Handbook of Aluminum

Volume 2

Alloy Production and Materials Manufacturing

edited by

George E. Totten
D. Scott MacKenzie

Handbook
of Aluminum

Volume 2
Alloy Production and Materials
Manufacturing

edited by

George E. Totten
G. E. Totten & Associates, Inc.
Seattle, Washington, U.S.A

D. Scott MacKenzie
Houghton International Incorporated
Valley Forge, Pennsylvania, U.S.A.

Preface

This second volume of *Handbook of Aluminum* addresses the physical metallurgy and processing technologies of aluminum and its alloys in a thorough and deliberate manner. Volume 1 covers a wide range of topics including pure aluminum and its properties, an extensive discussion of the physical metallurgy of aluminum and its alloys, and processes such as rolling, forging, casting, welding, quenching, super-plastic forming, and others.

There are 18 chapters in volume 2 and some of the topics discussed include:

- Production of aluminum alloys including extractive metallurgy, smelting, master alloy creation, and recycling
- Analytical methods used to characterize aluminum alloys
- Work hardening, recovery, recrystallization, and grain growth
- Modeling of microstructure evolution
- Microstructure-texture-property relationships
- Property prediction using quench factor analysis
- Mechanical properties and alloy chemistry
- Aluminum alloy corrosion
- Surface chemistry of adhesion processes
- Surface modification including nitriding
- Friction stir welding
- Aluminum intermetallics and metal matrix composites
- Environmental effects and toxicity of aluminum production processes

The information in these chapters, in addition to that in Volume 1, provides the reader with an extensive and rigorous reference to nearly all aspects of the production, physical metallurgy, and processing technologies encountered in the aluminum alloy industry.

We are indebted for the tremendous effort and patience shown by all of our contributors. We are especially indebted to Alice Totten and Patricia MacKenzie for their unending patience and assistance throughout the preparation of this text. We also acknowledge Houghton International for their support, without which this book would not have been possible.

George E. Totten
D. Scott MacKenzie

Contents

Contributors

Gilbert F. Bourcier, B.S. Old Dominion Engineering Services Company, Midlothian, Virginia, U.S.A.

T. David Burleigh, Ph.D. New Mexico Tech, Socorro, New Mexico, U.S.A.

Chris H. J. Davies, Ph.D. Monash University, Victoria, Australia

Alexis Deschamps, M.Eng., Ph.D. Institut National Polytechnique de Grenoble, Domaine Universitaire, Saint Martin d'Hères, France

Denise Crocce Romano Espinosa, Ph.D. University of São Paulo, São Paulo, Brazil

Georg Frommeyer, Dr.-Ing. Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf, Germany

Kiyoshi Funatani, Ph.D. IMST Institute, Nagoya, Aichi, Japan

Michael I. Gasik, D.Sci.Tech. National Metallurgical Academy of Ukraine, Dnipropetrovsk, Ukraine

Michael M. Gasik, D.Sc., D.Tech.Sci. Helsinki University of Technology, Espoo, Finland

Fathi Habashi, Dr.Tech. Laval University, Quebec City, Canada

Margaret M. Hyland, Ph.D.   University of Auckland, Auckland, New Zealand

Sven Knippscheer, Dipl.-Ing. Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf, Germany

Vladislav I. Mazur, D.Sci.Tech. National Metallurgical Academy of Ukraine, Dnipropetrovsk, Ukraine

D. Scott MacKenzie, Ph.D.   Houghton International Incorporated, Valley Forge, Pennsylvania, U.S.A.

Angelo Fernando Padilha, Dr.-Ing.   University of São Paulo, São Paulo, Brazil

Ronald Lesley Plaut, Ph.D.   University of São Paulo, São Paulo, Brazil

Dierk Raabe, Dr.-Ing. Max Planck Institut for Iron Research, Düsseldorf, Germany

Bala Radhakrishnan, Ph.D. Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.

Bert Reinhold, Dipl.-Phys.   ALD Vacuum Technologies AG, Hanau, Germany

Anthony P. Reynolds, Ph.D.   University of South Carolina, Columbia, South Carolina, U.S.A.

Robert E. Sanders, Jr., Ph.D.   Aluminum Corporation of America, Alcoa Center, Pennsylvania, U.S.A.

Gorti Sarma, Ph.D. Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.

Heinz-Joachim Spies, Dr.-Ing.habil.   Freiberg University of Mining and Technology, Freiberg, Germany

James T. Staley, Ph.D.   Consultant, Durham, North Carolina, U.S.A.

Jorge Alberto Soares Tenório, Ph.D.   University of São Paulo, São Paulo, Brazil

Yoshiki Tsunekawa, Ph.D.   Toyota Technological Institute, Nagoya, Aichi, Japan

Masayuki Yoshida Nihon Parkerizing Company Ltd., Hiratsuka, Kanagawa, Japan

1

Extractive Metallurgy of Aluminum

FATHI HABASHI

Laval University, Quebec City, Canada

1 HISTORY

Since Humphry Davy announced in 1808 his belief that the plentiful compound alumina was the earth (oxide) of an undiscovered metal, scientists had been making efforts to obtain this new metal. Davy never made any aluminum himself; but in 1825, the Danish scientist Hans Christian Oersted (1777–1851) published his successful experiment in producing a tiny sample of the metal in the laboratory by reducing aluminum chloride with potassium amalgam. Potassium was isolated a few years earlier by Davy.

Two years later, Friedrich Wöhler (1800–1882) in Germany produced tiny globules of aluminum by the same method, and was able to demonstrate the metal's lightweight and malleability. Henri Sainte–Claire Deville (Fig. 1) in France in 1854 showed that cheaper sodium could also be used, and the first commercial plant producing small quantities of aluminum was begun in 1855. Since potassium and sodium were produced electrolytically, the process was expensive.

In 1886, following the development of large-scale equipment for generating electrical power, Paul Héroult (Fig. 2a) in France, and Charles Hall (Fig. 2b) in the United States, independently developed a process for the direct electrolytic decomposition of Al_2O_3. They discovered that when an electric current is passed through molten cryolite containing dissolved Al_2O_3 at 980–1000°C, molten aluminum is deposited at the cathode and carbon dioxide is liberated at the carbon anode. This discovery, coupled with the process developed by Karl Josef Bayer (Fig. 3) in 1888 for the production of alumina, resulted in the modern process for the production of aluminum.

Figure 1 Henri Sainte–Claire Deville (1818–1881), the first to produce aluminum on commercial scale by reduction of $AlCl_3$ with sodium.

2   GENERAL REMARKS

Aluminum comprises 8% of the earth's crust and is, therefore, the most abundant structural metal. Its production since 1965 has surpassed that of copper and now comes next to iron (Fig. 4). Its unit price started very high and today is comparable to copper (Fig. 5). It is competing with copper in the electric industry and as a material of construction. Although the electrical conductivity of aluminum is slightly lower than that of copper, it is still economical to use in preference to copper in power cables because of its lighter weight. As a material of construction, aluminum can be anodized to get a protective oxide film, which can be dyed to give a colorful appearance. For the production of the metal, the following points should be taken into consideration:

- The electrowinning of aluminum from an aqueous solution is not possible because of the strongly negative deposition potential of this metal and the rapid hydrolysis of the aluminum ion.
- Oxides in general have high melting points. Therefore, for the electro-winning of aluminum from its oxide, a suitable low melting point electrolyte must be found in which the oxide is appreciably soluble.
- In the manufacture of aluminum, there are two main stages. The first embraces the production of pure Al_2O_3 from bauxite, and the second is the reduction of this Al_2O_3 to the metal in a bath of fused cryolite (Fig. 6).

Figure 2 (Top) Paul Héroult (1863–1914) and (bottom) Charles Martin Hall (1863–1914) invented simultaneously and independently the electrolytic process for reduction of Al_2O_3.

- In the electrowinning of aluminum from oxide melts, the carbon anodes are quantitatively consumed.
- The production of alumina from bauxite is the largest pressure leaching operation in the world.
- Similarly, the production of aluminum by the molten salt electrolysis of alumina in cryolite is the most important industrial application of molten salt electrowinning and is the largest electrolytic industry in the world.

Figure 3 Karl Josef Bayer (1871–1908) invented the process for producing Al_2O_3 from bauxite.

- The electrolytic step in aluminum production is the most energy intensive operation (Table 1).
- Material handling is a major cost factor in aluminum production. For each ton of aluminum produced, more than 3 tons of materials are transported to and from potlines.
- The electrolytic process produces aluminum with purity as great as 99.5–99.8%. Higher purity (99.99%) is obtained by an electrolytic refining process.

3 RAW MATERIALS

The production of aluminum by fused salt electrolysis takes place in an electrolytic cell which consists of a carbon anode, a molten cryolite–alumina electrolyte, a pool of liquid aluminum, and a carbon-lined container to hold the metal and electrolyte. Fresh alumina is added from time to time and the metal formed is siphoned out periodically. The reactions taking place can be simplified as follows:

Cathodic reaction: $Al^{3+} + 3e^- \rightarrow Al$

Anodic reaction: $O^{2-} + C \rightarrow CO + 2e^-$

Overall reaction: $Al_2O_3 + 3C \rightarrow 2Al + 3CO$
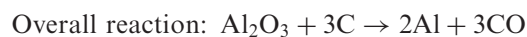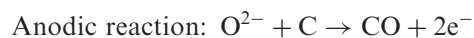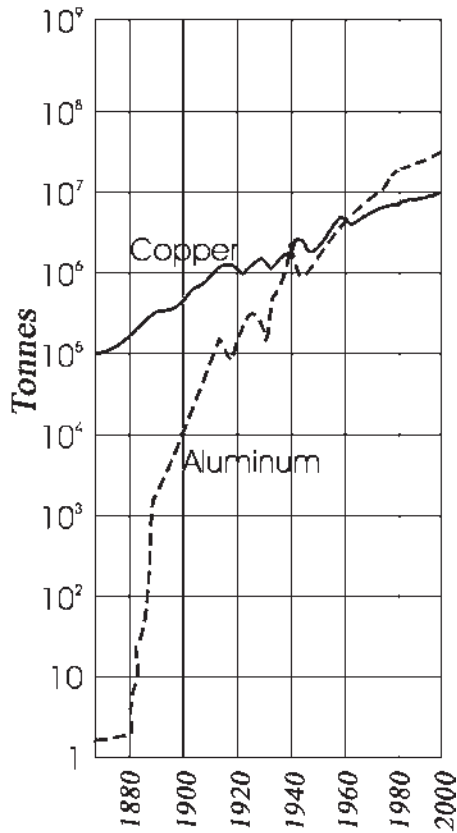
Figure 4 World production of aluminum vs. that of copper.

