providing answers to questions such as; what is the composition of the coating? How hard is the coating? are part of the characterisation process. Data specifying hardness, composition, thickness roughness and any other relevant parameter enable optimisation of coatings for specific usage conditions. For example, coatings resistant to abrasive wear require a minimum hardness, which is approximately equal to the hardness of the abrasive particles. Sufficient availability of hardness data on surface coatings is required before the coatings can be selected for their resistance to abrasive wear. Another example could be the measurement of chromium content in chromium alloy layers formed on steel for corrosion resistance. The suitability of a surface layer for corrosion resistance will be largely decided by whether the coating has minimum chromium content of approximately 10%.

Another aspect of coatings characterisation is quality control as coatings are vulnerable to defects to the same extent as that of bulk materials. Non-uniformity of coating thickness, impurities or undesirable crystalline phases, cracks, voids, substrate damage, residual stresses may have to be controlled to ensure reliable performance by the coatings.

The composition and structure of the coatings are established by using various physical and analytical techniques. Techniques range from the use rudimentary tool such as optical inspection to X-ray analysis. The special features of surface coatings as opposed to bulk materials have led to the modification of some analytical techniques or the development of special analytical techniques. The main objectives in surface analysis and coating characterisation are illustrated schematically in Figure 8.1.

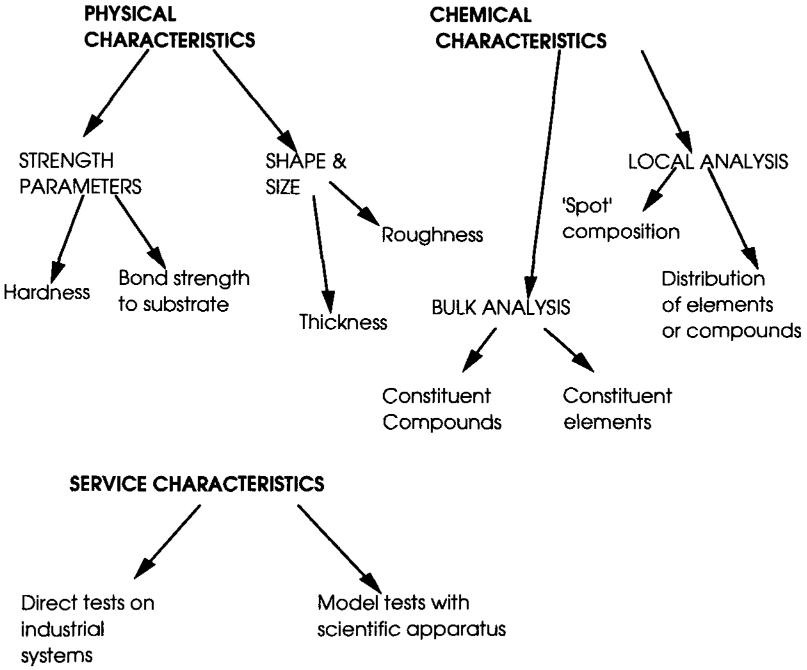


Figure 8.1 Objectives of surface analysis and coating characterisation.

The measurement technology serving each of the objectives illustrated in Figure 8.1 are described below together with the corresponding limitations in quality and practicality of measurement.

A further aspect of measurement relates to service characteristics rather than specific materials properties. It is usually not possible to accurately calculate wear or corrosion resistance from known materials properties. In most cases only very crude estimates are possible. Instead it is necessary to perform model wear or corrosion tests on purpose-built scientific apparatus or even perform tests on actual service equipment. The principles of testing of service characteristics are also discussed in this chapter.

8.1 Measurement of surface roughness and coating thickness

The geometrical parameters of a coating, i.e., its roughness and thickness exert a fundamental influence on the mechanical integrity of a coating. Accurate measurement of roughness and thickness are therefore critical to the evaluation of any coating. Coatings are produced in a range of thickness varying from thick coatings of approximately 1 mm thickness to very thin coatings that are about 10 nm in depth. A bar chart of typical thickness versus types of coating is shown in Figure 8.2.

The thickness of a coating has a strong influence on the choice of appropriate surface analytical technique and on the kind of information that needs to be collected. For very thick coatings, visible features such as cracks, hard phases and voids are important while for thin coatings microscopic and even atom-scale features such as missing atoms are critical.

8.1.1 Surface roughness

Surface roughness describes the morphological features on a real surface. Any real surface is not planar but covered with microscopic hills, valleys and even canyons (scratches). Surface roughness is a bulk measure of the average size of the hills and valleys. The scale of surface roughness is typically from 50 nm (very smooth surface) to 100 μm (extremely rough surface). This small size is the reason why surfaces appear smooth and planar (or spheroidal) to the naked eye when in fact the real shape is composed of a semi-infinite number of irregular forms.

Surface roughness is critical to the performance of wear-resistant and lubricant coatings, stress corrosion and fatigue life. Wear and friction are strongly affected by surface roughness and in many instances, an optimum roughness can be found which provides a minimum of wear and friction. In some applications of wear resistant coatings, e.g., pressing tools and moulds, a high quality of surface finish is required. For coatings intended to confer corrosion resistance excessive roughness can be a sign of poor quality and very severe surface roughness can generate local electro-chemical cell systems on the surface once the coating is immersed in an electrolyte.

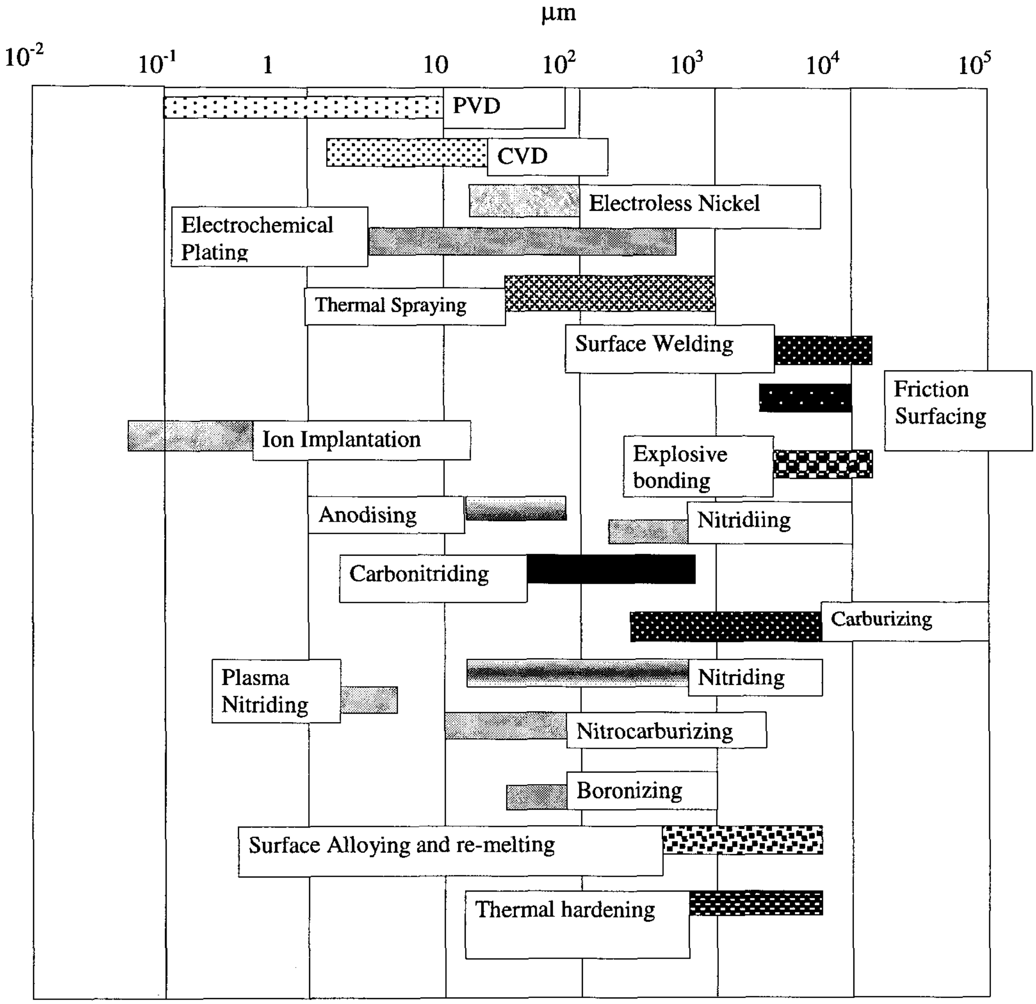


Figure 8.2. Thickness ranges of various coatings in current use (Adapted from *Roberts*).

There are two basic types of surface roughness measurement techniques: (i) The contact method which is based on a stylus traversing the surface, (ii) Non-contact methods based on phenomena such as the pattern of laser reflection from the rough surface.

Contact Method

In the contact method, a stylus traverses over a sample of the rough surface and any vertical movements of the stylus are recorded as a function of horizontal distance traversed. The vertical direction is defined here as normal to the plane of the surface and horizontal as planar to the surface since the vast majority of measurements are performed with this configuration of the test specimen. The contact method of surface roughness measurement is illustrated schematically in Figure 8.3.

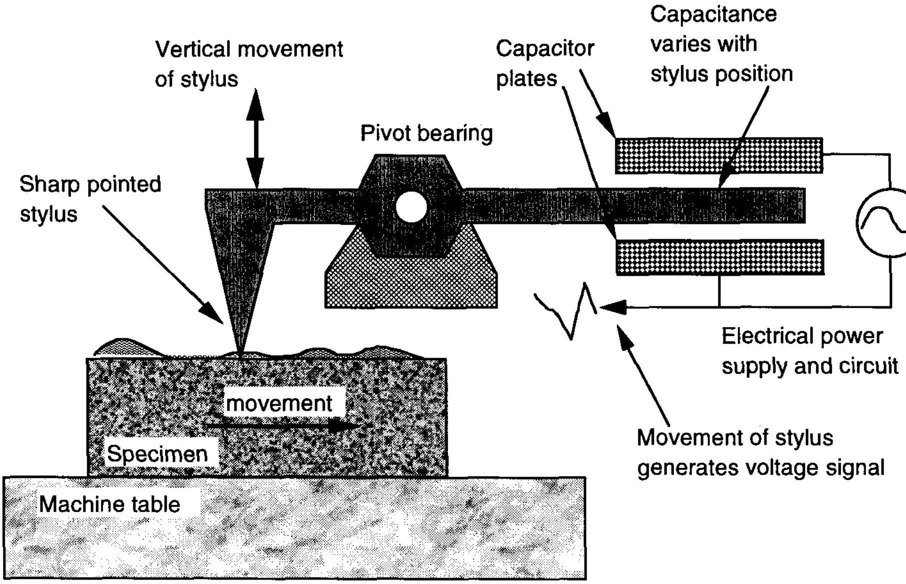


Figure 8.3. The contact method of surface roughness measurement.

The scale of horizontal movement is typically in the range of 1 to 100 mm and often many traverses of the stylus are performed to obtain a representative view of surface roughness. The stylus is driven by a precision screw thread while vertical movements are detected electrically. An enlarged view of surface features can be obtained by portraying adjacent passes of the stylus as a 3-dimensional image. This is useful for qualitative interpretation of roughness problems. For more quantitative work, the integral average of vertical movements as a function of horizontal distance is used as a measure of surface roughness. Contact methods of roughness measurement are widely used in laboratories and industry.

For the examination of surface features having an atomic order of size, the Atomic Force Microscope can be used. With this instrument, an extremely sharp stylus is used and extremely small horizontal movements obtained by using a piezoelectric table to drive the stylus. The piezoelectric effect is a phenomenon where the application of an electric voltage to certain materials causes a small change in their dimensions. A piezoelectric table is a device where small movements are generated by applying voltages to the piezoelectric table supports. The Atomic Force Microscope not only allows surface features of 1 nm or less to be observed but also provides some chemical data about the surface atomic structure. The most important restriction of the atomic force microscopy is that it cannot be applied to very rough surfaces because of the limited range of movement and sensitivity of the piezoelectric sensor. It is best suited for applications where the surface roughness is typically in the order of nanometers. The chemical data is derived from quantum-scale interactions between the stylus and test surface. An Atomic Force Microscope is illustrated schematically in Figure 8.4.

Non-contact methods

Contact methods are older than non-contact methods and provide a direct measure of roughness that is easier to analyse than data from non-contact methods. The reason for this is that the relationship between e.g. the reflection pattern of a laser and surface roughness amplitude is far less direct than the relationship between movement of a stylus and the size of the underlying feature. A major limitation of contact measurement is the slow speed of the stylus and the chance of damage to the surface by the stylus. The stylus must have a very sharp point to allow it to follow as much as possible of the surface detail yet the sharp point can easily cut the test surface. In many applications, a large area of possibly moving surface must be assessed quickly. An example of this type of application is the quality control of sheet metal as it emerges from the rollers. In these circumstances, non-contact methods become essential. The principle of surface roughness measurement by laser reflection is illustrated schematically in Figure 8.5

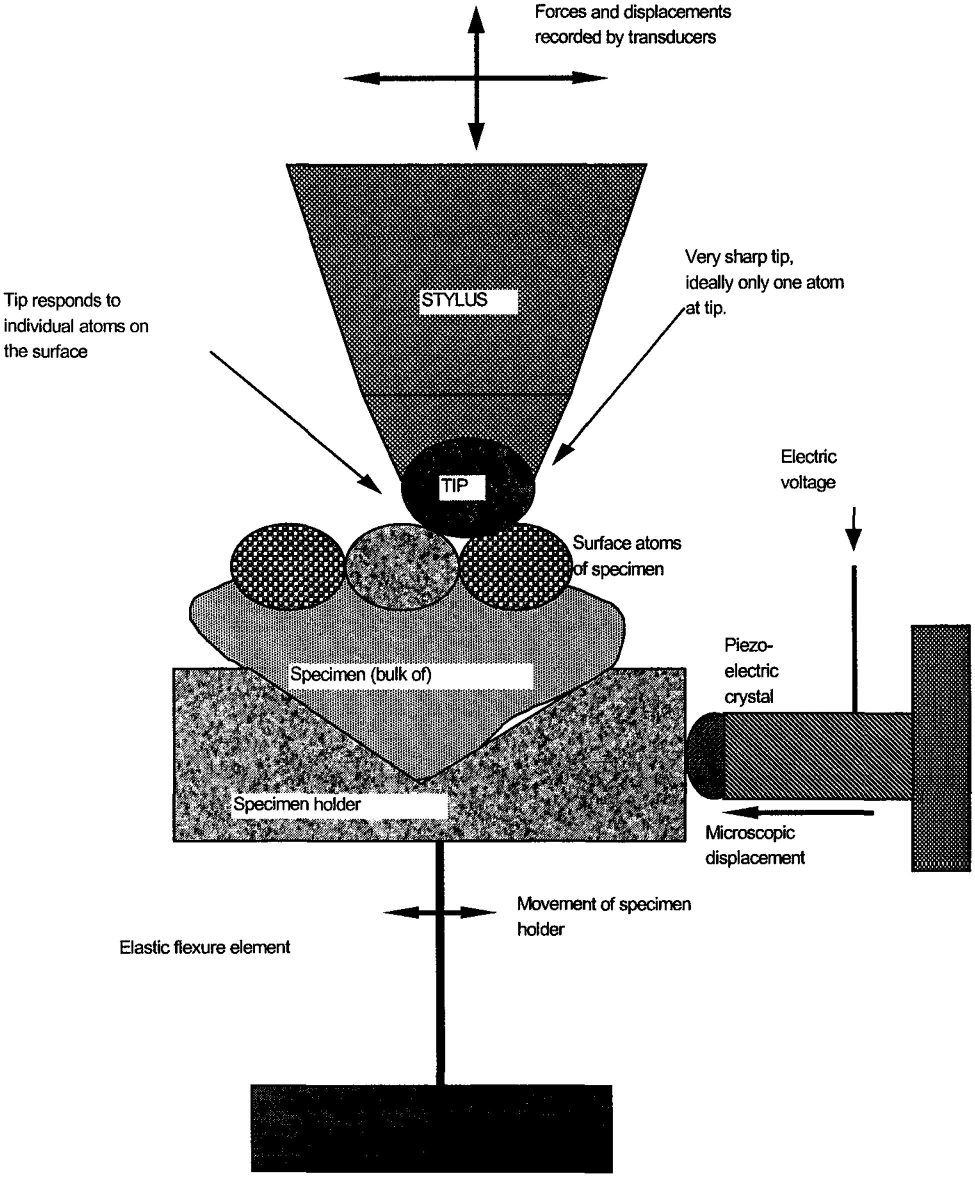


Figure 8.4. Application of atomic force microscope to measure atom-scale surface roughness.

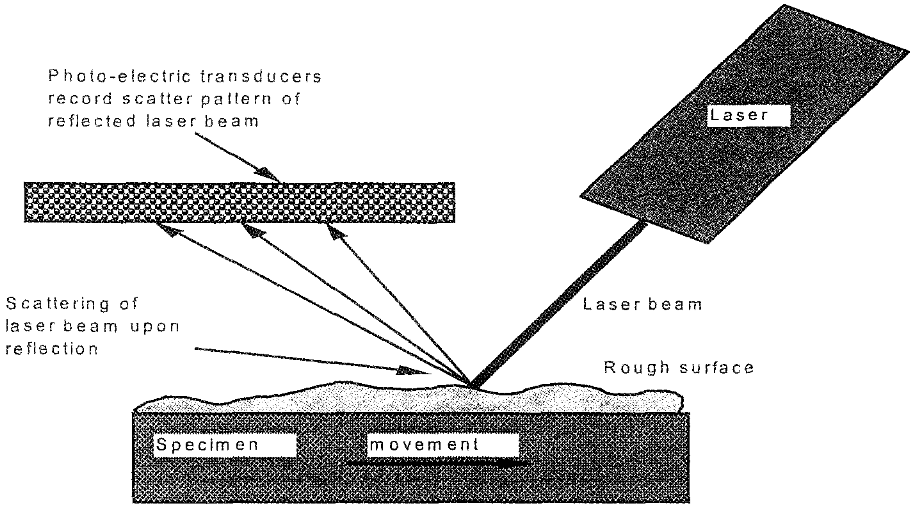


Figure 8.5. Measurement of surface roughness by laser reflection.

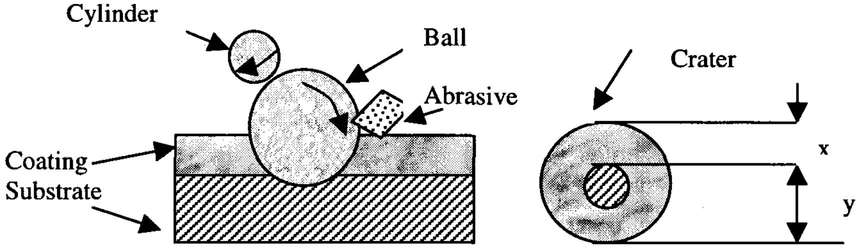
The critical advantage of non-contact methods is the possibility of continuous monitoring of the roughness of a moving surface or measurement of a remote surface. Changes in surface roughness values can be indicative of a change or imminent change in the wear rate. The greatest risk entailed with wear is a sudden rise in wear rate causing rapid destruction of machinery. A limitation of the laser reflection method is that the range of measurable surface roughness is less than the measurement range of a stylus. The roughness of very rough surfaces cannot be accurately measured because the reflected light no longer forms a distinct pattern and instead scattered diffuse reflection is observed. The minimum area required for measurement is larger for a laser than for a stylus. The laser beam has a diameter of approximately 1 mm while the tip of a stylus can have a radius as small as a few manometers. The resolution of a profilometer with stylus is usually limited by the precision of the X-Y table that moves the specimen.

Another recent development in the roughness measurement is the use of scanning confocal microscope. This uses a laser source, which is focussed on the surface to be measured. The reflected beam is allowed to pass through a slit whose aperture can be changed to restrict light from the plane of focus to pass through the slit. The microscope scans through the specimen surface over a predefined depth by sequential depth slicing based on the operator setting. After the scanning is performed the obtained images are rearranged based on the depth of the plane and a three dimensional image is reconstructed back to have an overall view of the surface features with the Z scale to indicate the roughness of the sample. This gives an advantage of providing the three-dimensional surface morphology besides the quantitative value of the roughness. Besides roughness, the confocal microscope also provides an idea of the pore size, interconnectivity between pores and pore distribution for surface coatings. This is especially useful in applications such as functionally graded coatings as the reflectivity mapping of different structures are also possible with the software on the 3-D image to reveal the variation of microstructure with depth. This instrument basically fills the wide gap between the conventional scanning electron microscope and the optical microscope for characterizing the surface features and has a overlap on both sides in applicability

8.2.2 Coating thickness

Coating thickness can be measured by several methods, each of which have their own merits and demerits. The gravimetric method where a specimen is weighed before and after coating is perhaps the simplest method. For accurate results, an analytical balance with precision to 1 mg or better should be used. The coating thickness is found from the gain in mass divided by the product of area and coating density. The average thickness over a large area, e.g. 1000 mm^2, is found with no information about localised variations from average thickness. The gravimetric method is most appropriate for discrete coatings on thin substrates such as foil. It is difficult to obtain meaningful thickness data by gravimetry for integral coatings since the extent of mixing between coating and substrate is not known precisely. For very thin coatings, attention to cleanliness is essential otherwise the data will be corrupted by the weight of dirt and greases attached to the specimen. Gravimetric thickness measurements are usually limited to coating thickness of greater than 100 nm.

The thinness of most coatings poses difficulty for direct measurement, a means of circumventing this problem involves the contact between a sphere and a plane. The method is known as the ball crater method where a spherical indentation is cut or indented into the surface of the coated material. The indentation is viewed by optical microscope to determine its size. The shape of the indentation is such that the diameter of the indentation is very large compared to its depth. This means that the diameter can be used as a sensitive indicator of depth, which is set to correspond to the coating thickness. The principle of measurement for the ball crater method is illustrated schematically in Figure 8.6.



$$\boxed{\text{Thickness of the coating} = \frac{x * y}{Dia_{ball}}}$$

Figure 8.6. Schematic illustration of the ball crater method for the determination of coating thickness.