3.1 Aluminum Oxide

Almost all the alumina feed is made from bauxite which contains about 50% Al_2O_3 in the form of hydroxides of aluminum. The ore is digested with caustic soda solution under pressure to dissolve the alumina as aluminate leaving behind a red mud containing iron oxides and other major impurities. At the same time, sodium aluminum silicate should also be brought down onto the mud to minimize the silica content of the liquor; otherwise, the final aluminum hydroxide is contaminated. The aluminate solution is then seeded with crystals of hydroxide to deposit this material in a form that is easily washed and handled. The precipitation is performed in cycles leaving behind a large quantity of crystals to seed the next cycle. Finally, the hydroxide is calcined to alumina at 1200°C to give a product >99% pure. To produce 1 ton of aluminum, about 2 tons of Al_2O_3 are required.

Russia is the only country producing Al_2O_3 from sources other than bauxite. Lacking large, commercially viable bauxite reserves, the former USSR has pioneered the production of alumina from nepheline, alunites, kaolin, shale, ash, and other alkali-free aluminosilicates, which are plentiful in the country. Production techniques developed in the former USSR are used at the Volkhovsky and Pikalevsky
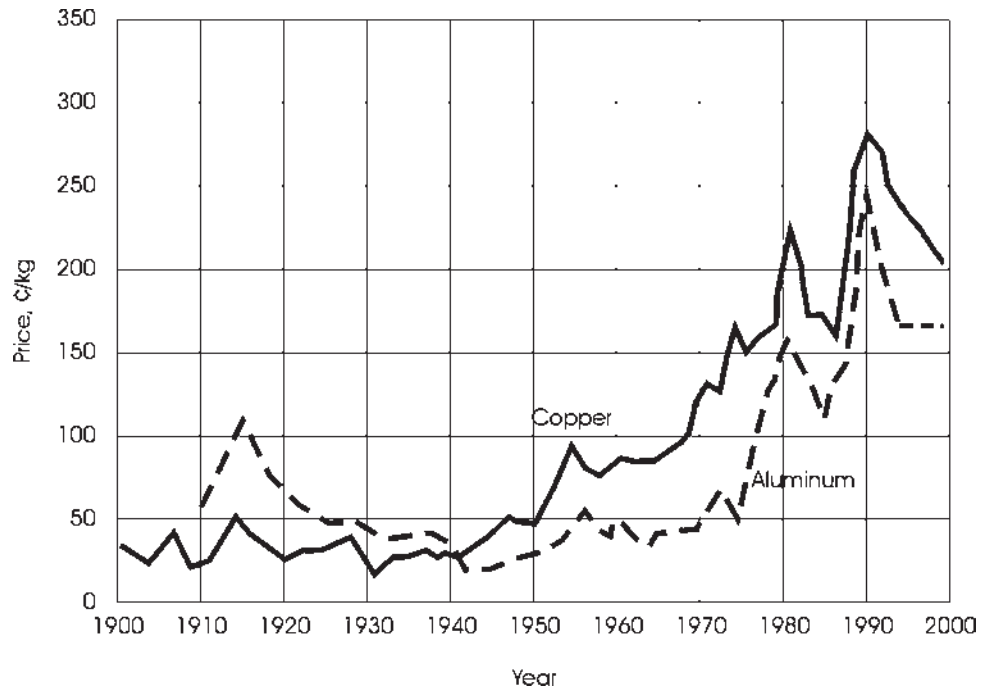
Figure 5 Price of aluminum as compared to that of copper.
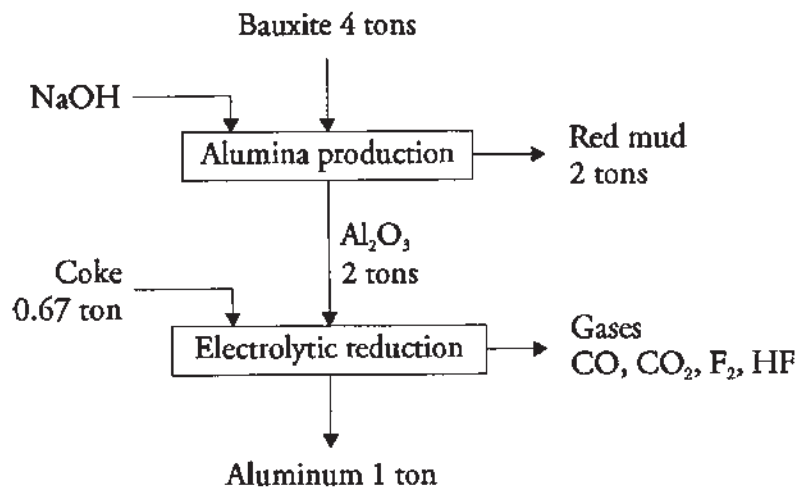
Figure 6 Production of aluminum.

Table 1 Aluminum Production Costs in an Integrated Plant

Sector	Percentage
Mining and milling	4.6
Al_2O_3 production	22.3
Metal production	73.1
Total	100.0

aluminum plants near Saint Petersburg and at the Achinsk alumina complex in Siberia. In producing alumina from nepheline, the only other raw material required is limestone. Four tons of nepheline are heated in rotary kilns with 12 tons of limestone to produce 1 ton of Al_2O_3, 0.7 ton of Na_2CO_3, 0.3 ton of K_2CO_3, and variable amounts of portland cement.

3.2 Carbon

Carbon is used for fabricating the anodes as well as the cathodes. It is a major item in making aluminum.

- Anodes are consumed in amounts of 0.4–0.5 kg of carbon/kg of aluminum produced; the theoretical consumption is 0.33 kg/kg Al. Since the ash from the carbon will contaminate either the aluminum produced or the electrolyte, high-purity carbon is desirable. The coke residue from petroleum refining has been the major source of carbon for anodes. This coke, produced at about 500°C, requires calcining at 1200°C to remove volatile constituents and increase its density.
- Cathodes are consumed in amounts of 0.02–0.04 kg of carbon/kg of aluminum produced due to disintegration during electrolysis. The cathodes are usually made of anthracite. Purity requirement is not as strict as in the anodes since contamination of the metal from the cathodes is negligible.

3.3 Cryolite

Cryolite has been found only in Greenland. It was mined there in the early 20th century and is now exhausted. Synthetic cryolite is produced by reacting hydrofluoric acid with alkaline sodium aluminate solution:

$$6HF + 2NaOH + NaAlO_2 \rightarrow Na_3AlF_6 + 4H_2O$$

Some cryolite is produced directly in reduction cells by reaction of the soda impurity in the alumina with added aluminum fluoride:

$$3Na_2O + 4AlF_3 \rightarrow 2Na_3AlF_6 + Al_2O_3$$
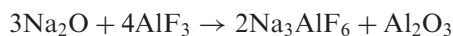
In modern plants with equipment for fume treatment, cryolite is a by-product rather than a raw material.

3.4 Fluorides

Although cryolite is not consumed during electrolysis, there are some losses due to volatilization as HF and different fluorine compounds estimated at 3–5% of the

metal produced. Some of these are captured and returned to the cell, others are lost. Many aluminum reduction plants have facilities to produce the hydrofluoric acid necessary for preparing cryolite and other additives as AlF_3, LiF, and MgF_2. Hydrofluoric acid is prepared by reacting fluorspar with concentrated H_2SO_4:

$$CaF_2 + H_2SO_4 \rightarrow 2HF + CaSO_4$$
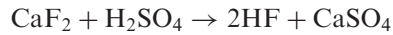
4   ALUMINUM OXIDE FROM BAUXITE

Bauxite,* named after the village *Les Beaux* near Marseille in southern France where it was first discovered, is not a mineral, but designates various kinds of aluminum ores consisting mainly of aluminum hydroxide. Three aluminum hydroxide minerals occur in bauxite: gibbsite, böhmite, and diaspore. They differ considerably in their physical properties, as shown in Table 2. A bauxite deposit consists mainly of either one of these types, although cases are known when mixed hydroxides are present in one ore. Bauxites vary in color from cream to dark brown when the iron content is high. Table 3 shows the composition of a typical bauxite. The main occurrences of bauxites are in Jamaica, Suriname, Ghana, Sierra Leone, Australia, Russia, and Hungary.

The treatment of bauxite to produce pure $Al(OH)_3$ from which pure alumina is obtained by the Bayer process using sodium hydroxide solution is the oldest and the largest pressure leaching operation in terms of the tonnage of raw material treated (Fig. 7). Aluminum minerals in bauxite are soluble in dilute H_2SO_4 but this acid is not used in large scale for the following reasons:

- Iron minerals and to some extent titanium minerals are also soluble; this will lead to an excessive reagent consumption and solution purification problem later.
- $Al(OH)_3$ precipitated from acid solutions is gelatinous and difficult to filter and wash.

Table 2   Aluminum Minerals in Bauxite

	Gibbsite (hydrargillite)	Böhmite	Diaspore
Formula	γ-$Al(OH)_3$	γ-AlOOH	α-AlOOH
Al_2O_3:H_2O ratio	1:3	1:1	1:1
Crystal system	Monoclinic	Orthorhombic	Orthorhombic
Hardness (Moh)	2.5–3.5	3.5–4	6.5–7
Specific gravity	2.42	3.01	3.44
Refractive index	1.568	1.649	1.702
Temperature of rapid dehydration	150°C	350°C	450°C
Product of dehydration	χ-Al_2O_3	γ-Al_2O_3	α-Al_2O_3
Solubility in 100 g/l Na_2O Solution at 125°C; g/l Al_2O_3	128	54	Insoluble

---

*Bauxite was originally spelled *beauxite*.

Table 3 Composition of Typical Bauxites

	Percentage
Al_2O_3	40–60
SiO_2	1–6
Fe_2O_3	2–25
TiO_2	1–5
$CaO+MgO$	0.2–0.6
Loss on ignition	10–30
Ga_2O_3	0.01
K_2O	0.01
P_2O_5	0.02–0.4
V_2O_5	0.01–0.1
$Ln_2O_3{}^a$	0.01
F	0.01–0.05

[a]$Ln=$ lanthanide.

Acid leaching is used only on a small scale to produce aluminum sulfate which is needed for water treatment.

The use of sodium hydroxide to leach bauxite was invented in 1892 by Karl Josef Bayer as a process for obtaining pure aluminum hydroxide which can be calcined to pure Al_2O_3 suitable for processing to metal. The process involves three steps: selective dissolution of aluminum hydroxide from bauxite, precipitation of pure aluminum hydroxide from the solution, and calcination of the hydroxide to Al_2O_3. Historically, the precipitation step was invented before the leaching step.

4.1 Leaching of Bauxite

About 90 million tons of bauxite are treated annually by this process. About 2 tons of bauxite yield 1 ton of Al_2O_3 from which 0.5 ton aluminum is produced. Also, 2 tons of bauxite produce 1 ton of waste minerals called *red mud*. Crushed bauxite is usually washed to remove fine particles of clay, dried in a rotary kiln, then ground to 60–100 mesh; the drying process is essential to facilitate grinding. Drying temperature should be less than the temperature of dehydration of aluminum hydroxides; otherwise, the solubility will be impaired. Figure 8 shows a flowsheet of the process and Fig. 9 shows an operating plant. The reactions in leaching are the following:

$$Al(OH)_3 + OH^- \rightarrow [AlO(OH)_2]^- + H_2O$$

$$AlOOH + OH^- \rightarrow [AlO(OH)_2]^-$$
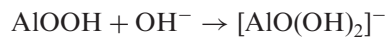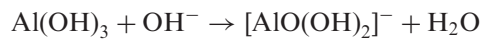
Leaching is usually carried out in mild-steel autoclaves, with direct steam injection for heating and agitation (Fig. 10). Operating conditions depend on the type of minerals in the bauxite. Bauxites containing gibbsite are leached at a lower temperature, with lower NaOH concentration, and for a shorter time than those containing böhmite and diaspore as shown in Table 4. The more concentrated the NaOH, the faster the rate; however, highly concentrated solutions will require excessive dilution in the later stage of precipitation, which presents difficulties in
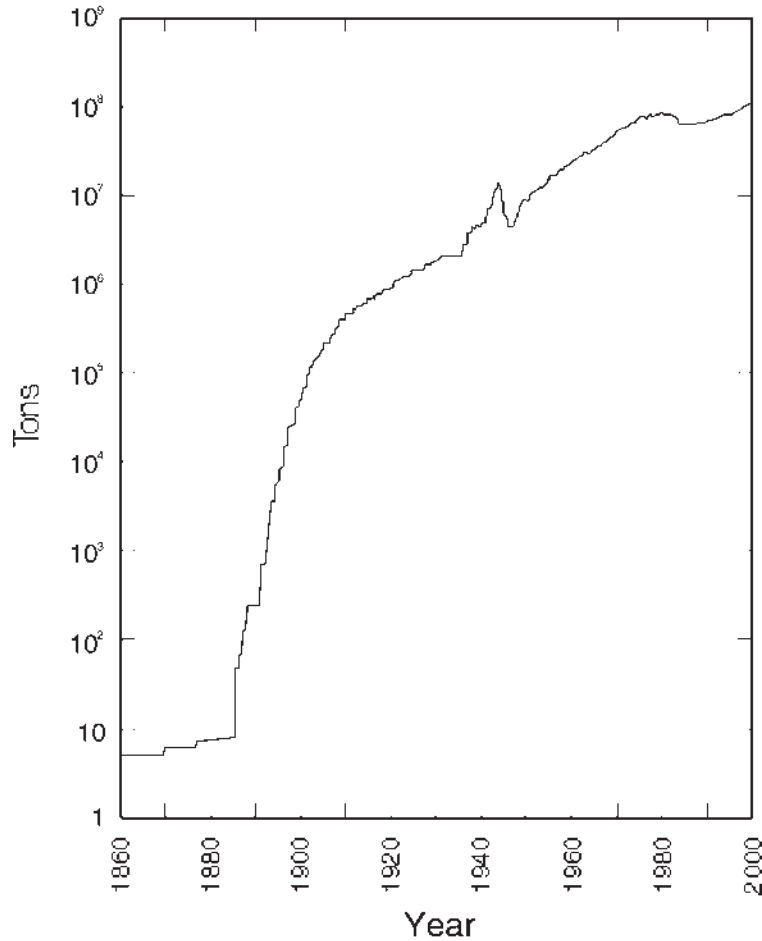
Figure 7 Annual processing of bauxite worldwide.

handling and filtration. Therefore, there must be an optimum concentration which compromises between digestion time and subsequent operations.

Leaching time could be shortened to 3–4 min if the process is conducted in tube autoclaves at 330°C and 25,000 kPa; also, the settling properties of the mud are improved. In tube autoclaves (Fig. 11), the slurry is pumped into an externally heated thick-walled tube about 30 cm in diameter and 30–50 m long. The major part of the heat is supplied by the slurry leaving the tube. Only at the extreme end of the tube, steam from an outside source is used for heating.

Sometimes, NaOH is formed in situ in the autoclaves by adding sodium carbonate and calcium hydroxide:

$$Na_2CO_3 + Ca(OH)_2 \rightarrow 2NaOH + CaCO_3$$
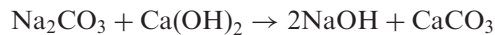
To a leach solution containing 200–250 g/l Na_2CO_3, enough lime is added such that the solution contains about 140 g/l NaOH. Leaching is carried out at 140°C for about 1 hr. This method applies only to gibbsite, because in the case of böhmite or
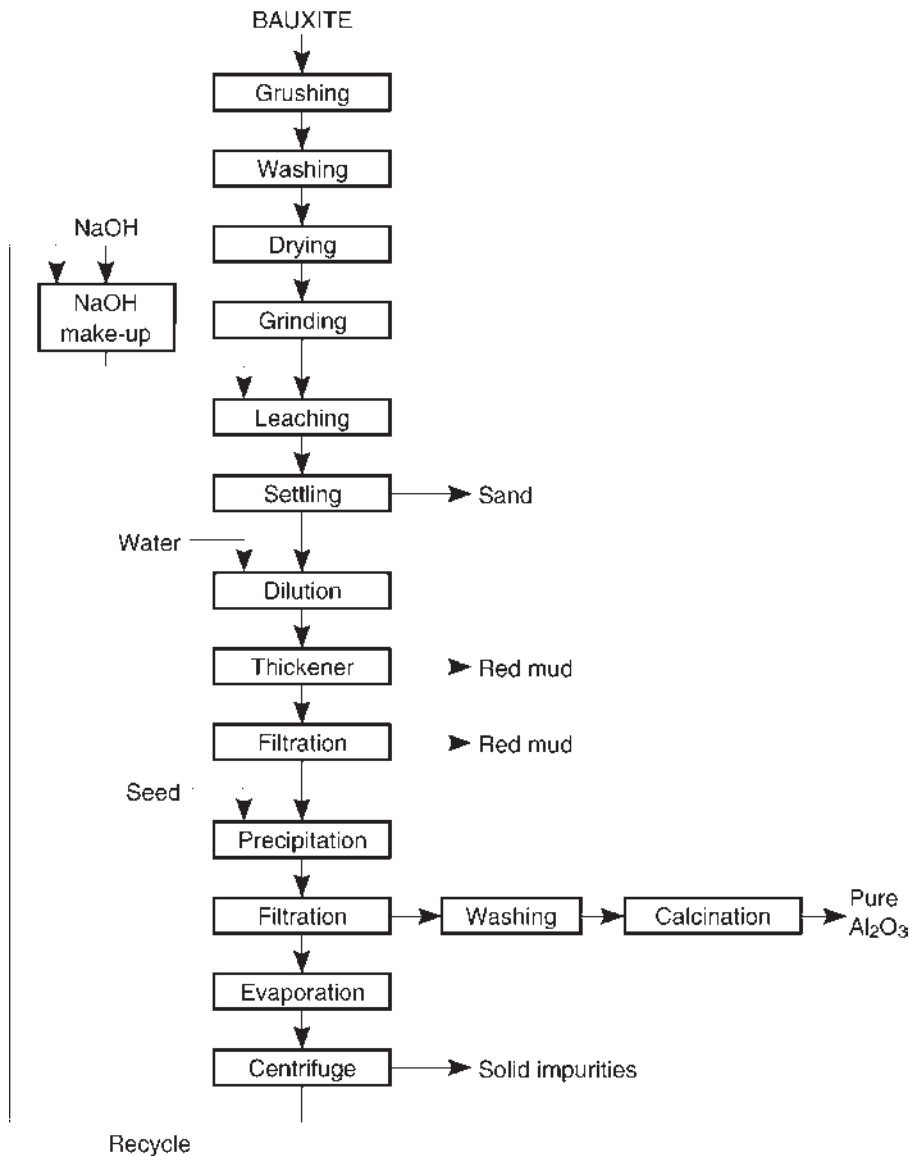
Figure 8 Flowsheet of the Bayer process.

diaspore, a sodium hydroxide concentration above that which can be obtained directly from Na_2CO_3 and lime is required. So, it is necessary to prepare NaOH separately and concentrate it by evaporation for use in leaching.