The thickness of the coating can be found from the equation:

$$l = (x*y)/D \qquad \{8.1\}$$

Where l is the coating thickness [m], x and y are the dimensions of the indentation that are coplanar with the coating, D is the diameter of the spherical indenter [m]. The diameter of the spherical indenter is typically 25 to 30 mm, which ensures that for coplanar dimensions of approximately 1 mm, a coating thickness of 30 μm can be reliably measured. The ball crater method is most often used for case hardened surface layers where the case is hard enough to remain intact and the case/substrate interface does not suffer gross deformation as a result of the indentation. The method depends on a clearly visible boundary between the coating and the substrate. Integral coatings,which lack a well-defined boundary cannot be accurately measured by this means. The sample of coated material should be able to function with the remaining coating after testing or else be disposable.

For very thin coatings of less than 100 nm thickness, indirect determinations of depth are more suitable. These involve surface analytical techniques, which are described in Section 8.6.

Besides the use of ball and crater method, ultrasonic method can be used for determining the thickness of the coating for discrete coatings. However precise calibration of the speed of the waves in the coating material and the substrate material is essential to obtain the exact thickness of the coating. This uses the principle of sending an ultrasonic wave using a probe and monitoring the reflection from the substrate–coating interface and the bottom surface of the substrate. The thickness is calculated based on the time gap between the emission of the wave and the sensing of reflected wave and the speed of the wave in the said medium, i.e. substrate or coating.

8.3 Hardness and Microhardness analysis

Hardness and microhardness measurement are very important parameters for the definition of the mechanical structure of a film or coating. Hardness is directly related to the mechanical properties of a material but the actual value of hardness is affected by the characteristics of the measurement method. Hardness and microhardness data are valuable for quality control of coatings as the data can be obtained from relatively simple and convenient tests. This facility for quality control has lead to the widespread use of hardness as a specification of material and coating quality.

The basic principle of both hardness and microhardness testing is to load a sharp pointed indenter or needle into a surface and then measure the deformed imprint after retraction of the indenter. The indentation is usually viewed with an optical microscope attached to the hardness test apparatus. Hardness data is defined in terms of the indenter geometry, e.g. spherical or pyramidal end, the load on the indenter and for visco-elastic materials the duration of loading. The operating principle of hardness measurement by indentation is illustrated schematically in Figure 8.7.

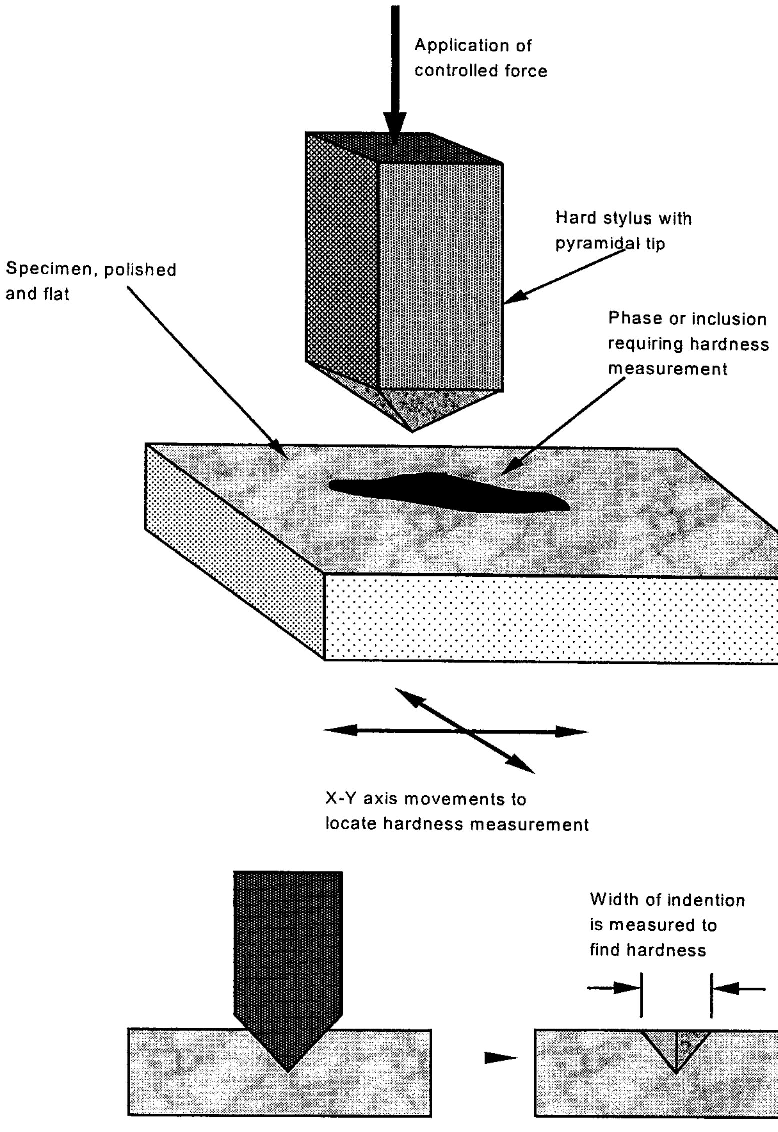


Figure 8.7. Operating principle of hardness measurement by indentation.

There are several hardness tests based on this principle, the Rockwell and Brinell tests, which use spherical indenters are widely used industrially for hardness measurement of metals. The Vickers test, which is based on a pyramidal diamond indenter, is often used for scientific purposes as it can provide hardness measurements for a wide range of materials from soft metals to hard ceramics. Unlike the Rockwell and Brinell tests, which express data in terms of proprietary scales, the Vickers test provides a measurement of the compressive flow stress, which is the parameter controlling hardness. Some standard methods of hardness testing are specified by the ASTM (American Society for the Testing of Materials), e.g. ASTM-E384. For most tests of bulk hardness an indentation load of 30 N or more is used and the size of the indentation is of the order of 300µm size in the plane of tested surface. The depth of penetration of the indenter is similar to the size in the plane of the tested surface. The bulk hardness test is unsuitable for most coatings, except for very thick coatings of several millimetres depth.

Indentation measurement is not the only means of measuring hardness; the Shore test for rubbers determines hardness by the rebound of a metal ball impacting on the rubber. Rubber hardness may be critical to the survival of rubber coatings for protection of steel pipes against corrosive-abrasive wear. The Shore test is specified in the technical standard ASTM - D 2240-91 "Standard test for rubber property - durometer hardness". This test can also be used for polymers with similar characteristics to rubbers, e.g. polyurethane. An important consideration is the thickness of the polymer sample, e.g. rubbers are often supplied as thin sheets. A minimum thickness of 6 mm is required for the ASTM test; this thickness can be achieved by stacking of thin sheets. If the polymer sample is too thin, then the depth of penetration by the hardness indenter (metal ball) will be reduced because of close proximity to the support plate. An erroneously high hardness may be measured as a result of this.

The distinction between hardness testing and microhardness testing is that with the latter, a much smaller, sharper indenter is loaded against a surface at a greatly reduced load. Typical loads for microhardness testing range from 10 mN to 10 N. The corresponding indentation is much smaller in size and can be considered to be representative of a small volume of the specimen material as opposed to the bulk of the specimen material. A commercial hardness measurement system is illustrated in Figure 8.8.

Figure 8.8. Commercial hardness measurement system.

Microhardness testing can be used to distinguish softer phases from harder phases in a microstructure as well as for investigations of surface coatings. The Knoop and Vickers tests provide similar but different measures of microhardness. The key difference between the Vickers and Knoop tests lies in the shape of the indenter. The Vickers indenter is in the form of a square-based pyramidal diamond while the Knoop indenters formed from a rhombic-based pyramidal diamond. The difference in shape causes the Knoop indenter to have a shallower penetration than the Vickers indenter for the same load. The Knoop test is therefore more suitable for hardness measurements of thin coatings. The Vickers formula for microhardness versus size of the indention is defined as:

$$\mathrm{VHN} = 1854 * P/d^2 \qquad \{8.2\}$$

For the Knoop test the corresponding definition is:

$$\mathrm{KHN} = 14229 * P/L^2 \qquad \{8.3\}$$

Where VHN is the Vickers Hardness Number [dimensionless] and KHN is the Knoop Hardness Number [dimensionless]. P is the contact load [gram force], d is the diagonal of a square [micrometers] and L is the long diagonal of a rhombus [micrometers]. The same depth of penetration is obtained with HVN = 0.3 and HKN = 0.718 but in general there is no exact correspondence between either test. The relationship between Vickers and Knoop indentation is illustrated graphically in Figure 8.9.

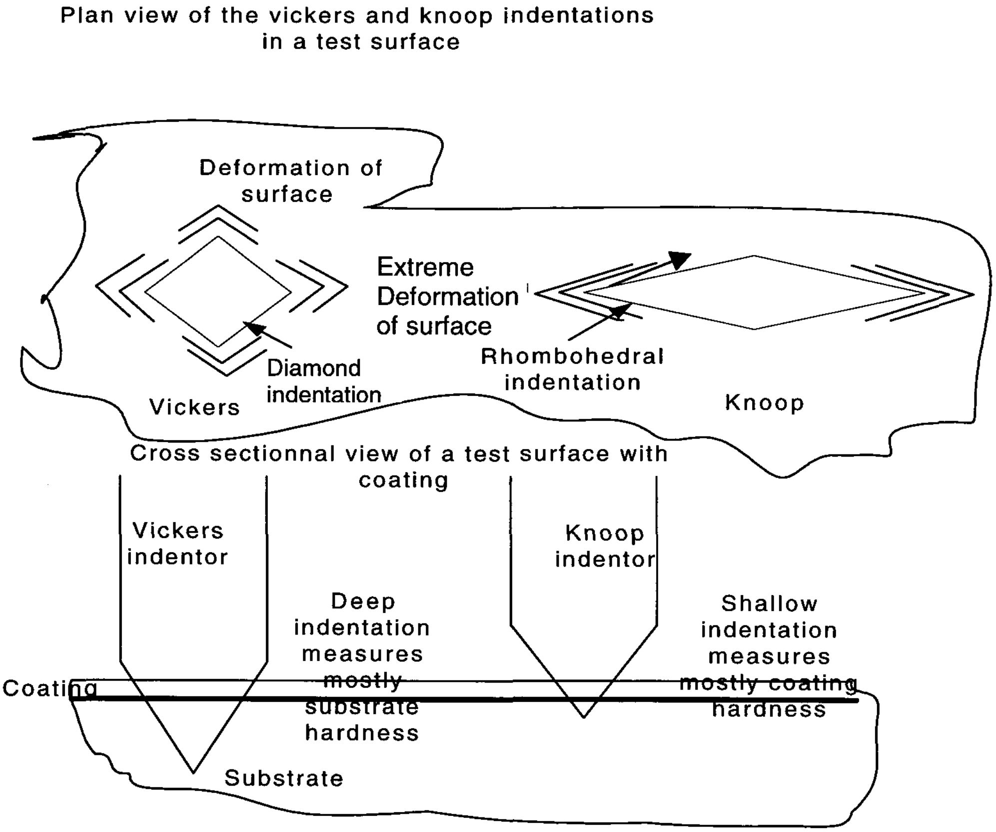


Figure 8.9. Comparison between Vickers and Knoop indentations.

Indentations can be located on the exterior surface of a material or else the material can be sectioned to reveal hardness as a function of depth below the surface. For very thin coatings, a taper section can be prepared, as this will enlarge the apparent depth of the coating by a factor equal to the secant of the angle of taper. The principle of tapering to expose more detail of a coating is shown in Figure 8.10.

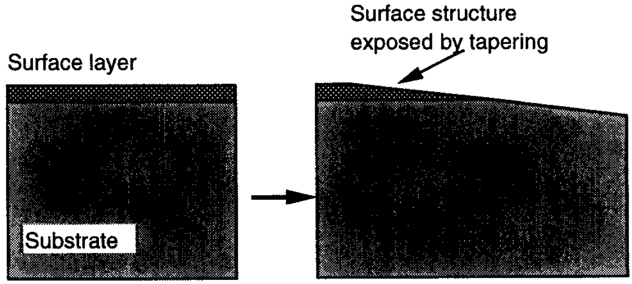


Figure 8.10. Taper-sectioning of a coating and substrate to reveal fine detail of the coating and substrate.

The measurement of the hardness of coatings can involve some difficulties when the coating thickness is very small compared to the size of the indentation. If the coating thickness and indentation depth are comparable, then the hardness of the substrate will influence the measured hardness. In these circumstances, the true hardness of the coating can only be obtained by numerical analysis of the stresses around the indenter and resulting plastic deformation. Such an analysis requires much detailed work and in practice is hardly ever performed.

Hardness testers can be used not only for hardness measurement but can in certain circumstances provide wear rate data. An indentation can be placed on the wear track of test specimen and the size of the indentation before and after wear testing measured. The size of the indentation declines in proportion to the amount of wear thus providing a direct measure of wear. Surface flow of material caused by sliding contact can also be determined by looking at the shape of the indentation. If there is surface flow, the shape of the indentation will become distorted. This method of measuring wear is convenient when displacement transducers are unsuitable. For instance, the flank wear of cutting tools has been measured from Knoop indentations.

A recent development in the coating characterization using hardness testing equipment is the Nano-indentation. This instrument has an extremely fine version of the Vicker's indentor. The surface to be examined is indented using a very light force in the order of mN ~ μN. This technique enables measurement of the hardness of the coating at different heights from the surface; allowing a depth profile of variation of hardness to be obtained. Though this method is able to predict the toughness and hardness of the coating with sufficient accuracy it is basically an approximate method to estimate the thickness of the coating within a few nanometers of error. This instrument is widely applied in advanced thin film characterization. Just as with the atomic force microscope, rough surfaces cannot be investigated with this instrument due to the fine tip of the indenter.

8.4 Adhesivity testing

The strength of adhesion between a coating and its substrate is critical to the survival of both coating and substrate. Many coatings, especially wear resistant coatings are subject to extremely high levels of mechanical stress and if there is inadequate adhesion between coating and substrate, the coating will simply detach to expose the substrate. If the coating displays strong adhesion to the substrate, then its failure mechanism is usually by wear or corrosion of the coating, which are much slower processes than uncontrolled detachment of the coating. It is also possible for the coating to fail if its melting point is lower than the substrate but this failure mode can usually be prevented by correct selection of coatings.

There are two important types of adhesivity tests, namely, the bending test and the scratch test. In the bending test, a coated specimen of the substrate is bent around a rod until the angle of bending reaches nearly 180°. The angle of bend at which the coating first shows signs of fracture or detachment is noted and used as an adhesivity parameter. The higher this angle of bending is, the stronger the adhesivity is considered to be. The principle of the bend test is illustrated schematically in Figure 8.11.

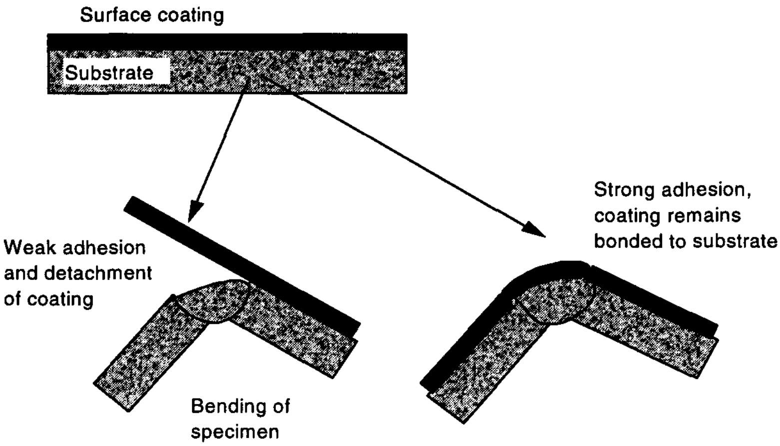


Figure 8.11. Principle of bend test for coating adhesion.

In the scratch test, a sharp stylus is drawn across the coating surface with the deliberate intention of scratching or gouging the coating. Tangential load of the stylus can be measured and recorded as a function of the normal load and the scratching speed. Coatings are evaluated by observing scratching for progressively increased normal loads on the stylus. On the basis of practical experience, the coating is expected to detach once a critical normal stylus load is exceeded. If there is no detachment of the coating, then a steady tangential force is observed for all values of contact load. If detachment occurs, then the tangential force is lower than would be expected from linear extrapolation of tangential forces at lesser contact loads. The tangential force also displays transient variation as the coating detaches in discrete lumps with peaks in the fracture force. The rapid variation in tangential force is associated with Acoustic Emission, which is a tiny amount of noise emanating from the cracking of the coating. A sensor that is in close contact with the substrate detects acoustic emission, which is then electrically amplified. The principle of scratch testing is shown schematically in Figure 8.12.

It is also possible that the coating never fails and instead deforms with the substrate to produce an even deeper groove as contact load is progressively increased. This kind of coating is extremely desirable since if the coating remains intact then the substrate is only likely to suffer wear from slower fatigue processes than by direct cutting.

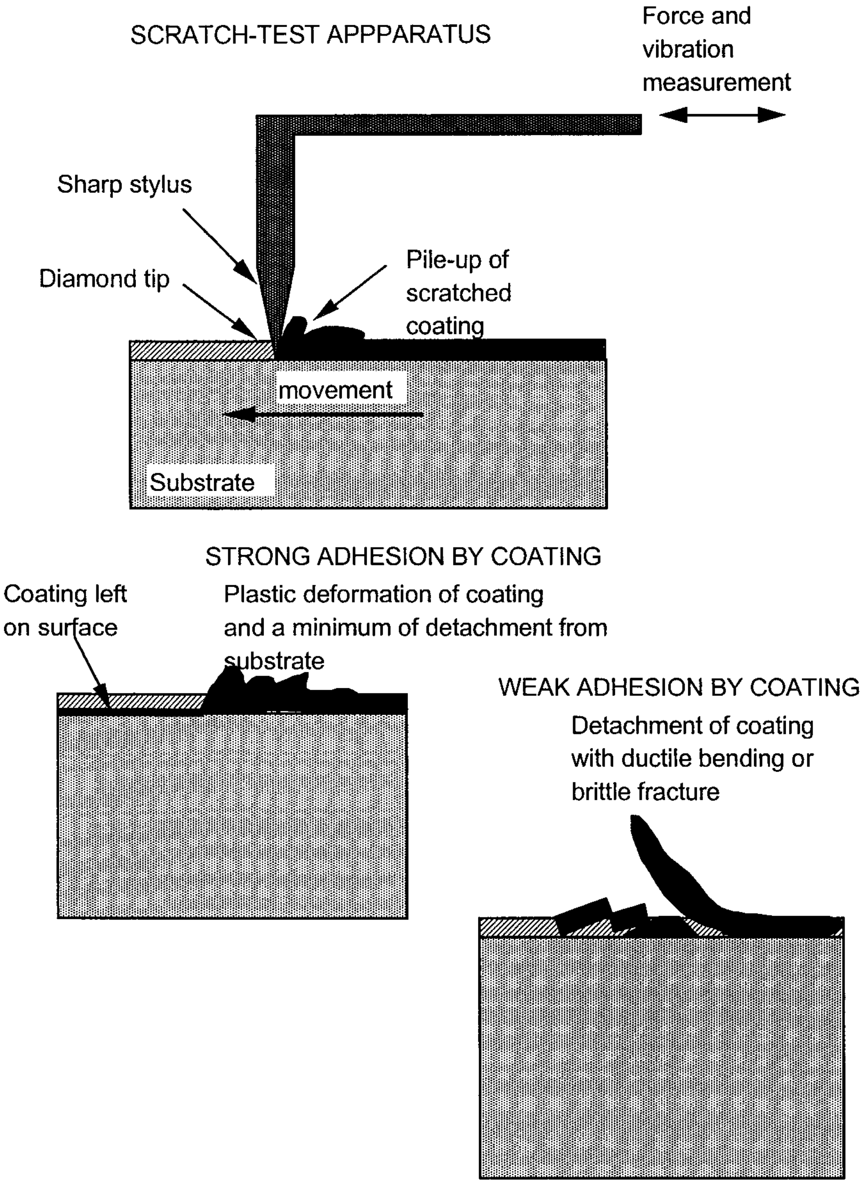


Figure 8.12. Principle of scratch testing of coating adhesion.

Even if a coating is not subjected to direct mechanical stresses, the growth of corrosion products, differential thermal expansion and ageing induced dimensional changes of the coating or substrate can all induce substantial stresses. For instance the utility of a paint is critically reliant on strong adhesion to the steel or iron substrate to prevent penetration by water under the influence of surface tension forces. Enamel coatings must adhere to a steel substrate even when the substrate is exposed to temperature changes of several hundred degrees Celsius. The corresponding strain in the coating from differential thermal expansivity could reach 1%.

There are other methods of evaluating adhesivity. For instance, in friction surfacing it is possible to leave the consumable attached to the end of a coating run. The substrate adjacent to the consumable is then separated from the rest of the substrate to be later subjected to a tensile test aligned with the centre-line of the consumable. Adhesion is then measured from the tensile fracture stress. This method of adhesion testing is illustrated schematically in Figure 8.13.