4.1.1 Behavior of Impurities During Leaching

Depending on their chemical properties, impurities in the ore may be found either in solution, or in the insoluble residue. Impurities that go into solution are either redeposited during a later operation, or remain in the mother liquor during
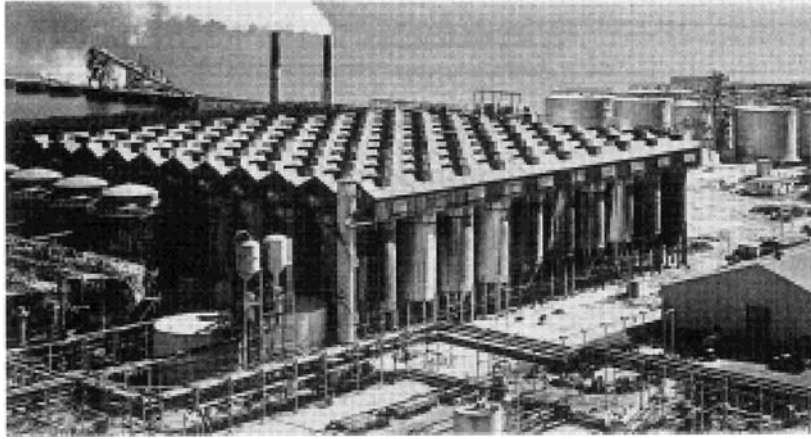
Figure 9   Bauxite treatment plant by the Bayer process in Kwinana near Perth in Australia produces 1,250,000 tons/year alumina (ALCOA).
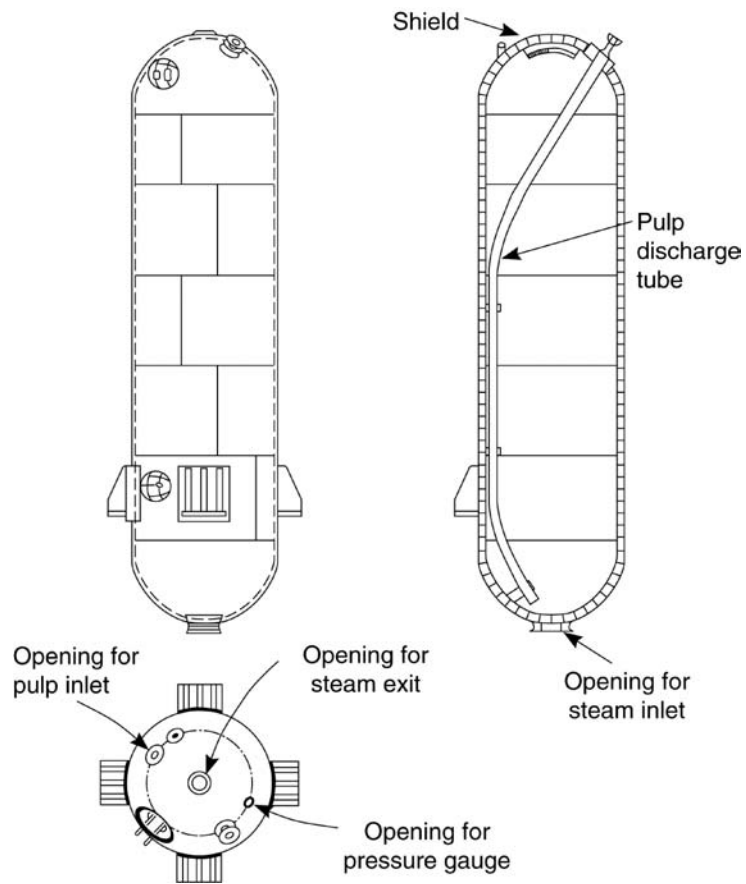
Figure 10   A typical autoclave for leaching bauxite.

Table 4 Typical Leaching Conditions of Bauxite with NaOH

Ore	Temperature (°C)	Pressure (kPa)	NaOH (g/l)	Time (hr)
Gibbsite	140	400	140	1
Böhmite	180	800	350–600	2–4
Diaspore	180	800	350–600	2–4

crystallization of aluminum hydroxide. However, due to recycling, they will accumulate in solution and would contaminate the product. Certain tolerable levels are, therefore, maintained by bleeding the solution at regular intervals. Table 5 shows the distribution of impurities during the processing of bauxite, and Table 6 shows the composition of the leach solution. Essentially, the Bayer process eliminates the three major impurities in bauxite Fe_2O_3, SiO_2, and TiO_2. Calcium and magnesium are mainly present as dolomite and are not dissolved. Apatite is partially soluble and is usually eliminated during leaching by adding $Ca(OH)_2$.

Organic Matter

Most bauxites contain about 0.1% organic matter. During digestion, some of this material is dissolved, causing the liquors to darken, while the remaining part is degraded and oxidized to oxalates. Some of the organic matter is, therefore, responsible for NaOH losses. Their presence may also cause liquors to froth, or may interfere with the subsequent process of hydroxide precipitation, or color the hydroxide. It may be largely eliminated during digestion by adding an oxidizing agent, e.g., MnO_2. In some cases, sodium oxalate is crystallized and removed. It should be noted, however, that the organic matter in the liquor may come from the flocculants added to assist the mud to settle.
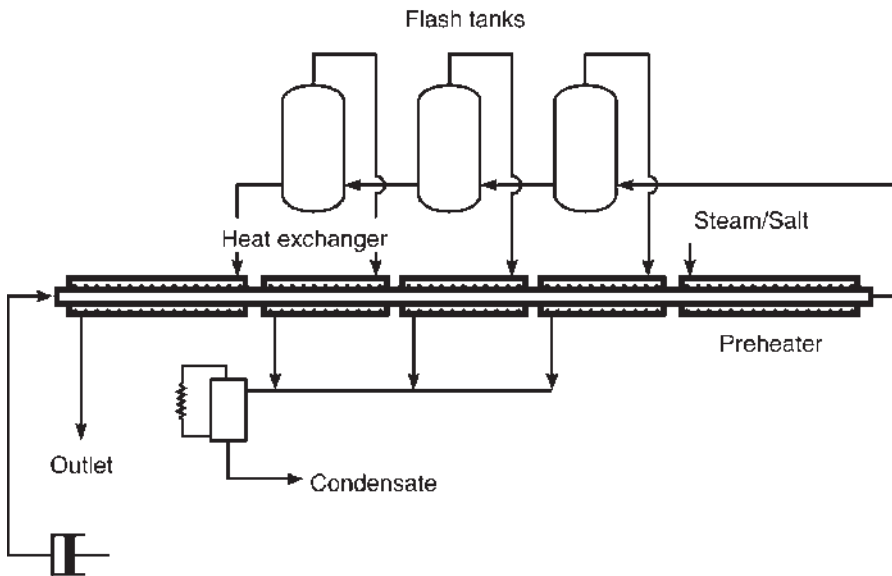
Figure 11 Tube autoclaves used mainly in Germany for leaching bauxite.

Table 5 Distribution of Impurities During the Processing of Bauxite

	Bauxite (%)	Red Mud (dried at 105°C) (%)	Calcined Al_2O_3 (%)
Al_2O_3	57.8	14.0	99.55
SiO_2	3.5	7.6	0.05
Fe_2O_3	24.3	57.6	0.04
TiO_2	2.5	5.7	—
Na_2O	—	7.4	0.18
Loss on ignition	12.5	7.7	0.18

Table 6 Composition of Typical Aluminate Solution

g/100 g free Na_2O

Al_2O_3	32.80[a]	Cl^-	1.00
CaO	0.12	CO_2	3.74
Fe_2O_3	0.05	F^-	0.03
Ga_2O_3	0.22	MgO	0.17
Na_2O	100.00	P_2O_5	0.90
SiO_2	0.60	SO_3	0.07
TiO_2	Trace	V_2O_5	0.45

[a]In the precipitation step, about 2/3 of the aluminum in solution is precipitated.

Iron

Iron occurs in bauxite mainly as hematite, Fe_2O_3, and is not attacked by the caustic leaching. Thus, the residue remaining after leaching has a high percentage of iron oxides, and, therefore, has a red color. That is why this residue is usually referred to as red mud. However, some ores contain ferrous iron in the form of siderite, $FeCO_3$. This is attacked by NaOH, forming colloidal ferrous hydroxide:

$$FeCO_3 + 2OH^- \rightarrow Fe(OH)_2 + CO_3^{2-}$$
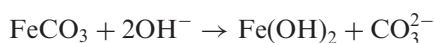
which is difficult to settle. It would, therefore, be advantageous to oxidize ferrous minerals to ferric during leaching.

Silicon

Silicon occurs as quartz, SiO_2, or as clays, e.g., kaolinite, $Al_2(OH)_4(Si_2O_5)$. Quartz is insoluble in NaOH under the conditions of leaching but the silicates are soluble. During digestion, silica that goes into solution combines with alumina and sodium hydroxide forming insoluble hydrated aluminosilicates such as $2Na_2O \cdot 2Al_2O_3 \cdot 3SiO_2 \cdot 2H_2O$, which are carried away in the red mud, thus causing losses. About 1 kg of NaOH is lost for each kilogram of soluble silica in bauxite. Although most of the soluble silica in bauxite is precipitated during digestion by forming sodium aluminosilicate, small amounts may still be found in solution, especially when concentrated NaOH solution is used. To precipitate the silica completely, addition of CaO is recommended, since insoluble calcium silicate can be formed. Lime addition during digestion has a further advantage: any Na_2CO_3 present in the solution due to

absorption of CO_2 from the atmosphere, and which has no dissolving action on bauxite, will be converted to NaOH.