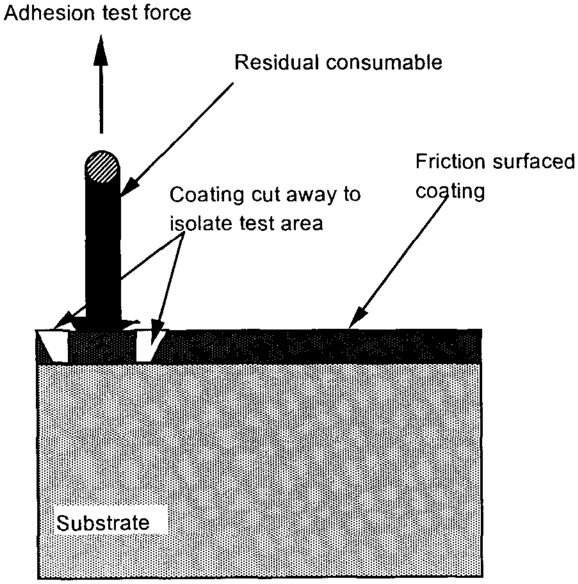


Figure 8.13. Adhesivity test for friction surfacing.

This kind of test is limited in scope to friction surfacing and even then is subject to controversy because adhesion directly beneath a stopped consumable may be stronger than adhesion beneath a moving consumable. In other words, the recorded adhesive strength may be higher than is actually found in steady-state friction surfacing.

8.5 Microstructural evaluation

There is no substitute for seeing a coating or material in full detail. Microstructural evaluation of a coating can reveal much about its functional capacity and durability in service. To evaluate a coating in detail, it is essential to determine parameters such as grain size, alignment of grains to the substrate, the presence of inclusions and voids between substrate and coating. For instance in laser surface alloying of carbon steel, if chromium is the alloying element, the surface alloyed layer transforms to martensite. If zirconium is substituted for chromium then the surface alloyed layer becomes austenitic. Since martensite is far more brittle than austenite, this change in microstructure has fundamental implications for crack growth in the surface alloyed layer.

The fundamental features to search for in a typical microstructure study of a coating are: (i) the grain structure and crystalline phase of the coating, (ii) the grain structure and crystalline phase of the substrate immediately below the coating, (iii) the interface between coating and substrate, (iv) the exterior surface of the coating.

(i) The grain structure and crystalline phase of the coating have a direct relationship to the service performance of the coating. In most applications of protective coatings, the harder or hardest phase is desirable but may not necessarily form except under a precisely controlled coating process. The grain structure affects the mechanical strength of the coating, lamellar grains aligned to the substrate are more durable than columnar grains positioned normal to the substrate. Gaps between the grains, particularly columnar grains should also be checked for.

(ii) The grain structure and crystalline phase of the substrate immediately below the coating should be evaluated for evidence of a Heat Affected Zone (HAZ). In most cases, a HAZ is undesirable unless a zone of intermediate hardness between a hard coating and softer substrate is required. For integral coatings, crystalline phases formed from a mixture of the coating material and the substrate will form and the possibility of weak or brittle phases should be checked. Intermetallic compounds arising from Surface Alloying are often much more brittle than the original metals.

(iii) The interface between the coating and substrate should ideally be only visible as a boundary between two different materials. In practice, voids and inclusions are often found at the interface. Microstructural evaluation should include detailed observation of the coating-substrate interface for any signs of mechanical weakness such as voids and inclusions. The interface is smooth and planar for most coatings with the exception of friction surfacing and explosion bonding where the interface is so rough that an exact definition of coating depth becomes impossible.

(iv) For almost all applications, the surface of the coating should be smooth with the minimum of surface cracks or depressions. The rough surface formed by the growth of columnar pyramidal grains is particularly undesirable. Surface cracking by laser surface alloying can result in sites for mechanical fatigue crack growth. Microstructural evaluation, which necessarily involves microscopy will enable the quality of the coating surface to be evaluated.

The purpose of microstructural evaluation is summarised schematically in Figure 8.14.

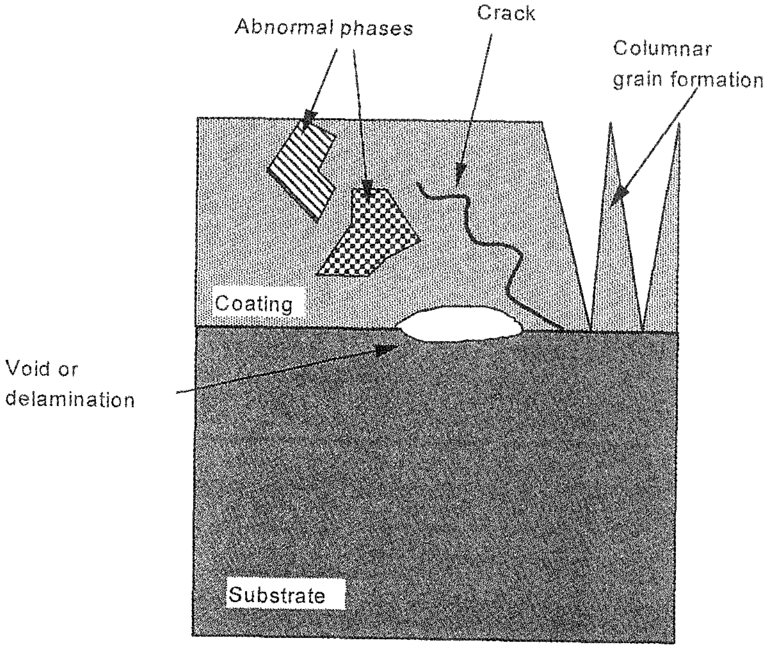


Figure 8.14. Microstructural evaluation of coatings.

Effective microstructural evaluation will depend in most cases on the combined use of several forms of microscopy, e.g. optical microscope and Scanning Electron Microscope. This is because not all microstructural features can be observed by one microscope since either the magnification is insufficient or else the visualisation medium (e.g. light or electron beams) are not sensitive to a particular feature.

Some examples of microstructural observations of materials degradation are presented below. In the first example, which is shown in Figure 8.15, thermal degradation in the form of decarburization and grain growth are identified. This is a typical consequence of service temperatures in excess of the designated level.

Figure 8.15. Microstructure of thermally decarburized steel.

Thermal decarburization is a more severe form of materials degradation than grain growth since a large decrease in surface hardness is caused. Grain growth can only be determined by comparison of grain size between the surface and interior of the substrate. In general, grain growth causes only a marginal decline in hardness and therefore is not of so critical as is decarburization, which also causes embrittlement of high carbon steels.

Microstructural observation enables determination of the retained austenite in steel. Austenite is often retained in the 'case' or surface hardened layer of a carburized steel component. Some weakening of the steel is caused by retained austenite due to the difference between the austenitic and martensitic crystal structures. Retained austenite is not however considered to be a form of thermal degradation since austenite retention is occurs in steels during the rapid quenching of steel from above the eutectoid temperature to the martensitic start temperature and depends on the cooling rate and the TTT (Time Temperature Tranformation diagram) diagram of steel in question. A micrograph of retained austenite in carburized steel is shown in Figure 8.16.

The austenite is the unetched white region and the martensite is seen as network of nodules.

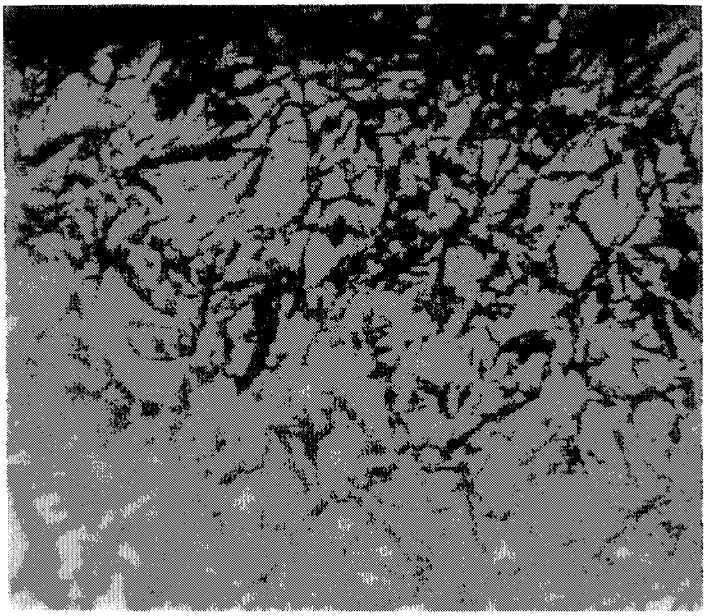


Figure 8.16. Micrograph of retained austenite in case hardened carburized steel.

Whenever a surface treatment or coating process involves intense heating or a rapid quenching of the substrate after holding the sample at high temperatures, it results in formation of an undesirable Wittmanstraten microstructure, which causes embrittlement. Although the hardness of this microstructure is high, there are also large induced tensile stresses, which render the substrate prone to microcracking. A micrograph of the Wittmanstratten microstructure is shown in Figure 8.17.

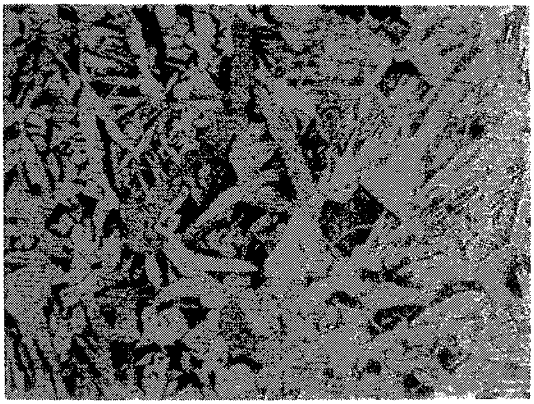


Figure 8.17. Micrograph of Wittmanstraten microstructure.

Oxidation, descaling, stress corrosion cracking, pitting, delamination wear and any other form of damage to the microstructure and integrity of a coating can all be observed by microscope in the same way as illustrated above.

8.6 Chemical analysis

The chemical composition of a coating exerts a vital role on the functioning of a coating. A coating can easily vary from its intended chemical composition because of either poor process control, dilution by the substrate or atmospheric contamination. Specialised analytical methods must be used to chemically analyse a coating, as coatings are almost always too thin to be investigated by bulk chemistry. There are two basic forms of chemical analysis of coatings: (i) determination of the chemical elements present and their relative proportions, (ii) determination of the chemical compounds present. Depending on the application, either or both types of chemical analysis may be required.

Most surface analytical techniques are based on the principle of directing either electro-magnetic radiation or an electron beam or a stream of ions at the surface of the coating. The response of the coating to radiation /electron/ion beams is to emit either radiation or electrons or ions. Measurements of the type and intensity of these emissions from the surface are then used to determine the chemical content of the coating by comparison with coatings of known composition. Most analytical instruments are equipped with a vacuum chamber for measurements as atmospheric gases usually block radiation or electron emission. Analytical techniques such as, Infra-Red Spectroscopy, X-ray Diffraction and Scanning Tunnelling Microscopy do not however require a vacuum. The principle of analysis of surface coatings by radiation or electron beams is illustrated schematically in Figure 8.18.

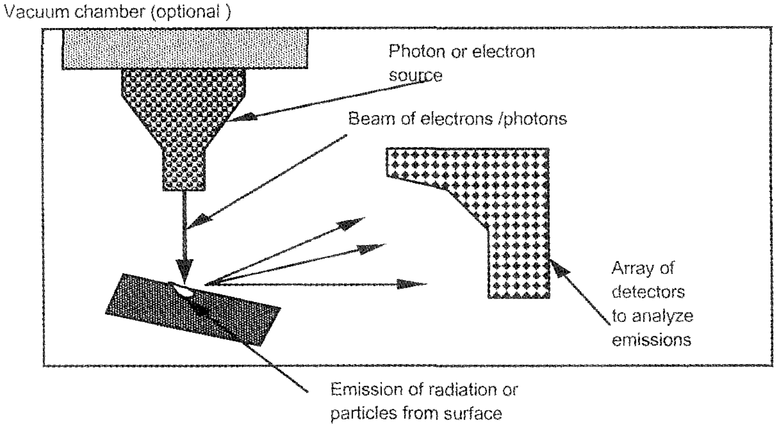


Figure 8.18. The basic principle of chemical analysis of a coating or surface.

There are many analytical techniques but only a few have gained wide acceptance because of relative convenience and range of data obtained. The classification of the main types of analytical methods is illustrated schematically in Figure 8.19.

The characteristics of Infra-Red Spectroscopy (IR-Spectroscopy), X-ray diffraction (XRD) , X-ray photo-electron spectroscopy (XPS), Energy Dispersive X-ray analysis (EDX) , Auger Electron spectroscopy, Secondary Ion Mass Spectroscopy (SIMS) and Scanning Tunnelling Microscopy (STM) are described below. STM is different from the rest of the analytical techniques since it is based on electron tunnelling between a mechanical stylus and the test surface. SIMS is also a special case since it employs ion beams not electron beams.

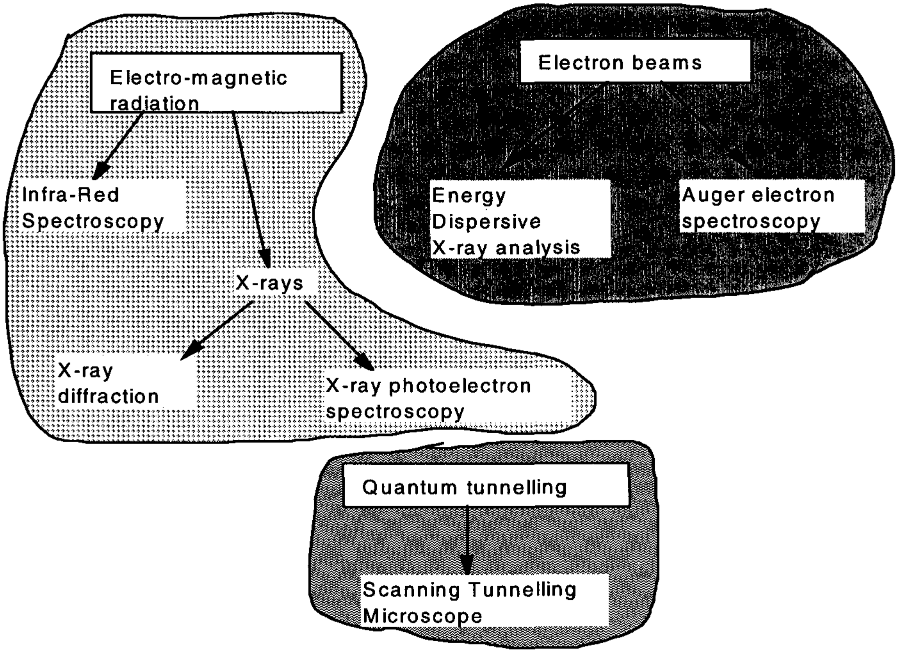


Figure 8.19. Classification of surface analysis techniques.

Infra-Red Spectroscopy

I-R Spectroscopy is an application of the characteristic of Infra-Red radiation whose wavelengths fall in the near, mid or far regime to pass through the sample. When the IR wave frequency matches the resonating frequency of the bond vibration, e.g., stretching, bending or twisting, the IR radiation is absorbed resulting in a strong absorption line in the spectrum at that wavelength. The frequency of resonant vibration is unique to each chemical bond. For example, the carbon-carbon bond has a different frequency to the oxygen-oxygen bond. Specific bonds are thus identified by I-R spectroscopy, which enables molecular structure to be deduced.

For example, if carbon-hydrogen bonds and carbon-carbon bonds are detected, then the test substance is a hydrocarbon such as ethane or higher molecular weight. Coordinated vibration of several atoms in a molecule also occurs and can provide information about molecular structure, allowing discrimination between e.g. ethane and butane. An I-R spectrometer consists of a heat source to generate Infra-Red radiation, a sample holder and semi-conductor detectors to collect the radiation after interaction with a sample. The collected I-R spectrum is usually plotted as absorbance versus wave number or frequency of radiation. At the frequency of resonance of any chemical bonds found in the sample, there will be a localised increase in absorbance, which is seen as a 'peak' on the plotted graph. The strength of the peak gives an indication of concentration while the wave number determines the bond type. There are 2 modes of I-R spectroscopy, transmission through a sample and reflectance off the surface of sample. The former is suitable for analysis of solids, gases and liquids while the latter is useful for analysis of coatings. Reflection of Infra-Red radiation affects only the outermost atoms of a coating, which is very useful for the analysis of thin coatings but has limitations when investigating thick coatings.

X-ray Diffraction

X-ray diffraction applies the ability of X-rays to be diffracted by a crystalline lattice since the wavelength of X-rays is comparable to inter-planar spacing. The diffraction pattern formed is characteristic to a specific crystalline structure and XRD is particularly useful for inorganic chemical analysis. Oxides, nitrides, carbides and sulphides of metals can usually be clearly distinguished by this technique. It is less effective for organic compounds. A diffractometer consists of an X-ray source, a sample holder with ability for rotation and a series of counter to monitor the diffracted intensity. The X-rays used are monochromatic (single wavelength) which is different from the X-rays used in imaging. Rotation of the sample facilitates detection of the diffraction pattern, which is a series of dark and light bands that denote destructive and constructive interference between scattered X-rays. The angular spacing between dark and light bands indicates the size of the inter-planar spacing, which is used to determine the composition. XRD is best applied to bulk samples but can be used to analyse thicker coatings and powder samples.

X-ray Photoelectron Spectroscopy

When mono-energetic soft X-rays are directed at a surface, electrons in the outermost shell (called as photoelectrons) are emitted from the surface. The energy required to lift an electron from the outermost shell of an element, which is plotted as a spectrum of binding energy versus the counts, is characteristic of the state of the element and the compounds that are present. For example, the iron-sulphide bond can be distinguished from the iron-oxide bond or from the iron-sulphate bond. As is the case with XRD, XPS is very useful for detailed determination of inorganic compounds and has the advantage of effectively analysis of coatings rather than bulk materials. Coatings of thickness less than 1 μm can be accurately analysed by XPS.

Energy Dispersive X-ray analysis

Energy Dispersive X-ray analysis (EDX) is used in conjunction with a Scanning Electron Microscope (SEM) as both share a common electron source. The SEM uses a flux of electrons instead of light to develop a magnified image of the observed object. When atoms on the surface of an object are bombarded by electrons, X-rays are emitted with photon energies that are unique to each chemical element present. The X-rays are generated by jumps of electrons between electron shells in an atom that occur during external electron bombardment. When the jumped electron returns to its equilibrium electron shell, a photon with energy dictated by quantum physics is emitted. The energy is sufficiently high to generate an X-ray photon. The penetration of electrons into a sample is typically in the range of 0.1 to 1 μm and the elemental composition of coatings with this range of thickness can be accurately identified. EDX can be used either to analyse composition at a point or over a wider area, specific elements can also be mapped over a test surface by scanning for X-ray emission at a specified energy. EDX is mainly used for inorganic coatings as carbon and hydrogen release X-rays with very low energies that cannot be easily detected. The detection problem is caused by the window between the SEM chamber and X-ray detector. A standard window excludes low energy X-rays and it is necessary to open the window before the X-ray detector can respond to X-rays from carbon or other light elements. Window opening entails a risk of damage to the X-ray detector and is usually only performed after reaching a very low level of residual gas pressure in the microscope chamber. The first five elements of the periodic table, hydrogen, helium, lithium, beryllium and boron are impossible to detect while carbon can only be detected with difficulty.

Auger Electron Spectroscopy

Auger Electron Spectroscopy (Auger) uses a focussed electron beam to etch a hole in the coating. This hole is tiny in size and deepens at the extremely slow rate of a few manometers an hour. A depth profile of chemical elements in the coating is obtained by spectroscopy of the atoms removed during etching of the hole. Depth profiling is so far a unique characteristic of Auger, which is extremely useful as a 3-dimensional view of a coating can be obtained as opposed to the 2-dimensional view provided by other analytical techniques. It is often found that coating composition varies with depth because of either corrosion of the coating or poor control of coating deposition. Auger can be used to show whether a metal coating has been partially oxidised or is separated from its substrate by a layer of impurities. The objectives of determining elemental distribution versus depth in the coating and elemental distribution over an area of the coating are shown schematically in Figure 8.20.

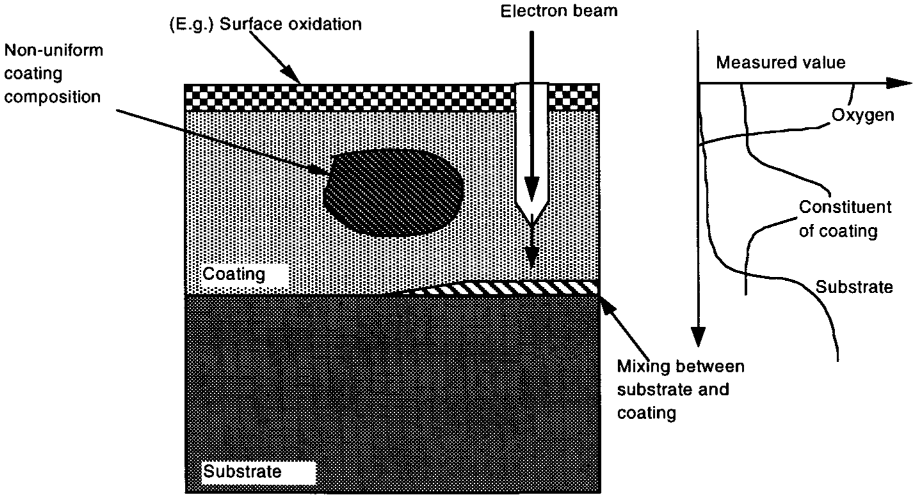


Figure 8.20. Significance of depth and area determination of chemical element distributions in a coating.

Secondary Ion Mass Spectroscopy

SIMS is often used as an attachment to the Scanning Electron Microscope where a common vacuum facility can be conveniently shared. During a measurement by SIMS, the specimen surface is obtained by bombarding the surface with a stream of ions, which are typically oxygen, nitrogen or argon ions. Ions of surface material are ejected by the bombardment and are collected in a mass spectrometer that is part of the SIMS system. The data obtained from the mass spectrometer is the elemental composition of the specimen surface. SIMS is very sensitive to small concentrations of elements and so permits reliable detection of trace impurities in a surface. Concentration differences of a factor of 10,000 or more can be accommodated by SIMS. In this way, SIMS is superior to EDS, which is effective for bulk constituents only. A larger area of specimen surface is required by SIMS than EDS so that localised variations in elemental composition are not easily detected by SIMS. A related but different analytical technique, Ion Microprobe analysis (IMA) employs a focussed ion beam to give analysis over a small area of a few micrometers in dimension. The limitation of IMA is that ions penetrate the surface more deeply than is the case with SIMS. This means that measurements of the composition of a coating will be probably affected by the composition of the substrate.

Scanning Tunnelling Microscopy

STM is a relatively new technique that has the special feature of permitting analysis of individual atoms when used at high magnification. STM is closely related to the Atomic Force Microscope described in Section 8.2 where the same stylus is used but electrical signals are collected instead of force variations. STM is exclusively used for analysis on the microscale to determine boundaries of atomic features on a surface. The data obtained does not provide absolute identification of chemical elements but can be used to distinguish between different elements on a surface. This instrument is suitable for detection of localised impurities or features on a coating such as minute inclusions or surface protrusions.

By employing two or more analytical techniques, a complete view of the elemental distribution in a coating can be obtained. It is often a surprise to learn how a purportedly uniform coating is revealed to contain numerous imperfections and impurities, which almost certainly affect its service performance. Defects such as holes in the coating and oxidation of the coating exterior are extremely common. Basic manufacturing problems such as poor surface finish of the substrate can be seen to prevent the formation of a uniform thin coating on a surface.

8.6.1 Comparison of surface analysis techniques

Each type of surface analysis, e.g. X-ray Photo-Electron Spectroscopy generates a specific type of data, which gives only a partial view of the composition of the surface. Specific applications are best served by some analytical techniques and not others. For example, if the area distribution of chromium on a corroded stainless steel surface is required, then Energy Dispersive X-ray analysis is more appropriate and Infra-Red Spectroscopy cannot be used for such analysis. Conversely, if the surface degradation of a rubber tyre needs to be investigated, Infra-Red Spectroscopy is preferable to Energy Dispersive X-ray analysis. When a comprehensive knowledge of the surface or coating is required, at least two analytical techniques should be used to eliminate any deficiencies in the analytical data.

The scale of magnification also imposes limits on the range of applicable analysis; Scanning Tunnelling Microscopy is suitable for discrimination of individual atoms but cannot be used for wide area analysis. Most of the other analytical techniques are suitable for investigation of areas of a few square micrometers. If larger areas must be analysed, extra time has to be expended in sequential analysis of many small areas of the specimen surface. For this reason, most of the specimen surface is neglected and analysis is focussed on small areas of surface, which are considered to be representative or hold critical detail. Expertise is required to ensure that critical evidence is not overlooked when this practice is followed.

A summary of the major types of surface analysis and their range of application is tabulated in Figure 8.21. The table is necessarily abbreviated and there may be some specialised exceptions to the presented rules.

Features Surface analysis	Elemental composition	Chemical bonding	Distribution of chemical elements	Depth of penetration
Energy Dispersive X-ray analysis	Yes, except for atomic number < 15	No information	Areal distribution	Fixed at 0.1 to 1 μm
Infra-Red Spectroscopy	Indirect data only	Detailed information	Unsuitable	Less than 100 nm
X-ray photo-electron Spectroscopy	Indirect information	Detailed information	Areal distribution	Less than 100 nm
Auger electron	All elements Semi-quantitative.	No information	Unsuitable	Variable, up to 1 μm
X-Ray Diffraction	All elements, quantitative	Detailed for inorganic compounds	Unsuitable	Fixed, about. 5 μm
Secondary Ion Mass Spectroscopy	All elements, with high sensitivity to trace elements	Unsuitable	Unsuitable	About 1 μm
Scanning Tunnelling Microscopy	Partial information only	Unsuitable	Very high resolution of analysis	Surface atoms only

Figure 8.21. Scope of application of the major surface analysis techniques.

It can be seen from Fig. 8.21 that with the application of several methods of surface analysis, a complete picture of the coating or surface can be obtained. Elemental composition, chemical compounds, area and depth distribution can be now known with considerable precision. For fine detail of e.g. the grain boundaries of a coating, the Scanning Tunnelling Microscope can be used. It is possible to accurately specify and control the characteristics of a coating or surface. This helps to ensure the reliability of coated and uncoated components when subject to materials degradation.

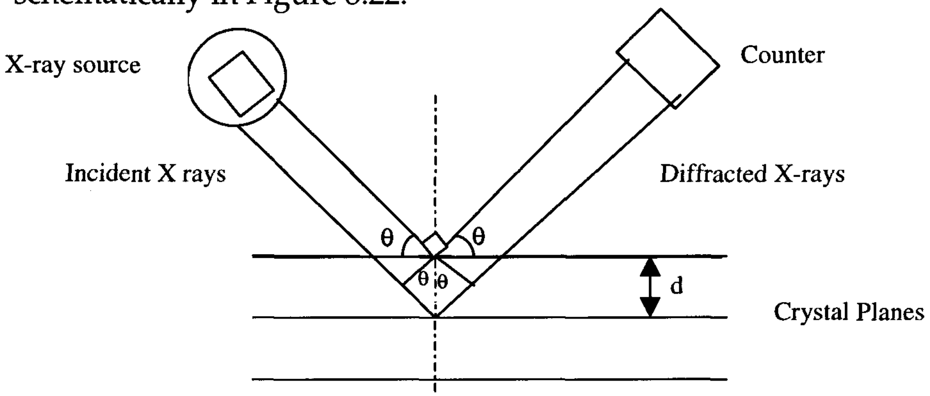
8.7 Residual stress analysis

Residual stresses are often generated in a component due to the various processing conditions. A typical example of the same may be found in castings and weldments where the core cools much later than the surface resulting in contraction of the core when the surface is already a rigid solid. This causes the core to remain always in tension due to restriction from surface atoms and surfaces to be in compression. Similarly most surface modification and coating techniques entail the induction of residual stresses at the surface of the substrate. Residual stresses may be generated by thermal, mechanical, compositional or a combination of these processing factors. In order to prevent fracture, it is of critical importance to be able to determine the nature, magnitude and distribution of any induced residual stresses in a component. Microscopic stresses arising from crystalline defects, dislocations or the anisotropy of individual grains are not of concern. Much more important are the macroscopic stresses which are induced over the entire microstructure close to the surface of the substrate. These residual stresses can be measured by the conventional Sach's boring method, interferometry or X-ray diffractometry.

Sach's boring method involves the attachment of a strain gauge rosette on the test surface followed by the boring of a small hole close to the rosette. Residual stresses are then deduced from the measured changes in strain upon completion of the boring. The main limitations of this method are (i) its destructive effect on the specimen, (ii) difficulty in distinguishing between the original residual stresses and those induced by boring. Specialised high-speed drills and air abrasive jets are used in an effort to suppress any additional residual stress. A semi-quantitative measure of residual stress is obtained by Sach's method.

Interferometry is based on the stresses observed in a plastic replica of the test specimen. A transparent plastic model of the specimen is prepared and placed in polarised light to reveal interference fringes. Residual stresses in the original specimen are simulated by appropriate heat treatment of the plastic model. The model is not subjected to any stress so that all observed stress is due to its own residual stresses. Highly stressed regions in the model are considered to correspond to regions of high residual stress in the original specimen. Interferometry is useful for the qualitative detection of residual stress sites. Improvement of interferometry to provide quantitative values of residual stress is the subject of current research.

X-ray diffractometry represents the only non-destructive method of quantitatively evaluating residual stress. The method of diffractometry requires the same equipment as is needed for X-ray analysis of elemental composition. A primary limitation of the method is that it is strictly applicable only for a material that is elastically homogeneous and isotropic. Polycrystalline metals represent a good approximation to this requirement. In a similar manner to strain gauges, X-ray diffractometry measures strain which is then used as data to calculate stress according to the elasticity theory. The principal of X-ray diffractomery is illustrated schematically in Figure 8.22.



Phase difference = xy +yz= d sinθ + d sinθ
= 2d sinθ

Figure 8.22. X-ray diffraction by a crystal lattice when irradiated with monochromatic X-rays.

X-rays penetrate the surface and are diffracted by the crystal lattice. The strongest diffraction effect occurs at the surface where absorption of X-rays is very weak. Diffraction patterns are generated when scattered rays reinforce one another in constructive interference which occurs when the incident angle of the X-ray, its wavelength and the lattice spacing of the crystal (that makes the specimen) conform to the relationship:

$$n * \lambda = 2 * d' * \sin(\theta) \qquad \{8.4\}$$

Where n is the order of the diffraction pattern, λ is the wavelength [m], d' is the lattice spacing of the crystal [m] and θ is the angle of incidence of the X-rays [radians]. This equation is commonly referred to as Bragg's law of diffraction. The inter-lattice spacing is dependent on the applied stress and so is effectively a measure of strain. In less exact terms, the interlattice spacing serves as a microscopic strain gauge. To determine residual strain, lattice strain parallel to the surface and at an inclined angle to the surface are both measured. The specimen is located at the centre of a circular frame, which holds both the X-ray source and the X-ray detector. The detector is moved around the circular frame to record X-ray intensities at various angles of reflection. The specimen can be pivoted about the centre of the circular frame to obtain measurements at different angles of inclination of the specimen surface to the X-ray beam. A schematic view of this measurement is shown in Figure 8.23.

The intensity of reflected X-rays usually displays a sine wave relationship to the angle of orientation of the detector. The angle at which the X-ray intensity reaches at maximum and the change in this angle with inclination of the specimen are the essential data. The relationship between X-ray intensity and angle of detector orientation for different angles of specimen inclination are shown schematically in Figure 8.24.

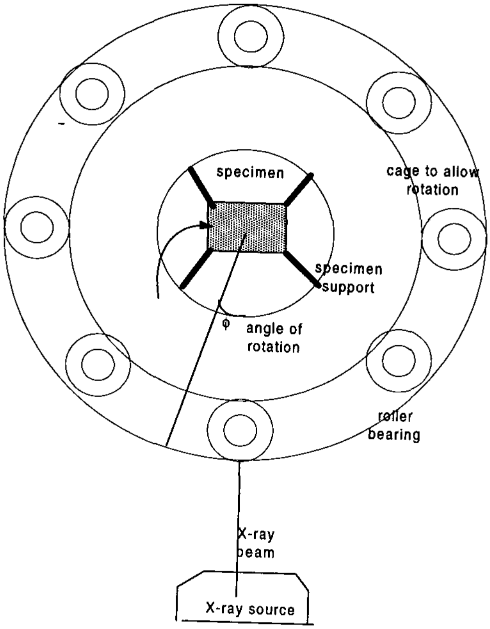


Figure 8.23. Strain measurement by X-ray diffractometry.

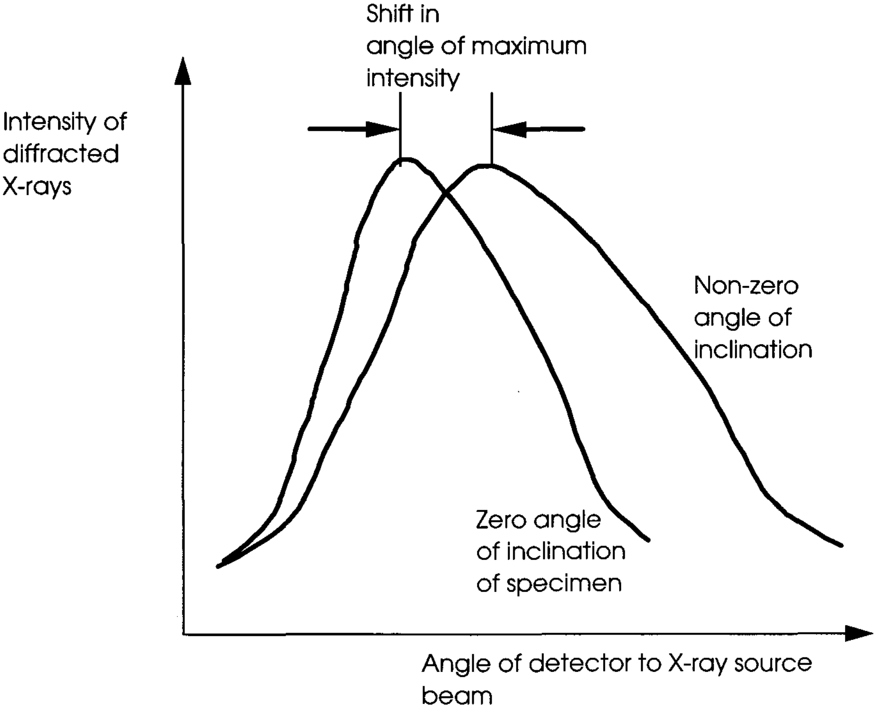


Figure 8.24. Distribution of intensity of reflected X-rays for various specimen inclinations.

Residual stress is deduced from X-ray intensity measurements by using the following analytical model.

According to the theory of elasticity, the relationship between strain and stress in any other direction is:

$$\varepsilon_\psi - \varepsilon_\perp = \sigma_\phi*(1+\nu)*\sin^2(\psi)/E \qquad \{8.5\}$$

Where ε_ψ is the strain at an inclined angle [dimensionless], $\varepsilon\perp$ is the datum strain [dimensionless], which is typically zero, σ_ϕ is the stress [Pa] in the designated direction, ν is Poisson's ratio [dimensionless], ψ is the inclined angle [radians] and E is the Young's modulus [Pa].

Equating strains in terms of the original lattice dimensions gives the following definitions:

$$\varepsilon_\psi = (d_\psi - d_o)/d_o \qquad \{8.6\}$$

$$\varepsilon_\perp = (d_\perp - d_o) / d_o \qquad \{8.7\}$$

Substituting these definitions into Eq. 8.5 gives the following relationship:

$$(d_\psi - d_\perp) /d_o = \sigma_\phi*(1+\nu)*\sin^2(\psi)/E \qquad \{8.8\}$$

The lattice spacing, $d_\perp$ and d_o are virtually identical thus resulting in negligible error when d_o is replaced by d_ψ. Differentiation of Bragg's law (Eq. 8.4) and its combination with Eq. 8.8 creates after some detailed manipulation, the following relationship:

$$\sigma_\phi = -2 * [E \cot(\theta)/(2 * (1+\nu)*\sin^2(\psi))]*\Delta\theta \qquad \{8.9\}$$

Where θ is the detector angle [radians] where the maximum X-ray intensity is observed and $\Delta\theta$ is the change in θ with shift in specimen inclination, ψ [radians].

It is customary to define a stress factor K as:

$$K = E \cot(\theta)/(2 * (1+v)*\sin^2(\psi)) \qquad \{8.10\}$$

This simplifies Eq. 8.8L* to:

$$\sigma_\phi = K *[\, 2*\theta_\perp - 2*\theta_\psi] \qquad \{8.11\}$$

As illustrated in Fig. 8.23, 2*θ is effectively the angle between the transmitted and the refracted X-ray beam which is easy to measure experimentally. To determine the stress in any given direction, i.e. σ_ϕ, intensities of transmitted X-rays are recorded for angles spanning the maximum in observed X-ray intensity. These measured intensities are then divided by the Lorentz polarisation factor and the absorption factor, which are:

$$LPF = (1+\cos^2\theta)/\sin^2\theta \qquad \{8.12\}$$

$$AF = 1 - \tan\psi * \cot\theta \qquad \{8.13\}$$

Where LPF denotes the Lorentz polarisation factor and AF denotes the absorption factor. The absorption factor is only applied for non-zero values of ψ. Data required to compile a quadratic curve that models the 2*θ maximum, is obtained by collecting measurements at ψ = 0°, 45° and 90°. For angles 45° and 90°, focussing of the X-ray beam is achieved by moving the detector towards the specimen using a radial sliding facility (as illustrated in Fig. 8.23).

Real metals display crystallographic anisotropy, which prevents calculation of the stress factor K from the physical properties of the bulk material. The metal crystals lying close to the surface exert the strongest influence on diffraction of the X-rays but these crystals have abnormal physical characteristics because they are only constrained in the subsurface direction. The stress factor K is often determined experimentally by applying various known levels of external stress and measuring the corresponding change in 2*Δθ by X-ray diffractometry.

The X-ray diffraction data gives only a single value of residual stress, when a stress profile is required, stress levels at various depths in the material must be measured. Residual stress profilometry is usually performed by chemically dissolving the metal to various depths and recording X-ray data. Mechanical removal is usually not practised because of the risk of adding extra residual stresses. A major problem with this technique is that when surface material is removed, the stress profile must change in response. For instance, when a surface layer of compressed material is removed, the subsurface layers in tension will then relax. Measurements of the subsurface residual stress must contain a correction to allow for stress relaxation. Calculation methods to predict or correct for stress relaxation are beyond the scope of this book.

8.8 Special techniques for dynamic testing conditions

One of the most important factors in materials degradation is that most of the associated mechanisms are often hidden in the substrate or the interface. It is thus difficult to initiate a preventive action to avoid catastrophic failure of the system. Typical examples could be phase separation or stress corrosion by the formation of an internal galvanic couple or wear. In the first two cases the mechanisms are rather slow and needs to be monitored using non-destructive testing techniques such as ultrasonic scanning or X-ray imaging.

Both these techniques are extensively used in engineering and other fields such as medicine for different purposes. Ultrasonic scanning is more useful in static circumstances or for degradation processes which are relatively slower like corrosion or phase changes. It is also important that the surfaces are relatively flat as the ultrasonic wave is sensitive to contour changes on the surface and a perfect contact between the probe surface and the surface to be examined is essential for accurate results.

X-ray imaging on the other hand forms a truly non-destructive technique, which does not impose stringent requirements on surface flatness. X-ray fluoroscopy or X-ray imaging, operates on the principle of attenuation during transmission through the sample and thus has the capability of observation of interfaces in coatings as well as in contacts between a stationary pin and a rotating disk in a sliding contact. In this technique the emitted X-rays are directed through a sample, which is either a static coated specimen or a static pin resting on rotating disk. In the first case of a static coated specimen, the X-rays are focussed on the specimen, which the X-rays are able to penetrate and reach the fluorescent screen placed below. The fluorescent screen emits visible light depending on the local intensity of the X-rays, thus forming an image. During the transmission through the sample the X-rays get attenuated differentially due to the local differences in thickness, defect concentration and density. Thus the received X-ray forms a true depiction of the features in the coating, interface and the substrate. However a serious limitation is that the effects are cumulative over the thickness on the transmitted X-ray. An experienced engineer would be able to overcome this limitation by varying the kV as the intensity is directly related to the applied kV and decides the depth of penetration before a dark field is seen on the screen. This method is rather a rough estimate and much depends on the experience of the engineer who is interpreting the X-ray images. In order to over come this lacuna, Computed Tomography was introduced which uses depth slicing using software to interpret the X-ray data from the CT sensor. Figures 8.25 & 8.26 show a typical X-ray image of the specimen during a wear test and the corresponding histogram profiles.

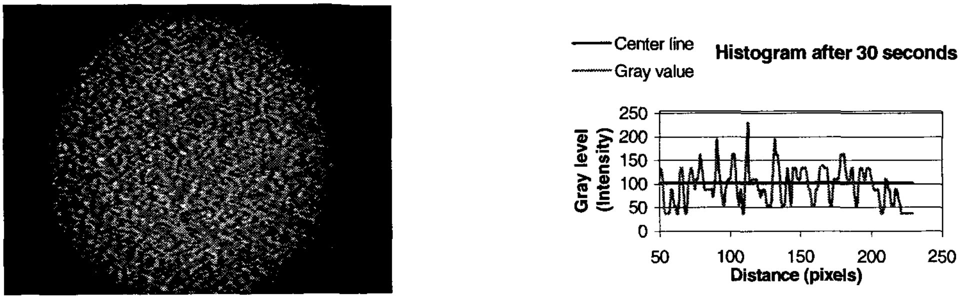


Figure 8.25 X-ray image of a sliding contact with the corresponding histogram, profile along the center line (horizontal) of the image.