Gallium

Gallium occurs in most bauxites, and is highest in French bauxites (0.0–0.05% Ga_2O_3). It dissolves completely during extraction. Recycling of NaOH in the process results in gallium enrichment up to 0.5 g/l. Such liquors are, therefore, an important source of gallium, from which it can be recovered, e.g., by solvent extraction or by electrolysis using a mercury cathode. If, however, gallium is not recovered and is left to build up in the leach liquor, it will reach a certain concentration beyond which it will be partially deposited, together with aluminum hydroxide during precipitation, thus causing contamination.

Vanadium

Vanadium in bauxite is partly soluble during digestion. In some ores, it is precipitated during evaporating the leach solution as complex salts such as $2Na_3VO_4 \cdot NaF \cdot 19H_2O$. This is especially the case for ores containing fluorine since fluorides are dissolved during leaching. Sometimes, these precipitate to form a hard scale in the evaporators which interferes with heat transfer. In other ores, vanadium builds up in the recycled NaOH to a concentration of about 0.5 g/l V_2O_5 and is recovered.

4.1.2 Mechanism of Dissolution

Because of their slightly ionic character, aluminum hydroxides in bauxite form minor amounts of ions when added to water. For example:

$$Al(OH)_{3(s)} \rightarrow Al^{3+}_{(aq)} + 3OH^-_{(aq)}$$
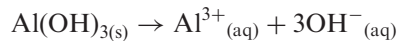
$$K_1 = [Al^{3+}][OH^-]^3 = 1.9 \times 10^{-33}$$

The value of the equilibrium constant is very small. Equilibrium is disrupted when the concentration of any of the ions Al^{3+} or OH^- is decreased; thus, more solid will go into solution to keep the value of K constant. Decreasing any of these ions may be the result of a neutralization reaction or a complex formation. In the presence of an acid, the OH^- ions are neutralized:

$$H^+ + OH^- \rightarrow H_2O \quad K_2 = 1 \times 10^{14}$$
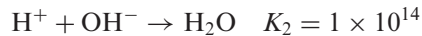
This shifts the equilibrium to the right and more hydroxide goes into solution:

$$Al(OH)_3 + 3H^+ \rightarrow Al^{3+} + 3H_2O$$
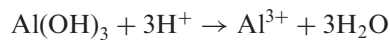
$$K = K_1 K_2^3 = 1.9 \times 10^9$$

Aluminum hydroxide is an amphoteric hydroxide, i.e., it may behave as an acid:

$$Al(OH)_{3(s)} \rightarrow AlO(OH)^-_{2(aq)} + H^+$$

$$K_1' = [AlO(OH)_2^-][H^+] = 4 \times 10^{-13}$$

In the presence of an alkali, the H^+ ions are neutralized:

$$H^+ + OH^- \rightarrow H_2O \quad K_2 = 1 \times 10^{14}$$
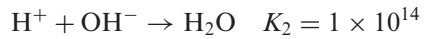
The equilibrium is shifted to the right and more hydroxide dissolves:

$$Al(OH)_{3(s)} + OH^- \rightarrow AlO(OH)^-_{2(aq)} + H_2O$$
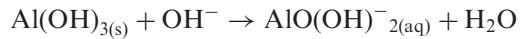
$$K' = K'_1 K_2 = 40$$

Thus, aluminum hydroxide is soluble in both acid and alkali.

4.1.3 Engineering Aspects

High-pressure membrane piston pumps (Fig. 12) are used for introducing pulps into autoclaves. Flash evaporators are usually installed after an autoclave. These are large, vertical tanks in which the hot slurry is introduced through a tube directed towards the bottom of the tank where protective baffles are installed to minimize the erosion of the tank due to impact (Fig. 13). This equipment serves three purposes: decreasing the pressure and temperature of the slurry, recovery of heat in the form of low-pressure steam, and concentration of the solution as a result of the flash evaporation of water. Slurries, to be introduced in an autoclave, are usually preheated by the steam generated in the flash evaporator.

Slurries obtained after leaching are then filtered to recover the leach solution, then washed to remove the entrained solution. The red mud generated is usually disposed of in ponds, and in some cases, transported by a pipeline under the sea.
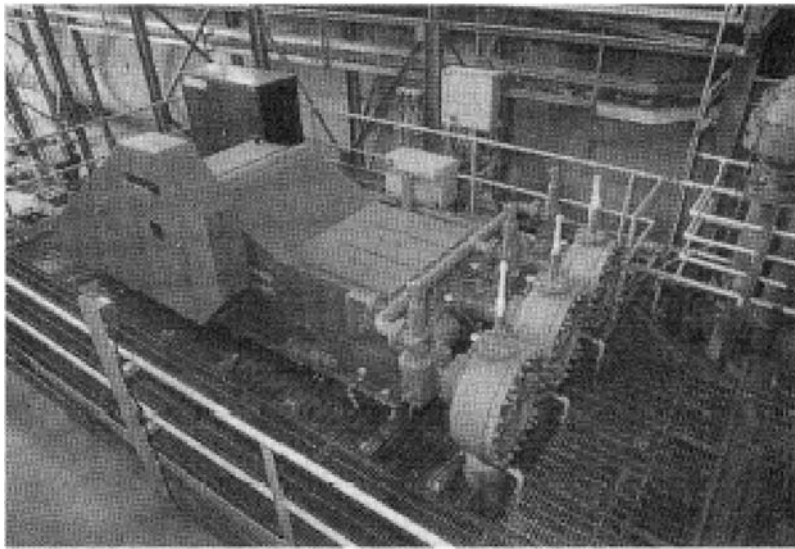
Figure 12 High-pressure pump.
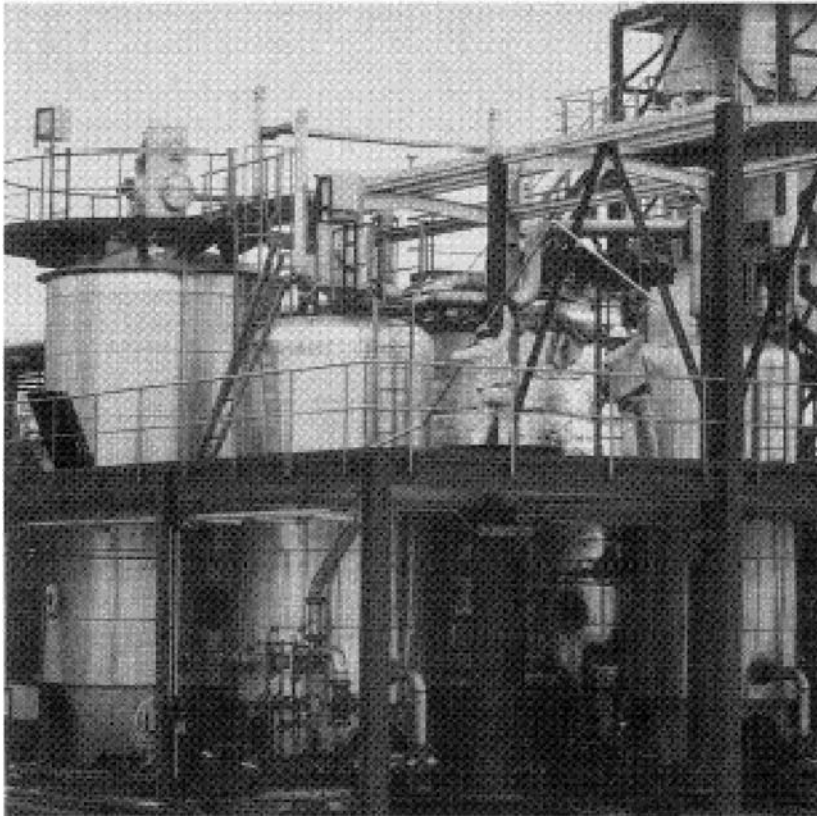
Figure 13 Flash tank.

4.2 Precipitation of Aluminum Hydroxide

There are a large number of aluminum hydroxides known:

— Gibbsite, γ-Al(OH)$_3$, occurs in huge quantities in bauxite and is produced in large scale by the Bayer process by cooling slowly a concentrated solution of sodium aluminate in the presence of crystal seeds. A crystalline product is obtained in the form of hexagons. The structure as revealed by x-rays emphasizes this hexagonal form. It is a layered structure composed of two planes of close-packed OH$^-$ ions with Al^{3+} ions sandwiched between them. The Al^{3+} ions are distributed in hexagonal rings. In the interior, each Al^{3+} shares six OH$^-$ with three other Al^{3+}, and OH$^-$ is bridged between two Al^{3+}. Three of the six OH$^-$ ions surrounding the Al^{3+} ion are from the upper layer, and the other three are from the lower layer (Fig. 14). Crystalline aluminum hydroxide is, thus, a giant molecule composed of layers of Al(OH)$_6$ octahedra packed in a cubic form.

— Bayerite, α-Al(OH)$_3$, is rarely found in nature, but is prepared in a small scale when a solution of sodium aluminate is neutralized by CO$_2$ at room temperature at pH 9–12. If precipitation is conducted at pH above 12, gibbsite is obtained. It is similar to gibbsite but the layers of Al(OH)$_6$
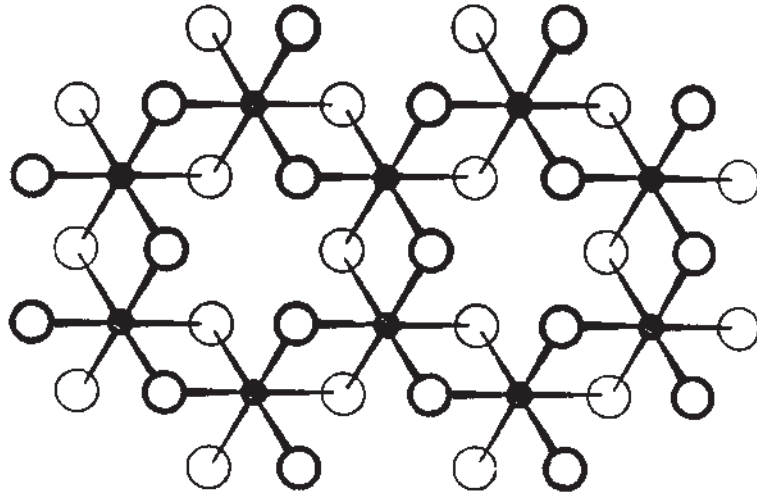
Figure 14 Part of a layer of Al(OH)$_3$. The heavy and light open circles represent OH groups above and below the plane of the Al atoms (shaded).

octahedra are packed in a hexagonal form. The Greek prefixes in this case are used systematically to differentiate between the different structures (and not for high-temperature and low-temperature modifications).

— Nordstrandite, Al(OH)$_3$, is obtained by precipitating a gel from aluminum salt solution with ammonia. Upon aging under the mother liquor at pH 7.5–9, the gel is converted to the crystalline phase that became known as nordstrandite. It was also identified in some soils. Its structure is inbetween that of gibbsite and bayerite.

— Diaspore, α-AlOOH, is aluminum oxide hydroxide that occurs as a major constituent in geologically old bauxites. It is composed of chains of double molecules arranged in a nearly hexagonal close packing:
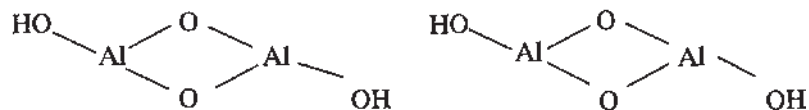
— Boehmite, γ-AlOOH, also occurs in some bauxites, and is composed similar to diaspore, but the chains are arranged in a cubic form.

When the different forms of aluminum hydroxides and oxide hydroxides are heated below 700°C, they yield alumina of high affinity to adsorb water, many organic molecules, and a number of ions from aqueous solutions. The fully dehydroxylated product, α-Al$_2$O$_3$, is inert and is obtained by heating at about 1000°C. During this process, a number of intermediate crystalline forms are obtained. The reaction path and kinetics of dehydroxylation are influenced by the particle size, the rate of heating, the water vapor pressure in the atmosphere surrounding the solid, and the presence of trace impurities.

Gelatinous aluminum hydroxide does not have the symmetry of a crystalline hydroxide. It is composed of Al(OH)$_3$ molecules agglomerated together through

shared OH^- ions and hydrogen bonds, which can include intermediate water molecules. Since the hydrogen bonds are much weaker than covalent bonds, the agglomerate is soft and nonsymmetrical. Alumina is, therefore, always calcined to the α-form known as corundum, which is the most dense, hard, stable and nonreactive aluminum oxide. γ-Al_2O_3 is cubic closest packed with respect to the oxygen ions and α-Al_2O_3 is hexagonal closest packed; aluminum ions occupy interstitial positions. The transformation of γ-Al_2O_3 to α-Al_2O_3 is endothermic ($\Delta H = 20.6$ kcal).

In the Bayer process for leaching bauxite, an alkaline sodium aluminate solution is obtained from which aluminum hydroxide is precipitated:

$$[AlO(OH)_2]^- + H_2O \rightarrow Al(OH)_3 + OH^-$$
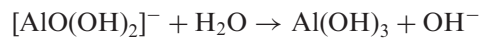
The basis of this process is obtaining a crystalline product (Fig. 15) which is easy to filter and wash, so that it is of the highest purity. To obtain a thoroughly crystalline precipitate, the aluminate solution is stirred with a large excess of aluminum hydroxide "seed". The seed has two functions:

- It reacts with the OH^- ions, thus, shifting the above equilibrium to the right and favoring precipitation.
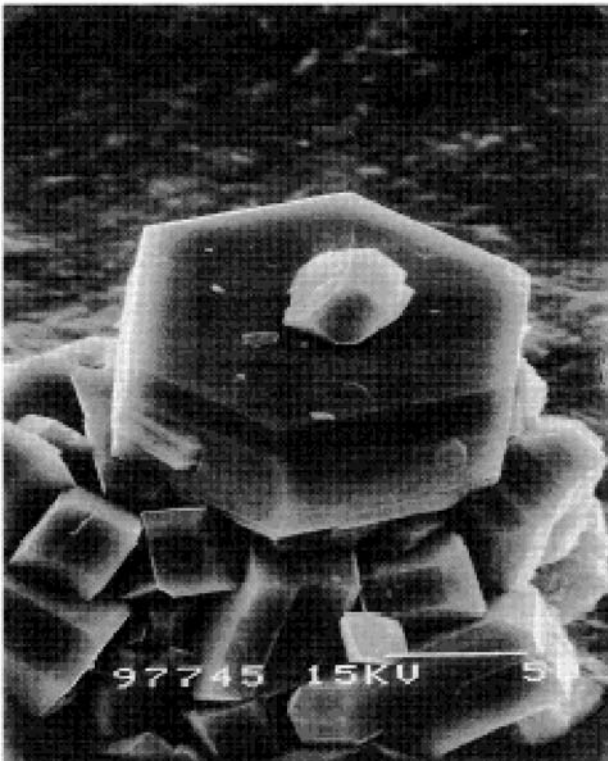
Figure 15 Electron photomicrograph showing the crystalline nature of aluminum hydroxide precipitated from the alkaline solution (Bayer process). The edge of the hexagon is about $8\,\mu$ and its thickness is about $4\,\mu$.
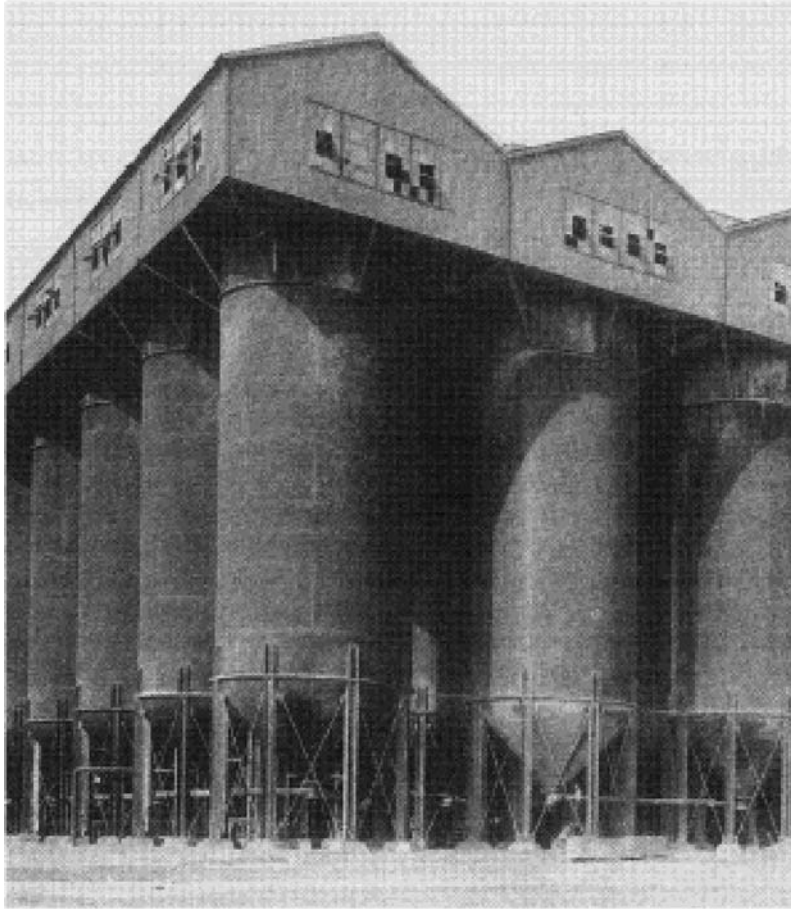
Figure 16 Precipitation tanks for aluminum hydroxide.

- It acts as a nucleus on which $Al(OH)_3$ precipitates.

The solution together with a seed is left in the vessel for about 4 days (Fig. 16). Continuous agitation and cooling to $25-35°C$ is necessary to permit the formation of coarsely crystalline products. A finely crystalline product is undesirable since it causes excessive dusting during calcination. In practice, the operation is conducted as follows: the slurry after precipitation is allowed to settle. The finely suspended crystals are recycled as seed, and coarsely settled crystals are withdrawn, filtered, washed, and calcined (Fig. 16). The mother liquor is evaporated, and its alkalinity is adjusted and recycled for leaching another bauxite batch.

4.3 Calcination of Aluminum Hydroxide

The final operation in the production of alumina is the calcination of the hydroxide:

$$2Al(OH)_3 \rightarrow Al_2O_3 + 3H_2O$$
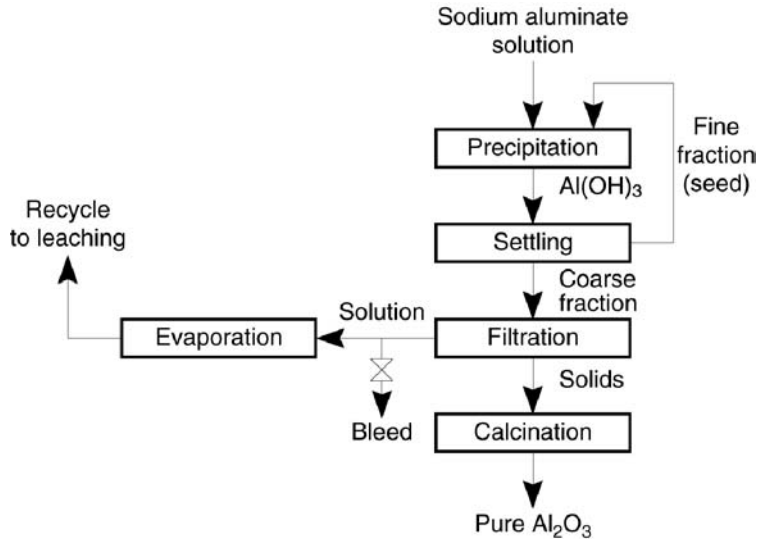
Figure 17 Precipitation flowsheet.