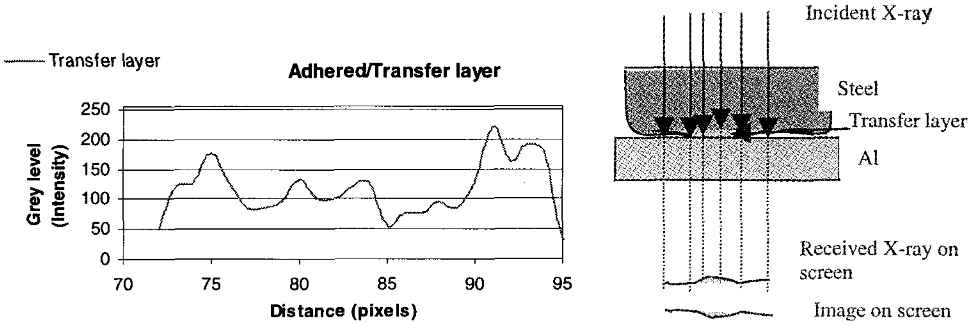


Figure 8.26 Histogram profile and a model depicting the mechanism

8.9 Analysis of service characteristics

While surface analysis, microhardness measurements and scratch tests etc. can reveal much information about the chemical and physical characteristics of a coating, the service life of a coating still remains unknown. There is still no mathematical model or equation that predicts service life in terms of basic physical and chemical descriptors such as hardness and galvanic potential. The closest approximation to such a model is the Tafel equation, which predicts aqueous corrosion rates. In theory it would be possible to predict the durability of a metal coating based on the rate of dissolution of the metal. In practice however, coating flaws such as holes and blisters will initiate failure of the coating before most of the coating has been dissolved by corrosion. For wear problems, there is no wear model comparable to the Tafel equation so that empirical measurement of service life is required. Fatigue, thermal and irradiative degradation share the same difficulties in prediction or at least cannot be predicted with precision. For any specific service application, a test has to be either selected from a technical standards organisation or devised before the service life can be known with precision. Standard tests, which are compiled by technical societies, provide basic information about materials degradation in a form that facilitates comparison between independent laboratories. Standard tests can also be used for the certification of materials and coatings. The American Society for Testing of Materials (ASTM) is an important source of standard tests. Other important societies include the American Society of

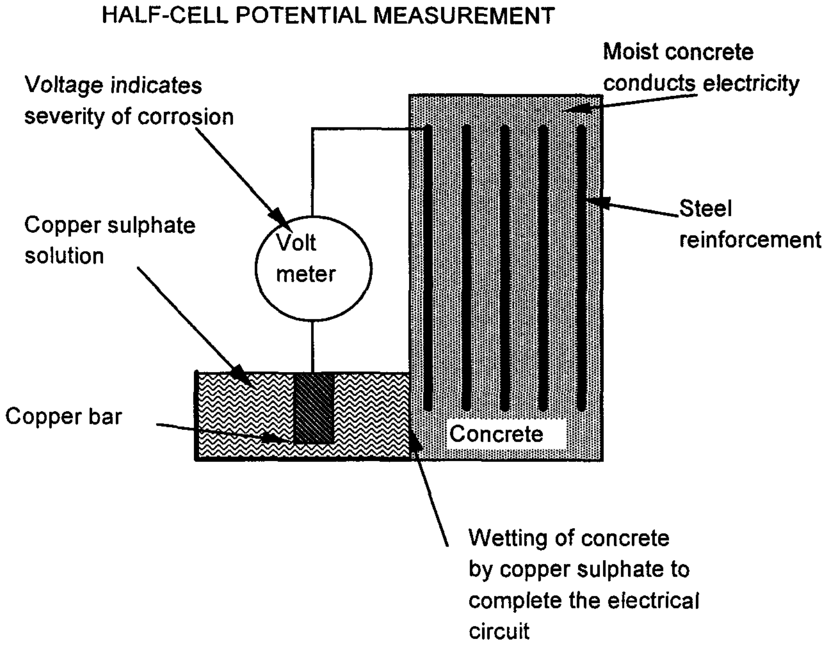
Mechanical Engineers (ASME) and the American Petroleum Institute (API). A listing of these tests and others is provided in Appendix 1. Tests developed by these societies and others are often used to provide legally binding evidence of the quality of resistance to corrosion, wear or fracture. It should be noted however, that no standard test could *guarantee* that a material or coating will give the required performance in any specific service environment. All tests are a compromise between practicality and experimental accuracy. Testing is expensive and time-consuming and it is often difficult to model long service lives. For example, the evaluation of service life of a biomedical implant intended to last for 20 years, is based on tests lasting perhaps 6 months. Imperfections in our understanding of materials degradation prevent the precise formulation of experiments to accurately replicate degradation phenomena. A model corrosion, wear or fracture test has to be developed which is simple and cheap to perform yet accurately simulates conditions in the real application. The most significant types of tests are described below.

8.9.1 Corrosion testing

Aqueous corrosion is more significant economically than other forms of corrosion and so is the subject of most corrosion tests. A basic form of test is the atmospheric exposure test where panels of test metal are laid out on a rack and exposed to open air for periods of time as long as several years. The prevailing temperature and rainfall are noted and compared with progress of corrosion. The extent of corrosion is assessed semi-quantitatively by noting the time required for the test panels to tarnish, develop rust or even disintegrate into powdery corrosion products. The choice of location for the test is important, coastal locations are suitable for monitoring the effect of salt spray, while industrial locations allow evaluation of corrosion resistance to acidic rain and sulphur compounds. As may be appreciated this type of test is quite slow and does not facilitate efficient optimisation of corrosion resistant coatings. In more advanced experiments, an artificial indoor environment is created where severe corrosion conditions occur continuously. Instead of long periods of dry sunny days punctuated by brief rainy periods, as is the case for natural weather, test materials can be continuously showered with water and held at a climatically high temperature to accelerate corrosion. Salt water can also be substituted for fresh-water to generate a salt spray resistance test, which is considerably more severe than natural rainfall. Some time limitations still remain and a large number of different specimen materials are often tested simultaneously to speed the search for the optimum material or coating.

More advanced corrosion testing involves monitoring the galvanic potential and corrosion current of a structure. This method of testing is particularly useful for hidden structures such as steel reinforcing bars inside concrete. The test structure is treated as a half-cell of an electrochemical cell (described in Chapter 3) while a reference electrode of known characteristics is used to complete the electrochemical cell. Copper in saturated copper sulphate is often used as the reference electrode for tests on reinforced concrete. The cell voltage or corrosion potential is then measured and used as an indicator of likely corrosion based on an empirical scale published by the relevant technical society. A major problem with this test is that the critical voltage for corrosion varies with type of structure and environment, e.g. whether or not the salt contamination is severe. The corrosion potential test and environmental corrosion test are illustrated schematically in Figure 8.27.

For more severe corrosive environments, e.g. molten salts, evaluation of coatings and materials is based on direct experimentation with less reference to technical standards as in most cases the relevant standard is unavailable. For exposure to strong acids, testing is arranged to determine the depth of corrosion in a specific time e.g. 100 hours for specific acidities (pH) and temperatures of the acid. It is important to ensure that the correct acidity and temperature is used for testing since corrosion rates and passivation phenomenon are critically dependent on both these parameters.



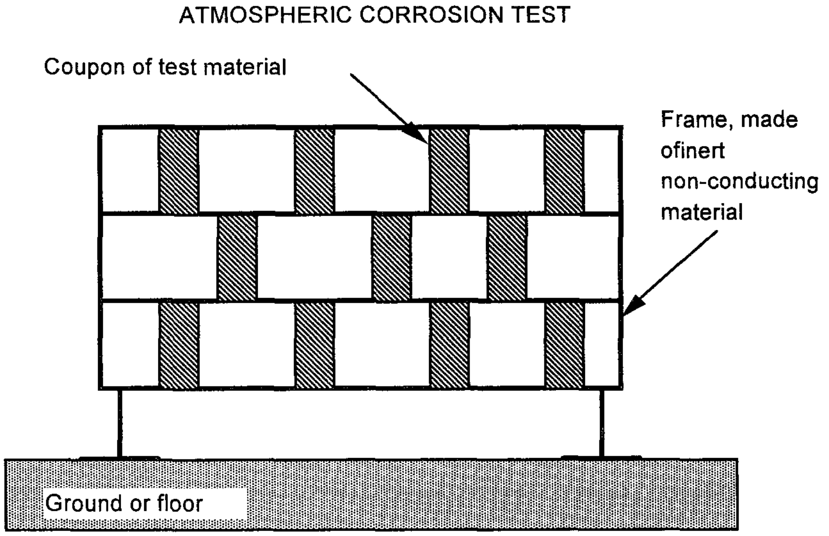


Figure 8.27. Basic principle of atmospheric corrosion tests.

8.9.2 Wear resistance tests

Wear testing is subject to a similar level of complexity as corrosion testing since wear is affected by many factors both physical and chemical. Mechanical parameters such as load and speed can combine with chemical parameters such as acidity and oxygen concentration to produce a complex wear characteristic that is hard to predict analytically. In most instances of wear testing, a standardised apparatus such as a pin-on-disc apparatus or a 4-ball apparatus is used to evaluate wear. A pin-on-disc apparatus consists of a cylindrical pin, which is placed perpendicularly to a rotating disc. A 4-ball apparatus consists of 4 steel balls (as used for rolling bearings placed together in the form of a pyramid. The lower 3 balls are usually held still while the top ball is rotated to generate wear. The friction coefficient is also measured when using these apparatus and others. Popular types of wear apparatus including the pin-on-disc and the 4-ball machines are shown in Figure 8. 28.

All of these apparatus are intended to represent either a practical wearing contact or else an extreme version of it, as is the case with the 4-ball test and Falex test. The pin-on-disc and pin-on-flat tests are widely used for scientific experiments while the 4-ball and Falex tests are mostly used for commercial quality control tests and technical standards. Wear rates are usually measured from changes in dimension of a worn specimen, either by the change in length or from the enlargement of the wear scar. Friction force is measured using a strain gauge load cell or a piezoelectric force transducer. Temperature is also recorded because of its strong effect on wear. In advanced tests, the temperature close to the contact is measured while in simpler tests; the overall apparatus temperature is substituted. External heating can also be fitted to allow wear tests at elevated temperatures and investigate the effect of thermal degradation or high temperature corrosion on wear resistance. Relative humidity should be checked as very low levels of humidity cause variations in wear from typical ambient conditions. Acoustic emission, which can be loosely described as 'micro-noise', is a very useful indicator of changes in a coating, e.g.

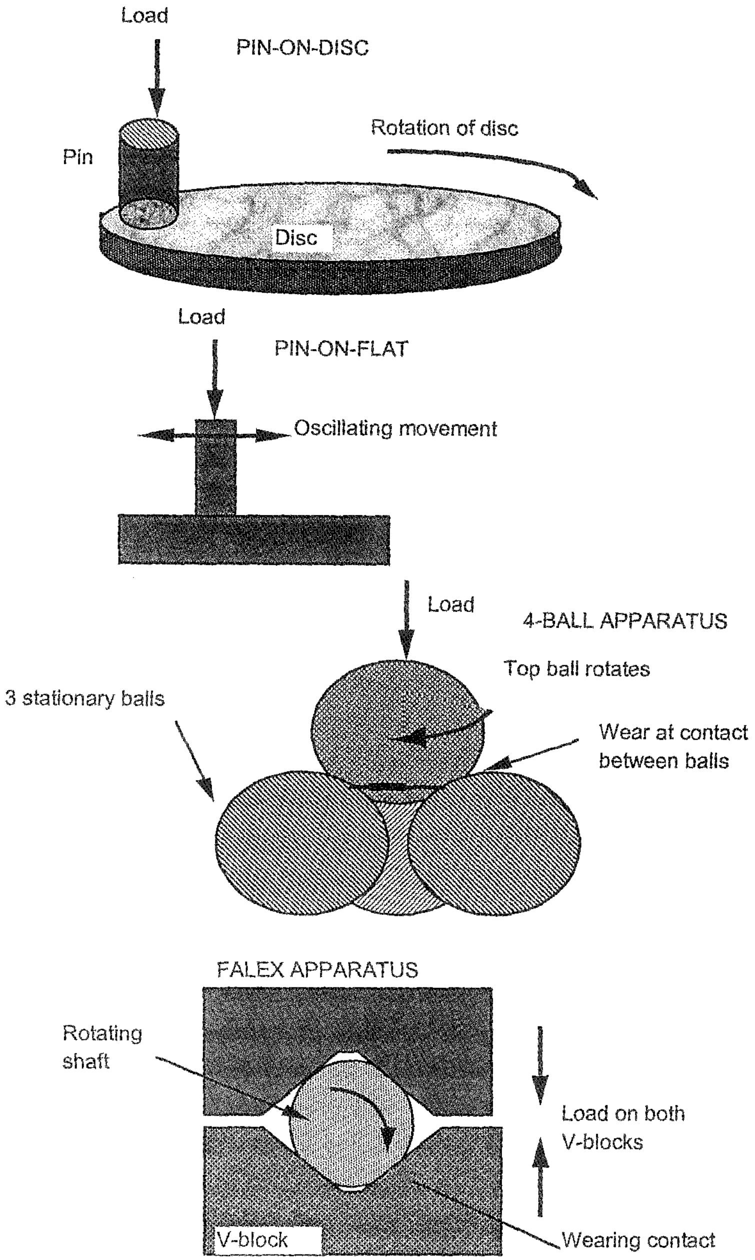


Figure 8.28. Configurations of standard wear tests.

when rapid mechanical destruction of the coating is occurring. Acoustic emission is monitored by a piezoelectric probe that is attached to the wearing test piece in a manner designed to ensure a good acoustic connection.

Standardised tests are widely used by industry to determine the wear resistance of materials and coatings. The performance of a test is closely specified in a standard method that is issued by technical societies such as the American Society of Mechanical Engineers (ASME) or the American Petroleum Institute (API). The method specification details the loads, sliding speeds, timing of tests as well as the preparation of materials. Compliance with a standard test allows immediate comparison of wear data obtained by independent research groups. Diverse data can then be collated to obtain a general understanding of the wear and friction characteristics of a material. Very careful performance of tests with the utmost respect for cleanliness and purity of materials is necessary for good agreement between data of different experimental laboratories.

There is only an imperfect correspondence between data from standardised tests and the observed wear characteristics of actual mechanical systems. Wear rates are extremely sensitive to factors such as the level of contamination in a lubricating oil, temperature variation, oxygen concentration around the wearing contact and mechanical vibration. The operating environment in a real mechanical system is usually unknown because of the cost and difficulty of measurement. For example, one might imagine fitting a gearbox with temperature sensors, vibration monitors and oxygen probes. For detailed and accurate measurements of wear, a full-scale test of the actual mechanical system is often performed. The most common example is an engine test where an engine is fitted inside a test-bed and motored with e.g. different lubricating oils or piston ring coatings. This type of test is extremely expensive both in terms of resources and time. The virtue of standardised wear tests is that they provide some indication of the quality of e.g. a wear-resistant coating much more quickly and cheaply. Tests with full-scale mechanical systems can then be performed on a short-list of near optimum test materials or coatings.

8.9.5 Fracture testing of coatings

Because of the weaker effect of coatings on fracture there has been much less development of fracture tests for coated materials. Tests on the fracture resistance of materials with coatings are usually the same as tests on uncoated materials. Rotating-bending and oscillating tensile load tests have been used to determine the quality of protection against fatigue fracture that is provided by coatings. Details of these tests can be found in any standard text on engineering materials and so are not described here. The protective effect (or lack of) for any particular coating may vary between different fracture tests so a comparison of different tests would be preferable for a reliable assessment of the coating.

8.10 Summary

Methods of experimental characterisation of materials and/or their coatings can be divided into two basic categories. The first category provides basic physical and chemical data about the coating and its substrate. Data such as hardness, uniformity of composition and physical structure can be measured by various analytical techniques. The second category provides an empirical measure of the usefulness of a coating in a specific service environment. The specific modes of materials degradation, e.g. wear, corrosion and thermal degradation are evaluated by corresponding experimental tests. The two categories are complimentary since although the final evaluation of a coating rests on its service life; there can be no understanding of how coatings function or fail without fundamental data such as hardness, chemical composition and bonding to substrate.

Recommended Reading

W.H.Roberts, Surface Engineering, Proc. Int. Tribology Conference, Melbourne, 2-4 December, 1987 , Publ. Inst. Eng. Australia Conference Publ No 87/18, pp. 438-451.

Ewald Heitz, Rolf Henkhaus and Alfred Rahmel, Corrosion Science :an experimental approach, Transulated by R.B.Waterhouse, NY, Ellis Horwood, 1992.

Peter J. Blau,Friction and wear transitions in materials : break-in, run-in, wear-in, Noyes Publications, 1989, Park Ridge, NJ, USA .

Hartman,W., "A Real Time X-ray Topography System", Real Time Radiologic Imaging: Medical and Inductrial Applications, Ed. Garrett,D.A., and Bracher,D.A., ASTM Special Technical Publication 716, Publ. American Society for Testing of Materials, Philadelphia, 1980, Pp.201-208.

This page is intentionally left blank

PART 3: APPLICATION OF CONTROL TECHNIQUES.

The task of selecting the optimum control technique for materials degradation is not as simple as it may appear. Firstly, there is the problem of specifying the exact form of materials degradation and then there is the task of enhancing the advantages and minimizing the disadvantages of the control technique, which in most cases involves surface engineering.

In practical engineering, the technical characteristics of materials degradation are given less priority than the cost of materials degradation. It is the real cost of materials degradation that should determine an engineer's response to the problems caused by materials degradation. Can the engineer afford to overlook a materials degradation problem or is he/she compelled to apply technical knowledge and produce a solution to the problem? The answer to this question is largely based on financial considerations. In this final part of the book, the application of surface engineering to a specific materials degradation problem is discussed in terms of its financial implications. The practical consequence of applying advanced control techniques, e.g. is there sufficient expertise in the workforce, are also discussed.

CHAPTER 9

CONTROL OF MATERIALS DEGRADATION

9.1 Introduction

The major part of this book has described the wide variety of different forms of materials degradation and a comparable variety of surface engineering technologies. A fundamental question is how to select the appropriate type of surface engineering for any specific application. An issue that is not often discussed by literature on surface engineering is that an unsuitable surface engineering technique can consume resources while failing to suppress the prevailing mode of materials degradation. Most literature on surface engineering deals with the process technology or the microstructure of the coatings produced. The selection of a surface engineering technique should be based on a good understanding of the prevailing mechanisms of materials degradation or else a poor choice will almost certainly result.

Once the optimum form of surface engineering has been selected, the next question is the degree of control that can be obtained. For some degradation problems, e.g. aqueous corrosion, extremely effective control can be obtained which slows the degradation by a factor of 1000 or more. For other types of degradation, e.g. wear at high temperatures, a reduction in wear rate by a factor of 10 is considered to be the limit of economic feasibility.

9.2 Methodology of analysing materials degradation

The underlying causes of materials degradation are rarely obvious and require detailed investigation before accurate understanding is achieved. A classic example of the need to carefully examine a materials degradation problem is provided by the delay in recognizing the true causes of corrosive-abrasive wear. As has been discussed in previous chapters, a synergistic interaction between corrosion and abrasive wear can inflict rapid materials degradation on a metal. For a long period of time, the role of corrosion was overlooked because the metal always appeared shiny without adhering corrosion products such as rust. Signs of abrasion such as scratches were evident on the metal surface and a common interpretation was to specify a harder metal to combat abrasion. In many cases, hardening did not reduce wear and sometimes even caused elevated wear as the hard metal was more easily corroded than the softer metal. A little over 50 years ago, Zelders observed that even if no corrosion products were present, a metal could still be rapidly corroding. Corrosion products were absent from the metal surface because they were immediately removed by wear. Zelders concluded that rapid wear was the result of interaction between corrosion and wear and his work lead to substitution of metals by softer but corrosion resistant polymers for many wear problems.

It can be concluded from the historical example above that the first step in any analysis of materials degradation is a careful compilation of evidence. Vulnerable materials should be identified, e.g. steel either in bulk or as a fraction of a composite. Possible corrosive agents should be noted, e.g. are there any sources of oxygen or nitrogen for high temperature nitride formation. Oxidation can still occur in a nominally inert gas if it is contaminated by even a few parts per million of oxygen. Possibilities for mechanical wear involving not only visible sliding contacts but also nominally stationary contacts should be checked.

Elastic deformation of a structure or looseness between components can generate much unforeseen sliding. The true load history on components should also be known if possible. Very often a structure or component, which sustains a nominally constant load is in fact subject to numerous cycles of fatigue loading. A classic example of this distinction between intended and actual mode of loading is the weighbridge, which is designed for static vehicle loads but in fact sustains fatigue loading each time a vehicle completes a weighing cycle. The significance of thorough analysis in establishing the true cause of materials degradation is summarized graphically in Figure 9.1

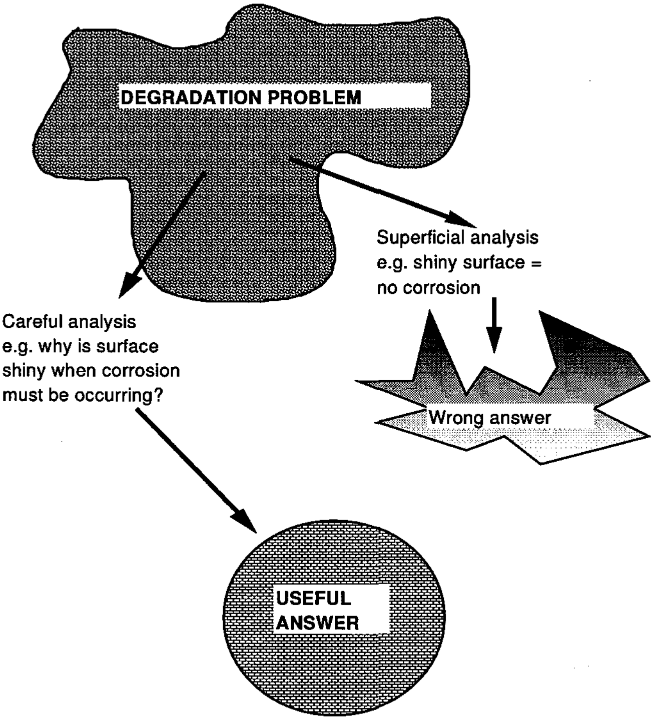


Figure 9.1 Importance of careful analysis in determining the true cause of materials degradation.

Evidence of materials degradation processes such as deposits of corrosion products , wear scars, cracks and fracture faces should be carefully compiled and observed. In many cases it is necessary to use advanced microscopy and surface analysis to adequately investigate the minute traces of corrosion products, wear or fracture debris that are found in practice. Reliable analysis of materials degradation problems depends on the collection of comprehensive evidence from the site of materials degradation. Such evidence should also include accurate records of temperatures, presence of air or other corrosive gases, sliding speeds, loads and any other relevant parameters.

9.3 Selection of optimal surface engineering technology

Once an analysis of the causes of materials degradation are completed, it is usually found that there is one primary mechanism of materials degradation accompanied by one or more secondary mechanisms. A prime example of this general observation is corrosive-abrasive wear in which corrosion is primary and wear secondary. Selection of the appropriate surface engineering should be directed at suppressing the primary mode of materials degradation with lesser concern given to the secondary modes. This is because it is difficult if not impossible to simultaneously optimize for two independent processes and so long as the type of Surface Engineering selected is moderately resistant to the secondary modes of degradation, effective control of materials degradation will be achieved. This is the reason why substitution of rubber for steel as the pipe material for acidic aqueous slurries of sand and other minerals was effective at controlling corrosive-abrasive wear. Rubber has far superior resistance to aqueous corrosion than steel while still maintaining some wear resistance to abrasion. Materials selection that gave priority to wear, e.g. hard alloy steels, was ineffective because of the priority given to the secondary mode of degradation, i.e. wear.

9.4 Control of wear by surface engineering

Most forms of wear can be effectively inhibited by surface engineering, the only exception to this rule is probably wear at extremely high sliding speeds where a compromise has to be accepted between surface melting for low friction and melting wear. Such problems are relatively specialized and are found mainly between an ordnance shell and gun barrel or a turbine blade and casing. The selection of surface engineering follows the basic modes of wear described in Chapters 2 and 5. Most types of surface engineering adequately control several but not all forms of wear and may even promote some wear modes.

9.4.1 Principles of coating selection for wear resistance

For wear protection, the critical qualities of a coating are hardness, non-metallic character and toughness. Hardness is essential for ensuring the mechanical stability of a coating under the high stresses found in a contact.

Abrasive and erosive wear

For abrasive wear, hardness is virtually the controlling parameter that determines the effectiveness of a coating. The hardness of the coating should be at least equal to the hardness of the abrading grits in order to suppress abrasive wear. For erosive wear, toughness as well as hardness determines the ability of the coating to survive and so prevent wear. A minimum thickness of the coating is also essential as the depth of the coating should be comparable to the size of the abrading grits. If the coating is too thin then the substrate immediately below the indentation by the grit suffers high stresses and fails by plastic deformation. The coating usually later detaches by fracture at the interface with the substrate and cracking within the coating. The high hardness and thickness requirements usually necessitate a discrete coating of hard metal or non-metallic compound rather than an integral coating. This finding does not mean that an integral coating for e.g. a metal will not suppress abrasive wear, it means however that the gains in wear resistance will be small compared to a very hard discrete coating.

When a hard, tough coating of sufficient thickness is used, it is possible to reduce abrasive wear to negligible levels compared to the uncoated substrate. Grits composed of extremely hard minerals, e.g. diamonds will however maintain a significant wear rate for almost all coatings.

Adhesive wear

The control of adhesive wear requires a combination of hardness and a non-metallic character to suppress metallic bonding between the sliding surfaces. The coatings can however be relatively thin as in most cases the substrate already has sufficient hardness. The possibility of using thin coatings allows a wider range of coating techniques to be used for preventing adhesive wear than is the case for abrasive and erosive wear. Both discrete and integral coatings are suitable for adhesive wear control. Wear resistance is not the only attribute of coatings that resist adhesive wear, control of the high friction levels associated with adhesive wear is also very important. In many instances, the coating is intended to reduce friction to the level where reliable operation of the machinery can be assured.

Fatigue based wear

The rationale for selecting coatings to resist fatigue-based wear is not as clearly understood as is the case for abrasive and adhesive wear. Sufficient hardness of the coating is needed to resist the subsurface shear stresses which cause fatigue wear but other factors that resist crack growth are also important. A hard, tough coating with excellent adhesion to the substrate is required. As is the case with adhesive wear, a small minimum coating thickness is acceptable. Since in many cases, adhesive and fatigue based wear are closely associated, coating requirements for the two modes of wear are merged. Integral coatings, which do not generate cracks and instead produce internal compressive stresses such as ion implantation, are usually effective for reducing fatigue-based wear. It is difficult to suppress fatigue-based wear to the same extent as other modes of wear because of the persistence of crack growth.

Corrosive and corrosive-abrasive wear

The property of conferring the worn surface with corrosion resistance while maintaining the mechanical strength of the substrate is the essential quality of coatings to resist corrosive and corrosive-abrasive wear. For metals, ceramic and plastic coatings with good adhesion to the substrate and the minimum of pores are suitable for this application. The minimum thickness can be small provided that an impermeable coating is produced. Where the degree of corrosion is high, e.g. wear of components immersed in an acid; the scale of wear resistance achieved by coatings can be very large. If an unsuitable coating is used, the wear may however be worsened because of anodic-cathodic coupling between the coating and substrate.

Fretting wear

Fretting wear is difficult to control or prevent by the use of coatings despite the low sliding speeds and small volume of wear. The problem is that the fretting wear is concentrated over a very small area of the substrate with a reciprocating motion that accentuates mechanical fatigue of the coating. Most coatings rapidly detach from the worn surface during fretting so substrate adhesion is a critical parameter. It appears that if the substrate adhesion is good and hardness is sufficiently high, a coating is very likely to be useful for the prevention of fretting wear. The key factor is durability of the coating inside the fretting contact. Both integral and discrete coatings are effective, integral coatings appear to allow a wider variety of alloying metals to show a positive effect in controlling fretting wear. The coating thickness for fretting wear resistance can be comparatively small

Melting wear and wear in liquid metals

Coatings for these applications are not as well developed as for other modes of wear. Melting wear is suppressed by ensuring that the maximum frictional temperature does not exceed the melting point of the material at the worn surface. Since hardness correlates well with bond strength, which in turn correlates with melting point, hard coatings can be effective for controlling melting wear. The depth of the coating would have to exceed the depth of extreme frictional heating, which for many contacts is typically about 10 μm.

Coatings for wear in liquid metals do not appear to have been studied in great detail. A promising type of coating is to coat a susceptible metal with a resistant metal, e.g. coat stainless steel with titanium for immersion in liquid aluminum. Evidently, the melting point of the metal coating must be higher than that of the liquid metal.

The factors by which a coating can increase wear resistance are not well known for these modes of wear.

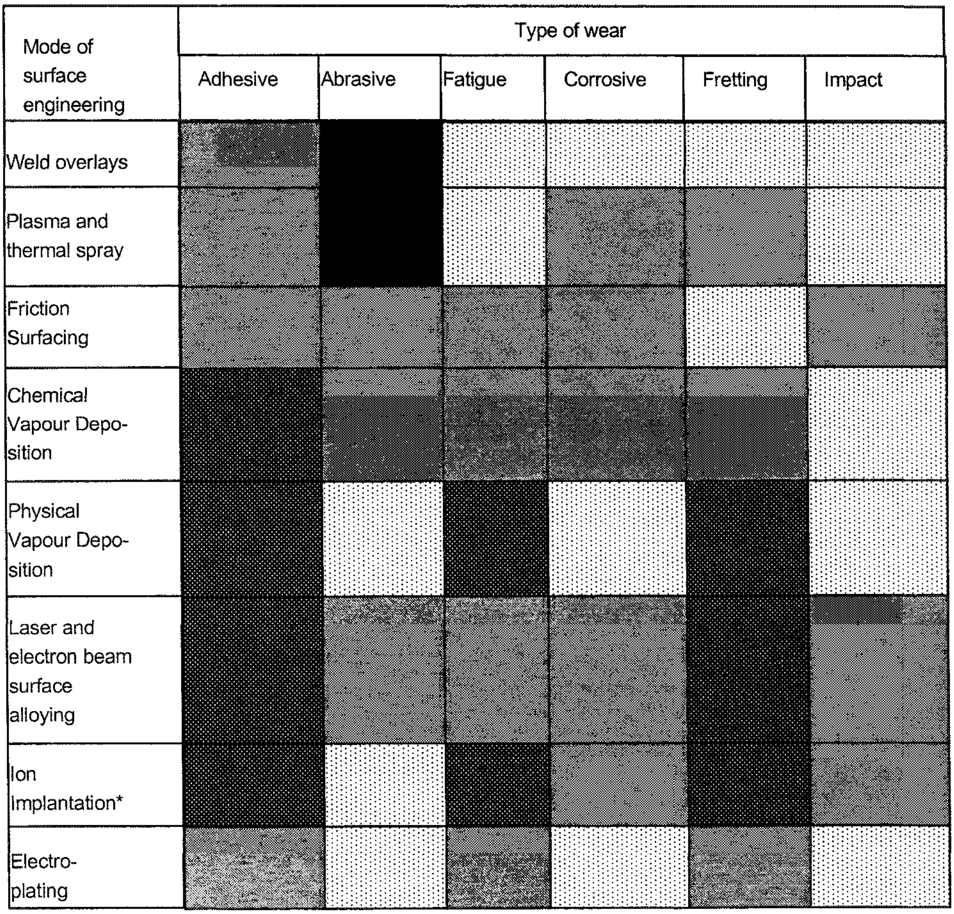
Impact and cavitation wear

Despite the different circumstances where these forms of wear occur, there is a fundamental similarity of impingement rather than sliding motion as the cause of wear. In both cases, wear mechanisms are dominated by fatigue driven crack growth and a tough fatigue resistant coating is most effective for wear control. Impact wear can occur at high temperatures, unlike cavitation wear, which is limited to temperatures below the boiling point of the liquid. At high temperatures in air, oxidation resistance becomes a controlling factor for coating durability against impact wear. For cavitation wear in water and aqueous solutions, corrosion resistance is also important. Discrete coatings are only suitable when adhesion to the substrate is good so that detachment of the coating does not occur. Integral coatings, which induce compressive stress to suppress crack growth, appear to be the most effective. Wear by impact and cavitation is very localized and complete elimination of all sites of localized wear on a component is very unlikely.

9.4.2 Selection of specific surface engineering techniques for specific wear mechanisms

Even when an understanding of the basic principles of coating selection for wear resistance are obtained, the wide variety of surface engineering techniques can still cause confusion. A simple guide, which covers the majority of cases is presented in tabular form in Figure 9.2. The limitations of space necessitate some simplifications so that exceptions to the rules illustrated graphically can always be found.

A basic feature of Fig. 9.2 is the high level of control of abrasive wear by plasma or thermal spraying, which can deposit thick hard coatings on metal surfaces. Other methods of coating such as Physical Vapor Deposition (PVD) are almost always inappropriate for abrasive wear control because only thin coatings are developed



Mode of surface engineering	Type of wear					
	Adhesive	Abrasive	Fatigue	Corrosive	Fretting	Impact
Weld overlays	▨	■	░	░	░	░
Plasma and thermal spray	▨	■	░	▨	▨	░
Friction Surfacing	▨	▨	▨	▨	░	▨
Chemical Vapour Deposition	■	▨	▨	▨	▨	░
Physical Vapour Deposition	■	░	■	░	■	░
Laser and electron beam surface alloying	■	▨	▨	▨	■	▨
Ion Implantation*	■	░	■	▨	■	▨
Electro-plating	▨	░	▨	░	▨	░

*includes plasma nitriding

■ Very effective for most conditions

▨ Effective for some conditions

░ Effective in only a few instances

Figure 9.2. Tabular guide to selection of wear control techniques.

by this means. PVD and ion implantation are effective at controlling adhesive wear with only a very thin coating. This allows coating of precision machinery without dimensional allowance for coating thickness. While adhesive and abrasive wear can now be effectively controlled by coatings, other forms of wear such as corrosive wear, fretting wear, fatigue wear and impact wear can not be so readily prevented. When new coating techniques are developed this trend may be reversed.

9.5 Control of corrosion by surface engineering

A key feature of corrosion, which strongly affects its control, is the large range of temperatures over which corrosion is found. Coating materials, which are effective at low temperatures may not be suitable at high temperatures. The most common example of this problem are paints which cease to be reliable beyond 100°C. As a general rule, organic coatings are cheap and effective at low temperatures while metal and ceramic coatings are required for high temperatures. Where corrosion is caused by liquids or moisture, acidity is also very important. For instance, galvanic zinc coatings are only useful at $pH > 5$. The presence of chemically very active elements such as sulfur place further demands on coatings and necessitate the need for specialized expensive coatings.

An inappropriate choice of coating can worsen a corrosion problem instead of remedying it. Selection of a coating for each of the basic modes of corrosion is discussed below.

Corrosion in neutral or near-neutral water

With the exception of superheated water and steam, the temperature range of aqueous corrosion is confined to near ambient temperatures. With this restricted temperature range, painting is an effective and cheap method of corrosion control. For flowing and impinging water, the mechanical strength of the paint becomes important since the shear stress generated by the viscous flow of the water and impact from entrained particles can fracture paint. Under severe conditions, high strength coatings of epoxies may have to be substituted. Where mechanical abrasion is significant, paints are unsuitable because of their vulnerability to crevice corrosion at scratched areas of paint. Galvanized coatings are usually used for this application since the electropositive galvanizing metal confers cathodic protection to the metal adjacent to scratches. For situations where mechanical abrasion coincides with acidity, galvanization is unsuitable and a thick tough polymer coating should be used instead. Enamel coatings are also suitable for corrosion protection and can withstand some mechanical abrasion and higher temperatures than paints. These features justify their use despite higher cost and unsuitability for coating renewal. Thin coatings of Titanium nitride deposited by Physical Vapor Deposition are currently being evaluated for corrosion combined with abrasion and erosion resistance as titanium nitride combines mechanical strength with resistance to most chemical reagents.

Corrosion in strong alkalis and acids

Strong alkalis and strong acids represent opposite extremes but share some similarities in terms of corrosion control. A common feature is that coatings, which are durable in neutral or near-neutral water, will probably fail rapidly in either strong alkalis or strong acids. For strong acids, non-metallic coatings such as enamel or polymers at low temperatures are suitable. Some acids present major difficulties for corrosion control. Hydrofluoric acid, which can be contained at low temperature by glass (enamel) coatings but not at high temperatures. Sulfuric and nitric acids are strong oxidants and will rapidly degrade polymer coatings. Galvanized coatings are completely unsuitable for strong acids because even if passivation of the metal coating occurs, a scratch on the coating can destroy this passivation.

Strong alkalis are very destructive to glasses and so enamel is unsuitable for protection against alkali corrosion. Metal coatings and polymer coatings (at low temperatures only) can be used for this application. Polymer coatings will sustain some degradation by strong alkalis so their regular replacement or maintenance should be considered.

Oxidative and High temperature corrosion

Oxidation and other forms of high temperature corrosion e.g. sulfidization are only significant at temperatures greater than 200°C which restricts protective coatings to metals, ceramics and glasses. Only a few rare metals such as platinum are highly resistant to oxidation at high temperatures so that metal coating for oxidation control is rarely practiced. A metal or metal alloy, which oxidizes slowly, is selected for an application and the metal is used in bulk form so that there is usually a generous margin for oxidation. Sulfidization is a very rapid form of corrosion, which limits a metal coating to a very short service life unless the coating is so thick that it becomes part of the bulk material. A potentially useful thin metal coating for sulfidization control is a molybdenum-aluminum alloy, which is deposited by Physical Vapor Deposition. The resistance of molybdenum to sulfidization is combined with the resistance of aluminum to oxidization to effectively protect the underlying substrate.

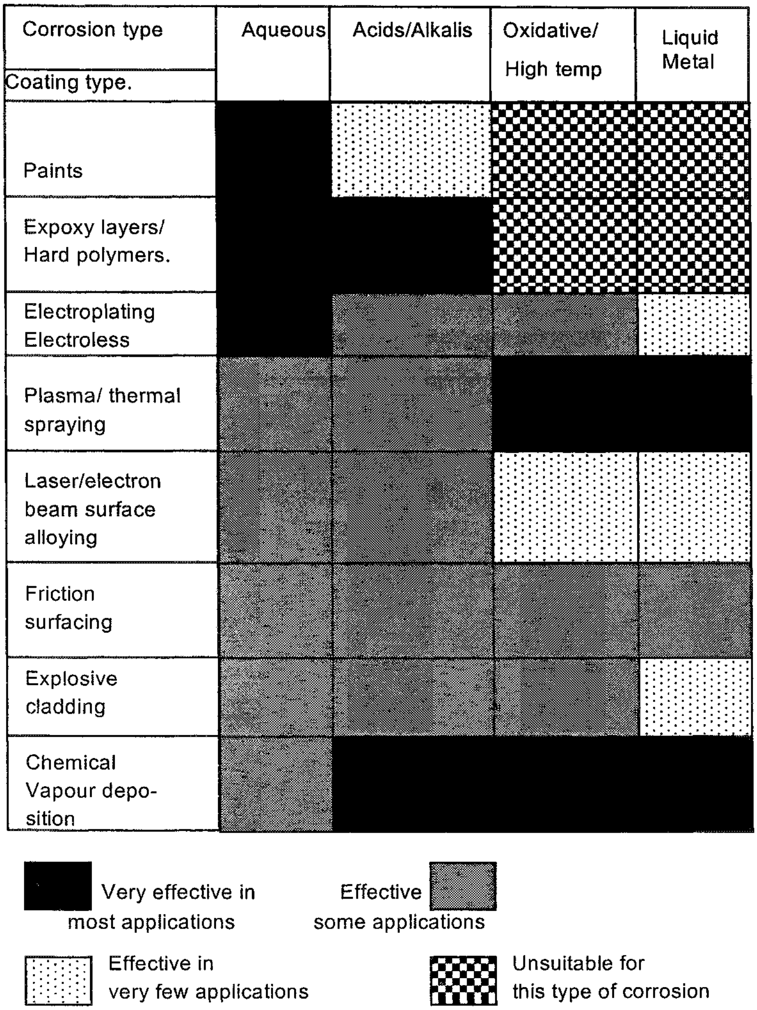
Ceramics are now viewed as very useful materials for coating despite the cost of deposition on metals. Oxide ceramics such as aluminum oxide are very resistant to oxidization and sulfidization while non-oxide ceramics, e.g. silicon carbide and silicon nitride can only resist oxidization for short periods of time. The thermal insulating capacity of ceramics is also a very useful property and thick ceramic layers are not used for what is referred to as 'Thermal barrier' coatings. Thermal barrier coatings protect the substrate against transient high temperatures. Such high temperatures usually coincide with oxidisation damage, which accentuates the usefulness of ceramic coatings.

Corrosion in molten metals and salts

Corrosion in molten metals and salts is a comparatively specialized form of materials degradation and so there are not many suitable coatings. For liquid metal corrosion, a coating of another metal, which is insoluble to the liquid metal, can adequately protect a metal. Coatings similar to those used to control wear by abrasive particles in liquid metal can be used.

Corrosion in molten salts is very rapid like sulfidization and so precludes the use of most coatings. Titanium nitride and titanium carbide coatings do however appear to be resistant to corrosion by e.g. molten lithium carbonate. Most other ceramic coatings, e .g. of silicon carbide suffer from corrosion problems, which probably originate at grain boundaries. Control of corrosion by molten salts is problematic and represents a major research topic.

A tabular guide to the selection of coatings for corrosion control is illustrated in Figure 9.3



Corrosion type Coating type.	Aqueous	Acids/Alkalis	Oxidative/ High temp	Liquid Metal
Paints	■	░	▦	▦
Expoxy layers/ Hard polymers.	■	■	▦	▦
Electroplating Electroless	■	▒	▒	░
Plasma/ thermal spraying	▒	▒	■	■
Laser/electron beam surface alloying	▒	▒	░	░
Friction surfacing	▒	▒	▒	▒
Explosive cladding	▒	▒	▒	░
Chemical Vapour deposition	▒	■	■	■

■ Very effective in most applications

Effective ▒ some applications

░ Effective in very few applications

▦ Unsuitable for this type of corrosion

Figure 9.3. Selection of coatings for control of corrosion.

It can be seen from Fig. 9.3 that while several different types of coatings can control aqueous corrosion, corrosion in more severe environments can only be controlled by a few specialized form of coatings. These coatings are usually more expensive than the general use coatings and so add to the cost of the system that requires corrosion protection.

9.6 Control of fatigue and fracture by surface engineering

The possibility of control or inhibition of fatigue and fracture by surface engineering remains an attractive possibility. If a relatively cheap coating could allow the extension of fatigue life of a component by e.g. 50%, the savings to industry would be very large. Much research has been expended on developing coatings of high strength metals, which under laboratory conditions suppress fatigue and fracture.

The basic principle to note is that there are two modes of fracture and fatigue, surface and sub-surface modes. In the surface mode, extreme roughness, scratching caused by wear, pit formation caused by corrosion provide an initiation site for fatigue crack growth or brittle fracture. In the sub-surface mode, an inclusion, void or internal crack provide the initiation site.

Surface engineering is only effective at suppressing the surface mode of fracture and fatigue because surface coatings and treatments have no effect on sub-surface microstructure. In most cases, the surface mode of fatigue and fracture occurs when components are exposed to harsh environments and sustain moderate loading for long periods of time. Rapid failure under severe loading is usually caused by the subsurface mode. Moderate loading means loads that are much less than the load where spontaneous fracture occurs (e.g. ductile yield). Severe loading implies a higher level of forces that are comparable to the loading at spontaneous fracture.

9.6.1 Discrete surface coatings

Discrete surface coatings can protect a substrate against fatigue or fracture by several means. There is the direct protection conferred by the coating against wear and corrosion, which suppresses the initiation of crack growth. The coating can also be protective if it is in a state of compressive stress (planar to the substrate) while the substrate sustains a tensile stress. Cracks initiated at the surface are inhibited from growing because of positive crack closure forces. A factor limiting the effectiveness of this form of protection is the limiting shear stress at the interface between coating and substrate. If coating adhesion is poor, this limiting shear stress will reduce the maximum compressive stress that can be generated in the coating without its detachment from the substrate. Figure 9.4 illustrates schematically protection by coatings against fatigue.