The presence of moisture in alumina for use in the reduction cell for aluminum production is highly undesirable, since it leads to the formation of hydrogen fluoride. Therefore, the desired product is α-Al_2O_3. Calcination of aluminum hydroxide is usually done in rotary kilns, and sometimes, in fluidized beds. Rotary kilns are long, horizontal, brick-lined cylinders slightly inclined to the horizontal to permit the gradual descent of solids at the upper end. They are heated by burning a carbonaceous fuel at the lower end. Aluminum hydroxide is charged to a kiln fired to about 1000°C. The solid product discharged from the other end of the kiln is then cooled in an air cooler (Fig. 17).

5 ELECTROLYTIC REDUCTION OF AL$_2$O$_3$

For the industrial aluminum production, the cathodic product is molten aluminum and the product discharged on the anode is largely CO_2 and minor amounts of CO. The overall cell reaction may be represented by:

$$Al_2O_3 + \frac{3}{2}C \rightarrow 2Al + \frac{3}{2}CO_2$$

$$Al_2O_3 + 3C \rightarrow 2Al + 3CO$$
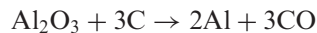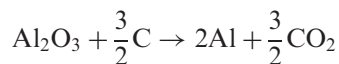
5.1 The Electrolyte

Cryolite, the double fluoride of sodium and aluminum, Na_3AlF_6, was found to be the best molten salt for electrolyzing Al_2O_3 for the following reasons:

- It is a good solvent of alumina.
- It has a greater decomposition potential than alumina.
- It is a good conductor.
- It can be free of metallic impurities which could be deposited.
- It has a low melting point.
- It has sufficient fluidity.
- It has a density less than that of aluminum at the working temperature.
- It has a low vapor pressure.
- It does not react with the electrodes or with the products of electrolysis.

The solubility of Al_2O_3 depends on electrolyte composition and temperature. Alumina and Na_3AlF_6 form a eutectic at 960°C with 11% Al_2O_3. The solubility limit of Al_2O_3 at 1000°C is 13% Al_2O_3. The solubility limit of Al_2O_3 in typical industrial electrolytes is $\approx$ 8%. For the dissolution of alumina in cryolite, the most important factor is the nature of the alumina used: the rate is faster with lower α-alumina content and larger specific area. The percentage of Al_2O_3 in the bath must be carefully controlled within narrow limits for the following reasons:

- Shortage in Al_2O_3 will cause operating difficulties − the anode effect (see later).
- Increased amount of Al_2O_3 will cause its deposition as a sludge under the metal layer. The electrolyte is agitated by:
- The bubbles of gas given off at the anodes.
- The magnetic field effect produced by the large currents passing through the anode bus bars, the cell walls, and the lining (see later).

Agitation is essential to help solubilize the alumina. Were there no agitation, the alumina would sink to the bottom of the cell and accumulate under the metal layer, thus, acting as an insulator to the cathodic carbon. However, too violent an agitation will increase the possibility of short-circuiting in the cell (anode–cathode distance decreases greatly).

5.2   The Cryolite Ratio

Aluminum fluoride is usually added to the electrolyte in excess of the Na_3AlF_6 composition. The ratio NaF/AlF$_3$ is called the cryolite ratio; it is 3 in pure cryolite. Although lowering the cryolite ratio can increase current efficiency of metal production, the change also increases the volatility of the electrolyte. It also adversely affects the electrical conductivity of the bath and the alumina solubility. The bath ratio is the mass ratio of NaF to AlF$_3$ because the molecular weight of AlF$_3$ is almost twice that of NaF, the cryolite ratio is double the value of bath ratio. Industrial cells operate at cryolite ratio between 2 and 3. Operating temperature is usually in the range 940–980°C.

The freezing point of pure cryolite is 1009 ± 1°C, and the cryolite-10% alumina eutectic point is 962°C. The freezing points of the aluminum fluoride–sodium fluoride system are given in Fig. 18. If a direct current be passed through the solution of alumina in cryolite or alumina in cryolite plus other fluorides, the alumina will be decomposed, the aluminum being deposited at the cathode in a molten condition (melting point 660°C) and oxygen at the anode, which
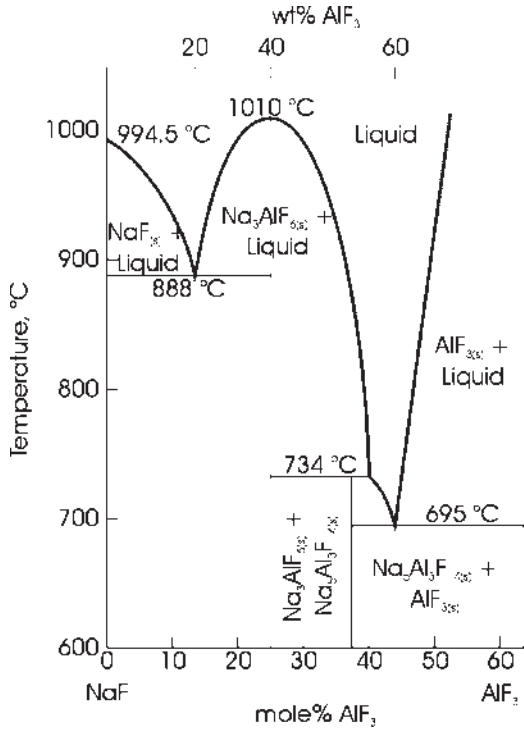
Figure 18 Phase diagram of the system NaF–AlF$_3$.

is carbon, with which the oxygen reacts to form CO_2 and CO. The thermal effect of the oxidation of the carbon anodes is to reduce the amount of electrical energy required to maintain the bath in a fused state. The bath itself is not decomposed by the current. The container is an iron box lined with carbon, referred to as a pot. The density of liquid aluminum at 950–1000°C is 2.3 g/ml (Table 7). Experience has shown the density of the molten electrolyte should be less than 2.1 g/ml to maintain good separation between the metal and fused salt phases. Additions that decrease the density of Na_3AlF_6–Al_2O_3 melts are beneficial.

5.3 Additives

The electrolyte contains calcium fluoride, aluminum fluoride, and sometimes, lithium fluoride or lithium carbonate to lower operating temperature and increase current

Table 7 Density of the Components of the Aluminum Cell

	Density
Molten cryolite	2.10
Molten aluminum	2.28
Solid Al_2O_3	3–4

Table 8 Effect of Additives to Electrolyte in Alumina Electrowinning. Downward Arrow Means a Decrease, Upward Arrow an Increase

	LiF	NaF	CaF$_2$	MgF$_2$	AlF$_3$
Melting point	↓	↓	↓	↓	↓
Al$_2$O$_3$ solubility	↓	↓	↓	↓	↓
Vapor pressure	↓	↓	↓	↓	↑
Density	↓	↓	↑	↑	↓
Viscosity	↓	↓	↑	↑	↓
Surface tension	↑	↑	↑	↑	↓
Electrical conductivity	↑	↑	↓	↓	↓

efficiency. A typical electrolyte consists of 5–15% excess AlF$_3$, 4–6% CaF$_2$, 1–6% Al$_2$O$_3$, and balance cryolite. Aluminum fluoride is added to:

- Neutralize the soda (about 0.6% Na$_2$O) present in the alumina. This results in the formation of cryolite.
- Reduce fog formation and hence increase current efficiency.
  However, it is consumed during operation by:
  - vaporization as sodium tetrafluoroaluminate, NaAlF$_4$.
  - Depletion by hydrolysis with traces of moisture leaking to the cell:

$$2NaAlF_4 + 3H_2O \rightarrow 2NaF + Al_2O_3 + 6HF$$
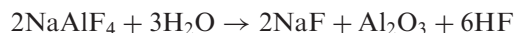
Fluoride losses are almost completely recycled to the cells. Nevertheless, aluminum fluoride consumption amounts to 0.02–0.04 kg/kg of aluminum product. Although the density of alumina is higher than that of cryolite, yet on the addition of alumina, the density of the cryolite melt decreases. The addition of alumina to cryolite lowers the conductivity. Table 8 summarizes the effect of the different additives on the solubility of Al$_2$O$_3$, density, conductivity, etc. The use of additives to lower freezing points and consequently operating temperatures, is always accompanied by a reduced ability to dissolve alumina. Calcium fluoride is seldom added intentionally, because owing to a small amount of calcium oxide impurity in the alumina ($\approx 0.05\%$), it attains a steady-state concentration of 3–8% in the melt. At this level, calcium is codeposited into the aluminum.

Other additives, sometimes, used are MgF$_2$, LiF, and NaCl; all lower the freezing point of the bath which is generally advantageous, but the first two introduce possibly deleterious quantities of Mg and Li into the product and the third, by giving rise to HCl in the gas leaving the cell, generates corrosion and working-condition problems. Deleterious effects on aluminum workability restrict the lithium concentration. Since the fluorides of calcium, barium, and strontium markedly increase the specific gravity of cryolite, the amounts of these substances present must be carefully kept below a point at which the bath would become so dense that the agitation produced in it during normal operation would cause aluminum globules to float and be carried to the anode, where they would be oxidized. In practice, barium and strontium are present only in traces, and the amount of calcium fluoride is kept below 15% so as to leave an adequate margin of safety. The CaO content of the alumina should not exceed 0.04%.