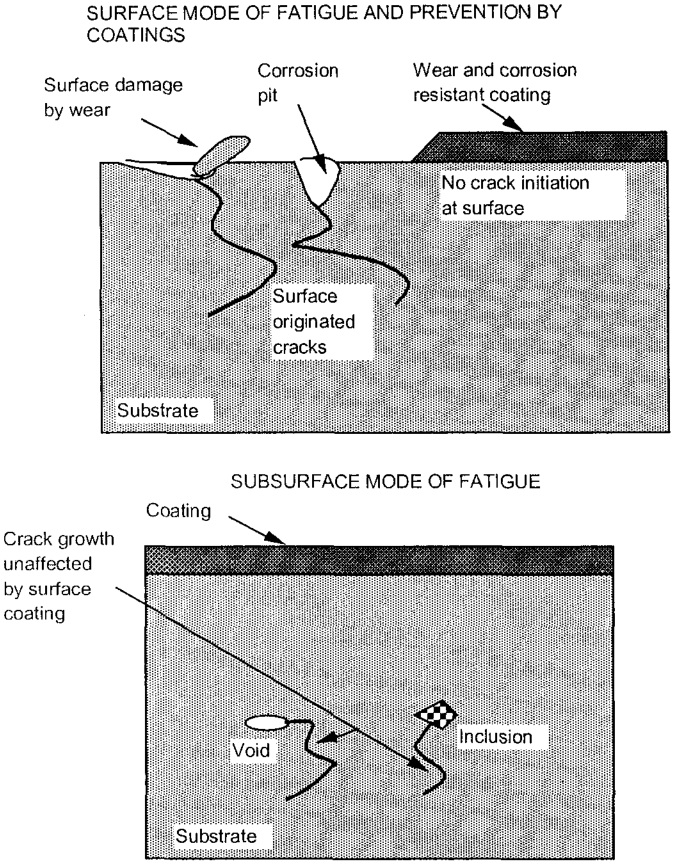


Figure 9.4. Mechanisms of protection by coatings against mechanical fatigue.

In general, tensile residual stresses are found in most artificially deposited thin coatings. The precise causes of tensile residual stress are complex and not fully understood except for a few test cases. The underlying reason is that coating formation is a form of condensation where shrinkage of the material to form a dense hard coating is involved. For some specific cases however, significant compressive residual stresses are observed in coatings. The occurrence of compressive residual stress depends on the coating material, the coating process and the coating temperature. For example, a sputtered tungsten film displays a large compressive residual stress close to - 1500 MPa when deposited at ambient temperature. If the temperature is raised to 300°C, there is no residual stress and if the temperature is raised further to 400°C, a tensile residual stress of approximately 1000 MPa is generated. The degree of crystallinity of the coating is also a significant factor. When silicon is deposited by Chemical Vapor Deposition at temperatures below 600°C, amorphous silicon is formed which has a compressive residual stress. At 600°C, crystalline silicon is formed and instead a large tensile residual stress is generated. The reason for this may be due to the fact that amorphous silicon is less dense than crystalline silicon. Residual stresses in a coating are always accompanied by an opposing residual stress in the substrate. While a compressive residual stress in the coating would suppress initial crack propagation, the corresponding tensile residual stress in the substrate would accelerate later crack propagation. It is important to ensure sufficient coating thickness so that cracks remain within the coating for a sufficient period of time. A thick coating with moderate compressive residual stress may be more useful than a thin coating with high compressive residual stress.

Protection against fracture follows similar considerations to fatigue except that the level of stress in fracture is generally higher than fatigue. This means that crack closure by induced compressive stress is less likely to be effective since the higher tensile stress of fracture will neutralize the induced compressive stress.

A further means of using discrete coatings to control fatigue failure is to repeatedly deposit the same coating on a loaded component. Each time a coating is deposited, the crack nucleation structure (formed early in the fatigue process) is covered in new material and ceases to be able to generate a crack. Copper specimens, coated with nickel at regular intervals (in this case, 25% of expected fatigue life) achieved virtually indefinite fatigue life [McKinnon]. However, the inconvenience of removing a loaded component for repeated coating has limited the application of this principle.

9.6.2 Integral coatings and surface modified layers

Integral coatings and surface modified layers offer some very interesting possibilities for control of fatigue and fracture. A traditional form of surface modification, shot peening, is useful for suppressing fatigue. As described in Chapter 7, shot peening generates surface compressive residual stress, which inhibits crack propagation. A technique closely related to shot peening; grit blasting is also effective for fatigue inhibition. Grit blasting involves a stream of water containing hard particles of alumina or other minerals, which impinge against the surface to be treated. The speed of impact of the alumina particles is slower than for the steel balls of shot peening in order to prevent erosive wear by the sharp-edged particles. It is found that there is an optimum impact speed of the alumina particles for maximum extension of fatigue life, at higher impact speeds, the surface damage caused by particle impingement creates an excessive number of fatigue crack initiation sites. The extension of fatigue life for a specific stress amplitude can be as high as a factor of 10 compared to untreated material. Both shot peening and grit blasting suppress fatigue at the initiation stage only, as soon as a crack grows longer than 1 mm, no further protection against fatigue is offered.

More advanced techniques of generating surface modified layers may also be useful for fatigue control. Ion implantation and ion plating are based on the forcible insertion of coating atoms into the substrate. This process of atom insertion generates large compressive stresses in the surface of the substrate, which promote crack closure. The bonding between integral coatings and the substrate is much higher than for discrete coatings. It is therefore possible to reach a much higher compressive stress without damage to the coating. The only risk is of subsurface cracking inside the substrate due to reactive tensile forces. Integral coatings based on surface alloying in a liquid state, e.g. laser and electron beam surface alloying, are liable to surface cracking because of tensile stresses caused by solidification of the surface alloyed layer.

This type of integral coating appears to be unsuitable for fracture control unlike ion implantation. A major limitation of ion implantation is the small depth of the implanted layer, which is typically about 1 micrometer. Crack growth beyond 10 µm depth below the surface is unlikely to be suppressed by ion implantation. Components, which are very thin or extremely small, are most likely to show improved fatigue lives with ion implantation.

9.7 Summary

Careful analysis of the true causes of materials degradation is essential to prevent wastage of time and resources on inappropriate control methods. Each mechanism of corrosion, wear or fracture has a different response to particular coating technologies. An unsuitable coating technology will give little protection and may even aggravate materials degradation. A comprehensive knowledge of the range of application of each coating technology is therefore essential for effective surface engineering.

RECOMMENDED READING

E.A. McKinnon, Journal of Testing and Evaluation, Vol. 4, 1976, page 57.

H.G. Zelders, La Corrosion Superficielle dans le Circuit de Lavage des Charbonnages des Mines de l'Etat Neerlandais, *Metaux et Corrosion*, 1949, Vol. 65, pp. 25-76.

CHAPTER 10

FINANCIAL AND INDUSTRIAL ASPECTS OF MATERIALS DEGRADATION AND ITS CONTROL

10.1 Introduction

So far in this treatment of materials degradation, the scientific understanding of degradation processes and control technologies have been discussed. The practicing technologist cannot however exclusively consider the technical aspects of materials degradation but must consider economic, ethical and legal aspects of materials degradation before the optimal solution to a specific problem can be devised. Control of wear and corrosion involves a comprehensive analysis that achieves the greatest reduction in costs of corrosion and wear consistent with legal and ethical requirements. One example being, a scarcely used sidewalk on a bridge may be costly to paint, but the sidewalk is however painted just in case somebody uses the sidewalk. If the sidewalk were not painted, that unfortunate person may fall through rusted metal on the sidewalk. To adequately understand the analysis of corrosion and wear control, a familiarity with basic financial concepts and ethics is necessary.

10.2 Financial analysis of materials degradation control

Detailed financial analysis of materials degradation is still at an elementary level as there is no fundamental theory that equates the large but variable costs of materials degradation with the costs of its control. Two basic methods of financial analysis are presented here; (i) an options model of the choice to control or neglect materials degradation, (ii) and evaluation of the financial benefit in controlling depreciation (this is the cost of materials degradation).

10.2.1 Options analysis of financial gain

Any anticipated corrosion or wear problem presents a technologist with an option to either ignore the problem or initiate measures to control the problem. Both choices of the option involve costs where the option to ignore entails direct materials degradation costs in terms of loss of material, loss of reliability and total failure. The option choice to initiate control measures may involve substantial costs, e.g. the dry-docking of a ship and repainting of its hull is costly with extra costs due to loss of profitable revenue from the working ship.

The value of protective measures can be summarized in the following equation.

Value of protective measures = Cost savings of reduced materials degradation - Cost of protective measures. {10.1}

Now according to the theory of finance, value is time dependent and in assessing any expenditure of funds, present value should be considered which leads to a modified form of 10.1 presented below.

Present value of protective measures = Present Value of cost savings of reduced materials degradation - Present Value of Cost of protective measures. {10.2}

The concept of Present Value involves discounting of cash-flows at a required rate of return where immediate cash-flows have higher values than cash-flows of the same nominal value at a later date. If materials degradation and its control are modeled as (i) a large initial cost for e.g. repainting a ship and (ii) a series of annual savings spread over a number of years then the value of protective measures is defined by the following equation.

$$\mathrm{VALUE} = -\mathrm{COST} + \mathrm{SAVINGS}^{*}\, A(N,R) \qquad \{10.3\}$$

Where , VALUE = Present value of protective measures
COST = Initial cost of protective measures,
SAVINGS = Annual savings of reduced materials degradation.
A= Annuity
N = Number of years
R = Required rate of return, e.g. 0.1 or 10% per annum.

The numerical value of an annuity is defined by the equation:

$$A(N,R) = (1 - (1+R)^{\wedge}\text{-}N)/R \qquad \{10.4\}$$

The value of A(N,R) defines the ratio of present value of all the annual savings to the value of just one annual saving. For example if $N = 10$ and $R = 0.05$ then $A(N,R) = (1-(1.05)^{\wedge}\text{-}10)/.05 = 7.72$. The value of this annuity, 7.72 is less than 10, which is the value in the absence of any capital discounting. This means that the value of the savings over 10 years is 7.72 times the annual savings not 10 times the annual savings. Large values of R always reduce the value of an annuity for any given value of N.

Suppose that the annual savings last for very many years so that the expression for A(N,R) can be approximated by the following simplification of 10.4.

$$A(N,R) = 1/R \text{ for large } N \qquad \{10.5\}$$

For constant annual savings over a long period of time, equation 10.3 will reduce to:

$$VALUE = -\,COST + SAVINGS/R \qquad \{10.6\}$$

Choice of the option whether to control or ignore materials degradation is based on the premise that only actions which generate a positive present value are appropriate. The minimum level of annual savings to generate a positive present value of materials degradation control over a large number of years is :

$$SAVINGS > COST{*}R \qquad \{10.7\}$$

Where a smaller number of years are involved, the minimum level of savings is dependent on the reciprocal of the annuity:

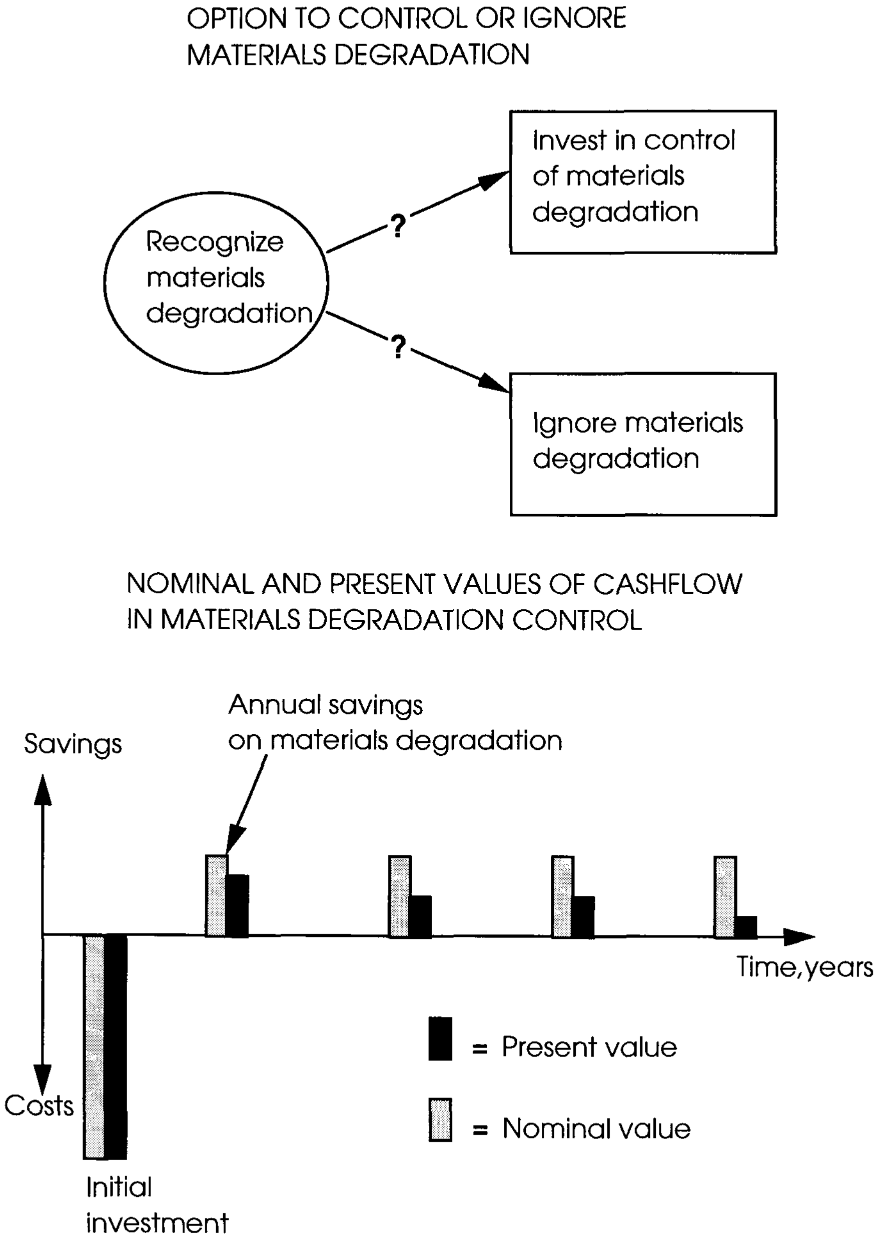
$$SAVINGS > COST/\,A(N,R) \qquad \{10.8\}$$

It can be seen that the minimum annual savings for viable materials degradation control increases in proportion with the required rate of return.

The relative value of the options to ignore or control materials degradation are illustrated schematically in Figure 10.1

How can we estimate the savings made when protective measures are applied to a system. One method of estimation is to consider the probability of systems failure caused by a specific mode of materials degradation, e.g. wear, and the cost of the failure. For a specific time period, the value of savings is equal to the product of the reduction in failure probability and the cost of the failure. This can be expressed algebraically as:

$$SAVINGS = (PROB(U) - PROB(P)){*}COST(F) \qquad \{10.9\}$$



OPTION TO CONTROL OR IGNORE MATERIALS DEGRADATION

NOMINAL AND PRESENT VALUES OF CASHFLOW IN MATERIALS DEGRADATION CONTROL

Figure 10. 1 Comparison of Present value of materials degradation control measures and the initial cost of control measures.

Where PROB(U) is the probability of failure when unprotected, PROB(P) is the probability of failure when protected, COST(F) is the cost of failure.

In many situations, one protection measure protects against several modes of materials degradation. In this case the total value of savings is the sum of savings on specific modes of materials degradation, e.g. for a wear and corrosion resistant coating. This can be described symbolically as:

SAVINGS(TOTAL) = SAVINGS(MODE1) +
SAVINGS(MODE2) + ...+ SAVINGS(MODE##)

{10.10}

In other cases, not only failure is involved but also perennial costs due to loss of system efficiency induced by materials degradation. A typical example is blow-by or loss of compression in an engine that is caused by crankcase cylinder wear. In this case COST(F) is substituted by COST(M) where COST(M) is the cost of malfunction. In the case of blow-by, this would be increased consumption of fuel oil and lubricant. Possible fines from excessive air pollution might also be relevant in this example.

As an example of the scale of costs involved, an industrial vacuum coating system is installed to produce corrosion resistant coatings on components. The system costs approximately US$ 1 million and has an annual running cost of approximately US$100,000. For the purposes of argument, the vacuum system is assumed to last indefinitely with financial viability defined according to equation 10.6. If the required rate of return is 15% per annum then the minimum annual savings for a Net Present value greater than zero is US$250,000, i.e., $100,000 + 15% of $1,000,000. If the vacuum coating system is used for 250 days a year and 100 components are coated each day, then the minimum saving on each component is US$250,000/(250 *100) = US$10.

What does a figure of US$10 minimum annual savings mean in terms of the cost of the component? Suppose the component is planned to last for 10 years in a corrosive environment that would destroy the component in less than one year without corrosion protection. Without performing the full financial calculations, it can be seen that the component and its installation should cost at least $10. This means that only valuable components are worth protecting or else, the component is very tedious to replace. An example of the latter might be a nut and bolt at the bottom of a deep well or on the seabed.

The required rate of return is controlled by the financial structure of a company or government institution not materials degradation parameters. The value of any materials degradation control procedure is not absolute but relative to the prevailing financial conditions. As may have been already appreciated from this treatment of materials degradation, estimation of rates of materials degradation is complex. Even if by sophisticated numerical modeling, an estimate of the amount and location of degraded material could be obtained, this information would not provide conclusive information about whether to control or ignore materials degradation. To reach this conclusion, it is necessary to follow the model given above or a more refined model to evaluate the appropriate response to materials degradation. In general, the more risky a company's business is, the higher its required rate of return is and the more likely that it is profitable to ignore materials degradation. The decline in value of an investment in materials degradation with increase in required rate of return is illustrated schematically in Figure 10.2.

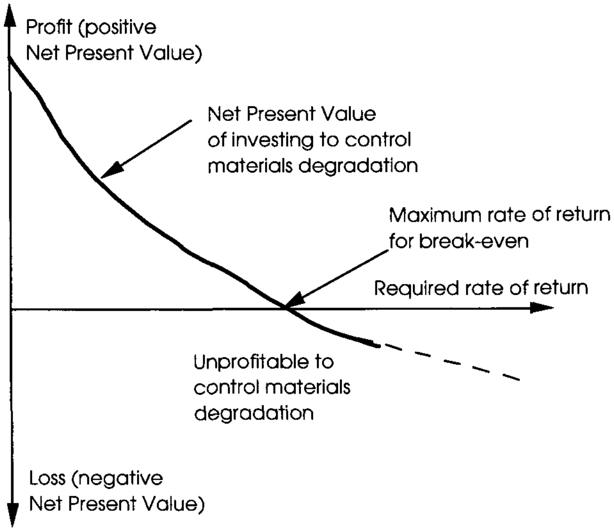


Figure 10.2. Dependence of value of materials degradation control on the required rate of return.

In a more detailed analysis of the available options to control materials degradation, the differences between cash flows of materials degradation control techniques become significant. For instance, with an aqueous metal corrosion problem that can be controlled by cathodic protection, there is a choice between the option of installing an impressed current system or using sacrificial anodes. The impressed current system has a higher capital cost than the sacrificial anodes but lower recurrent costs as the need to periodically replace the sacrificial anodes is avoided. The Net Present Value of sacrificial anode replacement can be modeled as an annuity just like annual savings on corrosion. Sacrificial anodes and impressed current systems offer very similar levels of protection against aqueous corrosion thus the corrosion savings from either of these control methods are modeled as equal. The Net Present Value of the sacrificial anode replacement (which is negative) will decline in the same proportion as annual corrosion savings decline when the required rate of return is increased. The initial capital cost of the impressed current system will not be reduced by increase in the required rate of return so that with any elevation in the required rate of return, the use of sacrificial anodes gains in financial viability at the expense of impressed current systems.

10.2.2 Modeling financial benefit in terms of reduced depreciation

Another model of the viability of materials degradation control, which does not involve financial calculations, is to look at the planned lifetime of a device or structure. In most cases, planned technical obsolescence will impose severe limits on the useful life of any engineering system. For example, a typical service lifetime is 5 years or less for a computer and 10 years for a car. Materials degradation control is useful only in so far as the functional life can be extended as far as the design life. Further extensions of working life by more advanced, costly materials degradation control are usually unjustified.

Estimation of materials degradation costs in terms of rate of depreciation can also be modeled in terms of discounted cash flows and Net Present Value. The financial consequence of prolonging the lifetime of equipment by control of materials degradation is to delay the time when replacement becomes necessary or else reduce the frequency of replacement when replacement occurs on a regular basis. Each company or industrial enterprise has its own required rate of return, which is the in-house discounting factor, e.g. 10% per annum. The required rate of return varies with the type of business and is usually higher for industries with short product lifetimes than for utilities where the product e.g. electricity is virtually unchanging. At high required rates of return, the financial benefit of reducing replacement frequency or delaying replacement is much less than at low required rates of return.

Technical standards societies have given some guidance to the financial analysis of materials degradation. The American Society of Materials has introduced the concept of Asset Loss Risk where the components of a system most likely to fail are identified and their replacement cost assessed.

10.2.3 Implications of behavioural finance

Another factor in the decision whether or not to invest in technology to control materials degradation, is human perception of cost and benefit. The field of study, which looks at the effect of human perception on financial issues, is called Behavioral Finance. In Behavioral Finance, it is found that a tangible cost is perceived far more acutely than an intangible gain. The tangible cost in this context is the high cost of special coatings, while the intangible benefits are 'savings' obtained by improved reliability or higher operating efficiency. In many industrial establishments, precise records for component lifetimes and consumption of energy or raw materials do not exist. The plant engineer is therefore unable to objectively determine whether surface engineering does generate benefits for the specific application. A large, advanced enterprise, which enjoys a large, almost unvarying demand for its products and where its process parameters are carefully recorded by a team of expert engineers, is the optimum type of enterprise for introducing advanced technology for controlling materials degradation. It is generally observed that a company will only invest in surface engineering or any other materials conservation technology if its own staff has found an unequivocal benefit. A smaller company where the plant engineers only ensure that the process machinery functions adequately and where demand for products is highly variable, will be far more unlikely to comprehensively apply surface engineering. There is no financial justification for conserving machinery by surface engineering if there are significant idle periods for the machinery. Another major rule of behavioral finance is that money for any corporate body or individual is divided into separate accounts where transfer between accounts in discouraged even though money can actually be transferred freely. Expenditure on materials degradation technology either means that the account for plant maintenance must be increased or else a new account has to be created. In both cases, a manager must propose the appropriate changes in accounts to the board of directors. If there is no reliable means of proving that this expenditure is balanced by a benefit or cost reduction, the manager is unlikely to propose the extra expenditure. In practice, a few advanced companies will vigorously apply measures to control materials degradation because of their precise monitoring systems. The majority of smaller companies will probably later adopt a surface engineering technique because of a fear of losing competitiveness.

10.3 Practical aspects of implementing new forms of surface engineering

Practical considerations often influence the final choice of surface engineering technology as much as technical evaluation of protection against materials degradation. A basic concern is the size of the component to be protected. Large components or structures that have a dimension greater than one metre are very costly to coat under vacuum, as a very large vacuum chamber is required. Conversely, plasma spraying of miniature components is impractical because of the size of the plasma spray. The level of skill required by technical staff for reliable quality of surface engineering should also be evaluated. Hazards such as emission of toxic substances, from electro-plating, high electrical voltages and microwave radiation from sputtering or X-ray radiation from electron beams must be carefully controlled to remove any danger to technical personnel. Large vacuum chambers require elaborate safety interlocks to prevent personnel from being trapped inside the chamber when the pumps begin extracting the air. In many instances, surface engineering procedures are fitted into the overall production process for a component where it is important to match the speed of the surface engineering process with the speed of production. The production rate will be restricted if the speed of the surface engineering process is less than the production rate without surface engineering. Speed and labor requirements still limit most of the advanced surface engineering components to only small numbers of high value or critical components.

10.4 Evaluation of control of materials degradation in terms of value to the company or institution

As described in 10.1, evaluation of a project or new practice is determined by financial considerations in most companies. A common problem in this evaluation is to include all the hidden costs and benefits as well as the apparent costs and benefits. Difficulties in determining hidden costs and benefits often lead company management to adopt a strategic approach where judgement instead of precise financial calculation is used to decide whether to proceed with a project which in this context relates to surface engineering. Examples of direct costs are; extra factory space for the surface engineering equipment, cost of training to learn how to operate the new equipment, extra power and water consumption. Examples of hidden costs or benefits are changes in marketability and new applications for a product enhanced by surface engineering. There is still no quantitative theory or model to predict the monetary value of materials degradation controlled by surface engineering. It is therefore necessary to learn either by empirical test or actual experience the effectiveness of surface engineering at enhancing company profitability.

In most cases, confident prediction of financial benefit is not the prime reason for developing surface engineering or other advanced control measures for materials degradation. As mentioned above in Behavioral Finance, most technological companies feel compelled to adopt new technologies such as surface engineering in order to ensure that their products do not become old-fashioned or obsolescent compared to the products of other companies. More stringent industrial standards and legal requirements may also force companies to adopt new surface engineering technologies before the companies can perceive any benefits.

10.5 Summary

The practical and financial aspects of controlling materials degradation are discussed in this chapter. The financial implications of materials degradation can be evaluated in terms of the comparative costs of controlling materials degradation or ignoring it. The cost of materials degradation is not only determined by the physical mechanisms involved but also by the financial structure of a company and its type of business. For the same corrosion problem, the optimum solution may vary with the enterprise under consideration. Practical considerations such as component size, level of staff skills, marketing of products are all relevant in a decision to apply surface engineering for materials degradation control. The lack of precise data concerning the costs of materials degradation and the benefits of its control also affects the final decision.

APPENDIX 1: List of some relevant standards used for testing corrosion, coatings and wear

Topics	Standard used	Equivalents available in other standards	Test title
Corrosion	ASTM G15		Terminology relating to corrosion and corrosion testing
	ASTM G1	ISO/DIN/AFNOR /BSI	Practice for preparing, cleaning and evaluating corrosion test specimens
	ASTM F1110		Test method for sandwich corrosion test
	ASTM F1089		Test method for corrosion of surgical implants
	ASTM D4350		Test method for corrosivity index of plastics and fillers
	ASTM D3310		Test method for determining corrosivity of adhesives materials
	ASTM D2251		Test method for metal corrosion by halogenated organic solvents and their admixtures
	ASTM D2649	ISO	Test method for corrosion characteristics of solid film lubricants
	ASTM D1748		Test method for rust protection metal preservatives in the humidity cabinet
	ASTM D1654	ISO/DIN/AFNOR /BSI	Test method for painted or coated specimens subjected to corrosive environments
	ASTM D1374		Test method for aerated total immersion corrosion test for metal cleaners
	ASTM B117	ISO/DIN/AFNOR /BSI	Test method of salt spray (Fog) testing
	ASTM B457		Test method for measurement of impedance of Anodic coatings on Aluminum
	ISO 10271-2001		Dental metallic materials -- Corrosion test methods
	ISO 11845-1995		Corrosion of metals and alloys -- General principles for corrosion testing

	ISO 11782-1-1998		Corrosion of metals and alloys -- Corrosion fatigue testing -- Part 1: Cycles to failure testing
	ISO 11782-2-1998		Corrosion of metals and alloys -- Corrosion fatigue testing -- Part 2: Crack propagation testing using precracked specimens
Coatings for corrosion & wear resistance	ISO 2079:1981		Surface treatment and metallic coatings -- General classification of terms
	ISO 2081:1986		Metallic coatings -- Electroplated coatings of zinc on iron or steel
	ISO 2082:1986		Metallic coatings -- Electroplated coatings of cadmium on iron or steel
	ISO 2093:1986		Electroplated coatings of tin -- Specification and test methods
	ISO 2177:1982		Metallic coatings -- Measurement of coating thickness -- Coulometric method by anodic dissolution
	ISO 2178:1985	ASTM (B244-97)	Non-magnetic coatings on magnetic substrates -- Measurement of coating thickness -- Magnetic method
	ISO 2360:1982		Non-conductive coatings on non-magnetic basis metals -- Measurement of coating thickness -- Eddy current method
	ISO 2361:1982		Electrodeposited nickel coatings on magnetic and non-magnetic substrates -- Measurement of coating thickness -- Magnetic method
	ISO 2819:1980		Metallic coatings on metallic substrates -- Electrodeposited and chemically deposited coatings -- Review of methods available for testing adhesion
	ISO 3497:2000		Metallic coatings -- Measurement of coating thickness – X-ray spectrometric methods
	ISO 3543:2000		Metallic and non-metallic coatings -- Measurement of thickness -- Beta backscatter method
	ISO 10308:1995		Metallic coatings -- Review of porosity tests
	ISO 10309:1994		Metallic coatings -- Porosity tests -- Ferroxyl test

	ISO 9211-1-4:1994		Optics and optical instruments -- Optical coatings -- Part 1: Definitions Optics and optical instruments -- Optical coatings -- Part 2: Optical properties Optics and optical instruments -- Optical coatings -- Part 3: Environmental durability Optics and optical instruments -- Optical coatings -- Part 4: Specific test methods
	ISO 9588:1999		Metallic and other inorganic coatings -- Post-coating treatments of iron or steel to reduce the risk of hydrogen embrittlement
	ISO 9717: 1990		Phosphate conversion coatings for metals -- Method of specifying requirements
	ISO 10074:1994		Specification for hard anodic oxidation coatings on aluminium and its alloys
	ISO 10111:2000		Metallic and other inorganic coatings -- Measurement of mass per unit area -- Review of gravimetric and chemical analysis methods
	ISO 1462		Metallic coatings – coatings other than those anodic to the basis metal - Accelerated Corrosion tests – Method of evaluation of the test results
	ISO 4536		Metallic and Non Organic coatings on metallic substrates – Saline droplets Corrosion Test (SD test)
	ISO4538		Metallic coatings – Thioacetamide Corrosion test (TAA test)
	ISO4540		Metallic coatings – Coatings Cathodic to the substrate - Rating of Electroplated specimens subjected to corrosion tests
	ISO 4541		Metallic and other non organic coatings – Corrodkote test Corrosion test (CORR test)
	ISO 4541		Metallic and other non organic coatings – General rules for stationary Outdoor exposure corrosion tests
	ISO 4543		Metallic and other non organic coatings – General rules for Corrosion tests Applicable for storage conditions

	ISO 10289-1999		Methods for corrosion testing of metallic and other inorganic coatings on metallic substrates -- Rating of test specimens and manufactured articles subjected to corrosion tests
	ISO 11997-1-2000		Paints and varnishes -- Determination of resistance to cyclic corrosion conditions -- Part 1: Wet (salt fog)/dry/humidity
	ISO 11997-2-2000		Paints and varnishes -- Determination of resistance to cyclic corrosion conditions -- Part 2: Wet (salt fog)/dry/humidity/UV light
	ISO 11998:1998		Paints and varnishes -- Determination of wet-scrub resistance and cleanability of coatings
	ISO 12686:1999		Metallic and other inorganic coatings -- Automated controlled shot-peening of metallic articles prior to nickel, autocatalytic nickel or chromium plating, or as a final finish
	ISO 10713:1992		Jewellery -- Gold alloy coatings

	ISO 12944-1-8:1998		Paints and varnishes -- Corrosion protection of steel structures by protective paint systems -- Part 1: General introduction Paints and varnishes -- Corrosion protection of steel structures by protective paint systems -- Part 2: Classification of environments Paints and varnishes -- Corrosion protection of steel structures by protective paint systems -- Part 3: Design considerations Paints and varnishes -- Corrosion protection of steel structures by protective paint systems -- Part 4: Types of surface and surface preparation Paints and varnishes -- Corrosion protection of steel structures by protective paint systems -- Part 5: Protective paint systems Paints and varnishes -- Corrosion protection of steel structures by protective paint systems -- Part 6: Laboratory performance test methods Paints and varnishes -- Corrosion protection of steel structures by protective paint systems -- Part 7: Execution and supervision of paint work Paints and varnishes -- Corrosion protection of steel structures by protective paint systems -- Part 8: Development of specifications for new work and maintenance
	ISO 8251:1987		Anodized aluminium and aluminium alloys -- Measurement of wear resistance and wear index of anodic oxidation coatings with an abrasive wheel wear test apparatus
	ISO 8252:1987		Anodized aluminium and aluminium alloys -- Measurement of mean specific abrasion resistance of anodic oxidation coatings with an abrasive jet test apparatus
	ISO 14713:1999		Protection against corrosion of iron and steel in structures -- Zinc and aluminum coatings - Guidelines
	ISO 14921:2001		Thermal spraying -- Procedures for the application of thermally sprayed coatings for engineering components

	ISO 15724:2001		Metallic and other inorganic coatings -- Electrochemical measurement of diffusible hydrogen in steels -- Barnacle electrode method
	ISO 15730:2001		Metallic and other inorganic coatings -- Electropolishing as a means of smoothing and passivating stainless steel
	ISO 15741:2001		Paints and varnishes -- Friction-reduction coatings for the interior of on- and offshore steel pipelines for non-corrosive gases
	ASTM C633		Test Method for Adhesion or Cohesive Strength of Flame-Sprayed Coatings
Wear	ASTM G 40-99		Termninology Relating to Wear and Erosion
	ISO 6601:1987		Plastics -- Friction and wear by sliding -- Identification of test parameters
	ASTM G 77-98		Standard Test Method for Ranking Resistance of Materials to Sliding Wear Using Block-on-Ring Wear Test
	ASTM G 119-93 (1998)		Standard Guide for Determining Synergism Between Wear and Corrosion
	ASTM G 65		Test Method for Measuring Abrasion Using the Dry Sand/Rubber Wheel Apparatus
	ASTM G75		Test Method for Determination of Slurry Abrasivity (Miller Number) and Slurry Abrasion Response of Materials (SAR Number)
	ASTM G 83		Test Method for Wear Testing with a Crossed-Cylinder Apparatus
	ASTM G 99		Test Method for Wear Testing with a Pin-on-Disk Apparatus
	ASTM G 132		Test Method for Pin Abrasion Testing
	ASTM G 105		Test Method for Conducting Wet Sand/Rubber Wheel Abrasion Tests
	ASTM F2205-00		Standard Practice for Gravimetric Measurement of Polymeric Components for Wear Assessment
	ASTM F 732-00	ISO	Standard Test Method for Wear Testing of Polymeric Materials for Use in Total Joint Prostheses

	ASTM F 1714		Guide for Gravimetric Wear Assessment of Prosthetic Hip-Designs in Simulator Devices
	ASTM F 1715		Guide for Gravimetric Wear Assessment of Prosthetic Knee-Designs in Simulator Devices
	ASTM F 1426-94 (1999)		Standard Practice for Identifying Tire Tread Surface Irregular Wear Patterns Resulting from Tire Use
	ASTM B 611-85 (2000) e1		Standard Test Method for Abrasive Wear Resistance of Cemented Carbides
	ASTM B 794-97		Standard Test Method for Durability Wear Testing of Separable Electrical Connector Systems Using Electrical Resistance Measurements
	ASTM C 501-84 (1996)		Standard Test Method for Relative Resistance to Wear of Unglazed Ceramic Tile by the Taber Abrader
	ASTM C 808-75 (2000)		Standard Guideline for Reporting Friction and Wear Test Results of Manufactured Carbon and Graphite Bearing and Seal Materials
	ASTM C1243 – 93 (1999)		Standard Test Method for Relative Resistance to Deep Abrasive Wear of Unglazed Ceramic Tile by Rotating Disc
	ASTM D2625-94 (1998)		Standard Test Method for Endurance (Wear) Life and Load-Carrying Capacity of Solid Film Lubricants (Falex Pin and Vee Method)
	ASTM D2266-01		Standard Test Method for Wear Preventive Characteristics of Lubricating Grease (Four-Ball Method)
	ASTM D 3181-95		Standard Guide for Conducting Wear Tests on Textiles
	ASTM D 6425-99		Standard Test Method for Measuring Friction and Wear Properties of Extreme Pressure (EP) Lubricating Oils Using SRV Test Machine
	ISO 4378-2:1983		Plain bearings -- Terms, definitions and classification -- Part 2: Friction and wear
	ISO 8251:1987		Anodized aluminum and aluminum alloys -- Measurement of wear resistance and wear index of anodic oxidation coatings with an abrasive wheel wear test apparatus

	ISO 9183-1		Tools for pressing -- Wear plates for press dies -- Part 1: Type A
	ISO 9183-2		Tools for pressing -- Wear plates for press dies -- Part 2: Type B
	ISO 10825		Gears -- Wear and damage to gear teeth -- Terminology
	ISO/TS 14569-1		Dental materials -- Guidance on testing of wear -- Part 1: Wear by toothbrushing
	ISO/TS 14569-2		Dental materials -- Guidance on testing of wear -- Part 2: Wear by two- and/or three body contact
	ASTM G 133-95		Standard Test Method for Linearly Reciprocating Ball-on-Flat Sliding Wear
	ASTM D3702		Test Method for Wear Rate of Materials in Self-Lubricated Rubbing Contact Using a Thrust Washer Testing Machine
	ASTM D 2714		Test Method for Calibration and Operation of the Falex Block-on-Ring Friction and Wear Testing Machine

Acronym used

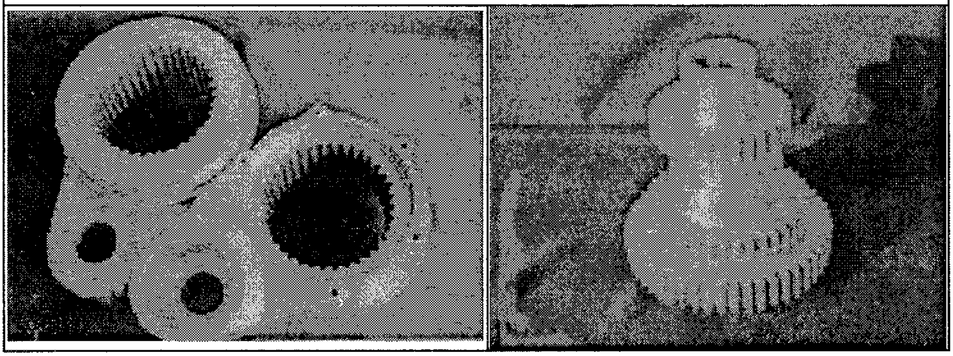
ASTM: American Society for Testing and Materials
AFNOR: Association Franchaise De Normalization
BSI: British Standards Institute
DIN: Deutsches Institut Fur Normung
USCAP: Unified System of Corrosion and Aging Protection
ISO: International Standards Organization

Other associated organizations

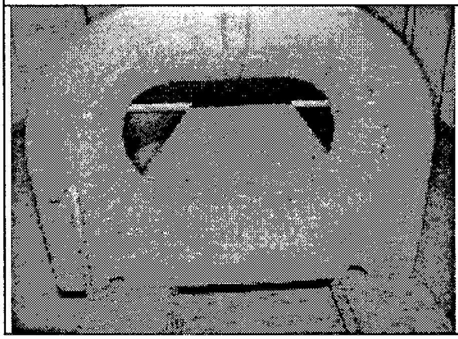
IEC: International Electro technical Commission
NACE: National Association of Corrosion Engineers

APPENDIX 2: Examples of surface engineered specimens

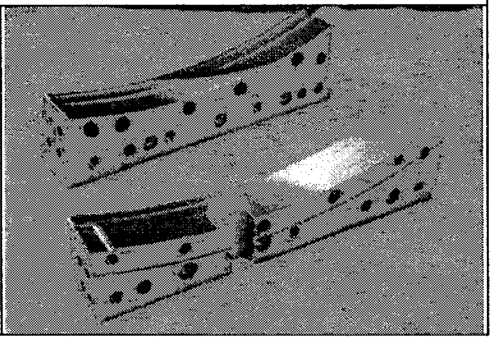
Typical plasma nitrided automotive parts



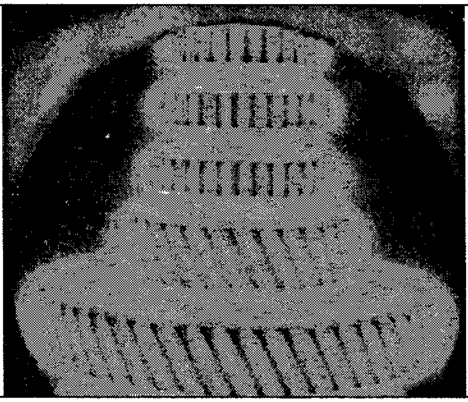
A Induction hardened marine component

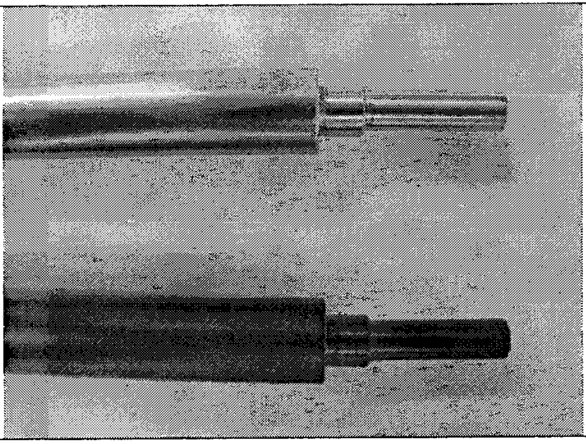
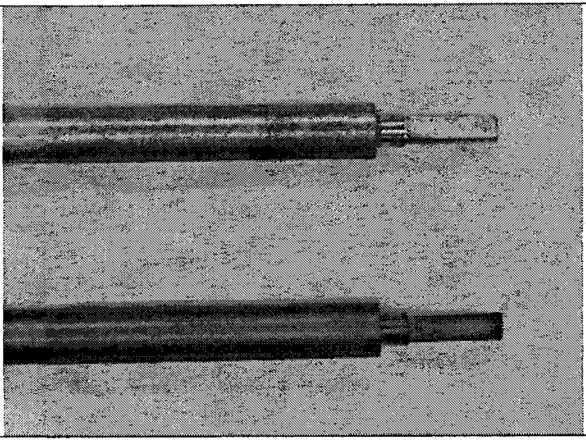
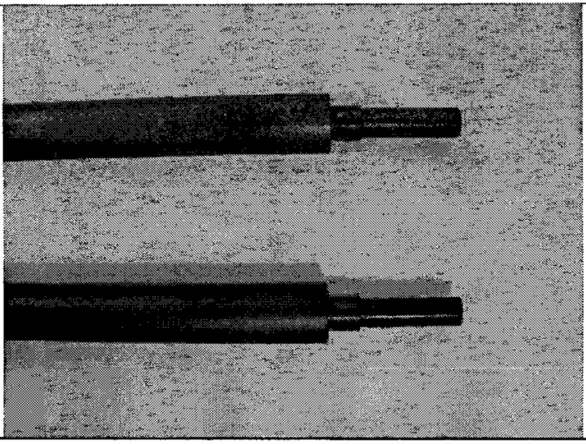


A typical plasma nitrided aerospace component



Plasma nitriding process of a helical gear component



A comparative evaluation of unnitrided and plasma nitrided medium carbon steel after day 0 of atmospheric exposition

A comparative evaluation of unnitrided and plasma nitrided medium carbon steel after day 1 of atmospheric exposition

A comparative evaluation of unnitrided and plasma nitrided medium carbon steel after day 13 of atmospheric exposition


INDEX

Materials Degradation and its Control By Surface Engineering

The second edition of Materials Degradation and Its Control by Surface Engineering continues the theme of the first edition, where discussions on corrosion, wear, fatigue and thermal damage are balanced by similarly detailed discussions on their control methods, e.g. painting and metallic coatings. The book is written for the non-specialist, with an emphasis on introducing technical concepts graphically rather than through algebraic equations. In the second edition, the graphic content is enhanced by an additional series of colour and monochrome photographs that illustrate key aspects of the controlling physical phenomena. Existing topics such as liquid metal corrosion have been extended and new topics such as corrosion inhibitors added.

"Guidelines for applications of surface engineering techniques to individual degradation mechanisms are covered. This does a concise job of suggesting basic selection criteria to be followed for specific degradation mechanisms ... The authors present a good overview of the interaction of surface engineering treatments for control of material wastage from various causes."

Corrosion

P271 hc

ISBN 1-86094-334-9

9 781860 943348

Imperial College Press

www.icpress.co.uk