5.4 Mechanism of Reduction of Al_2O_3

Cryolite is believed to be completely ionized into sodium ions and hexafluoro-aluminate ions which partially dissociates further as follows:

$$Na_3AlF_6 \rightarrow 3Na^+ + AlF_6^{3-}$$

$$AlF_6^{3-} \rightarrow AlF_4^- + 2F^-$$
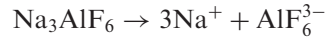
The high solubility of alumina in cryolite is attributed to a chemical reaction to form a complex species such as $Al_2O_2F_4^{2-}$ and $Al_2OF_6^{2-}$:

$$2Al_2O_3 + 2AlF_6^{3-} \rightarrow 3Al_2O_2F_4^{2-}$$

$$Al_2O_3 + 4AlF_6^{3-} \rightarrow 3Al_2OF_6^{2-} + 6F^-$$

Anodic reaction:

$$2AlOF_3^{2-} \rightarrow 2AlF_3 + O_2 + 4e^-$$

Cathodic reactions:

$$Na^+ + e^- \rightarrow Na$$

$$3Na + AlF_3 \rightarrow Al + 3NaF$$

Overall reaction:

$$2Al_2O_3 \rightarrow 4Al + 3O_2$$
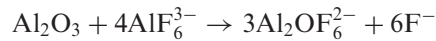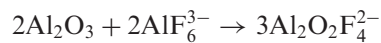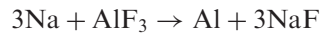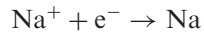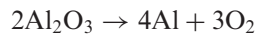
According to this scheme, AlF_3 accumulates at the anode while NaF accumulates at the cathode. This was confirmed experimentally. Theoretically, oxygen depositing onto carbon at the cell operating temperature should form CO with little CO_2:

$$C + O^{2-} \rightarrow CO + 2e^-$$
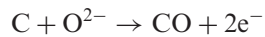
However, based on the net carbon consumption, the anode product is essentially all CO_2. It seems, therefore, that there is an accumulation of oxide ions at the anode so that

$$C + 2O^{2-} \rightarrow CO_2 + 4e^-$$
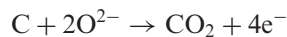
The oxidation of carbon at the anode can be considered to take place according to

$$\frac{3}{2}O_2 + xC \rightarrow (3 - x)CO_2 + (2x - 3)CO$$
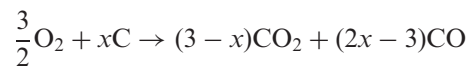
where x is between 1.5 and 3.

5.5 Anode Effect

Gas generated during electrolysis is continuously discharged out of the cell. When, however, the accumulation of gas bubbles at the anode surface becomes excessive, it results in what is known as "anode effect" . It was found that the anode gas composition changes from CO_2 to CO, i.e., more carbon is consumed with significant quantities (3–25%) of CF_4 and minor amounts of C_2F_6. This usually takes place when the bath is depleted of Al_2O_3 and hence fluoride ions are discharged at the anode together with the small amount of oxide ions generated from Al_2O_3.

Gas accumulation at the anode causes decreased current efficiency. The molten electrolyte becomes separated from the anode which is no longer wetted. A number of small arcs are formed between the molten electrolyte and the anode, the resistance of the cell increases markedly, and in consequence, the applied electric tension rises and the current density falls. Further, as a result of the formation of electric arcs between the anode and the electrolyte, the gas film adhering to the electrode will be strongly heated, and as the gas expands the electrolyte is forced further away from the electrode. The anode effect mainly takes place in the Söderberg cell.

5.6 Current Efficiency

Theoretically, 1 kAh of electric current should produce 0.3356 kg of aluminum, but only 85–95% of this amount is obtained. The reasons for this decrease in efficiency are:

- The solubility of aluminum in the electrolyte (fogging phenomenon) which is about 0.1% in a typical industrial electrolyte. This imparts electronic conductivity to the melt, thereby lowering the current efficiency. In addition, the dissolved aluminum reacts as follows:
  - With sodium fluoride to form sodium according to

  $$Al + 3NaF \rightarrow 3Na + AlF_3$$
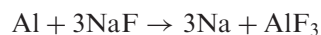
  Metallic sodium has been found in cell lining.
  - With AlF_3 to form aluminum monofluoride gas according to

  $$2Al + AlF \rightarrow 3AlF$$

  - With CO_2 to form alumina:

  $$2Al + 3CO_2 \rightarrow Al_2O_3 + 3CO$$
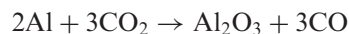
- Reaction of molten aluminum with the cathode lining to form aluminum carbide. The current efficiency is governed by the following parameters:
  - bath temperature;
  - cryolite ratio;
  - anode–cathode distance;
  - current density.

Adjustment of the anodes for the equalization of the current passing through them is an important factor in the operation of the cell. Anodes set too low may

allow projecting points to touch the metal layer in the bottom of the cell, thus short-circuiting it; anodes set too high may, because of the resistance of the thicker layer of electrolyte between them and the cathode, fail to carry their proper share of the current. The anode–cathode distance influences the diffusion of the redissolved aluminum or reducing agent towards the gaseous anodic products. It also determines the power dissipation due to electrical resistance in the electrolyte.

6   ENGINEERING ASPECTS OF ELECTROLYTIC REDUCTION

6.1   Cell Design

- Cells typically range 9–12 m long, 3–4 m wide, and 1–1.2 m high. Thermal insulation surrounds the carbon lining of the cell to control heat losses and is protected by a layer of frozen electrolyte called "ledge".
- Steel collector bars in the carbon cathode conduct electric current from the cell. These bars are either inserted into holes carefully sized so that thermal expansion forms a tight electrical contact, cemented in place with a carbonaceous cement containing metal particles, or bonded in place with cast iron.
- As a result of intensive research and development in the aluminum industry in the period 1920–1990:
  - Cell production increased from 90 to about 1000 kg aluminum/day (Fig. 19).
  - Energy consumed per kilogram of aluminum decreased from 25 to nearly 12 kWh (Fig. 20).
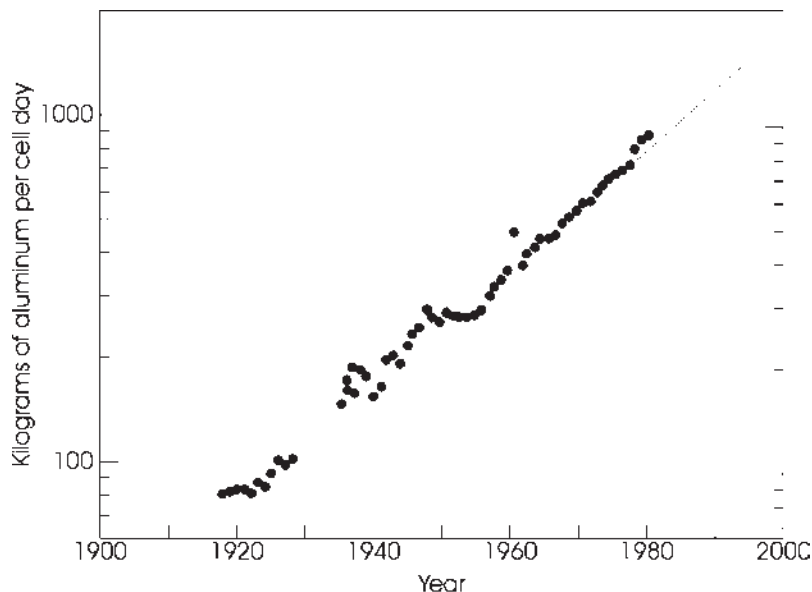
Figure 19   Trend in the production capacity of a single electrolytic cell for aluminum production.
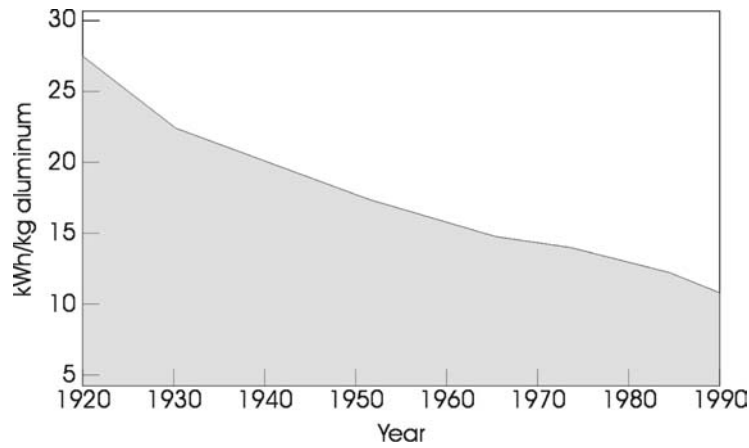
Figure 20   Energy consumption/kilogram of aluminum.
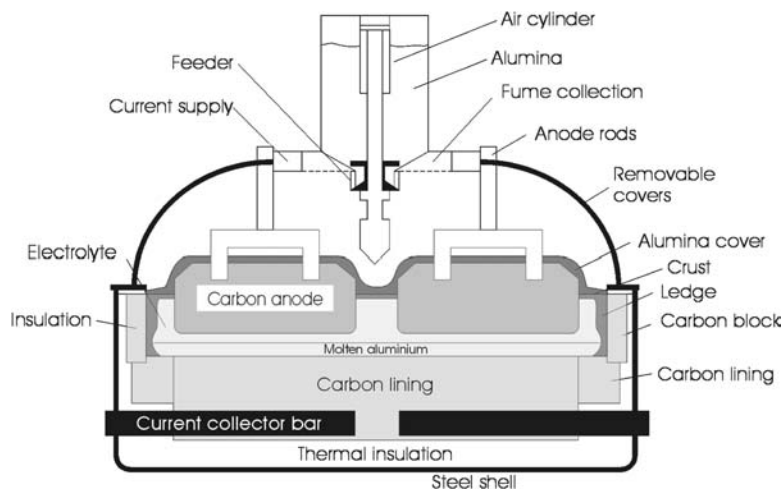
Figure 21   Electrolytic cell with prebaked anodes.

- The molten aluminum layer is 3–25 cm thick, the electrolyte layer 15–30 cm, the anodes project through the frozen crust so that the anode–cathode distance is 4–8 cm.
- There are two types of reduction cells, based on the form of the anode: the prebaked (Fig. 21) and the self-baking (Fig. 22). Historically, the aluminum industry started with the prebaked electrode which was found to be labor-intensive because of the frequent changing of electrodes. The continuous, self-baking anode was developed by Carl W. Söderberg (Fig. 23) in Norway during World War I. Their peak period was in the 1940s and 1950s.
- The prebaked anode cell became the preferred design for two reasons:
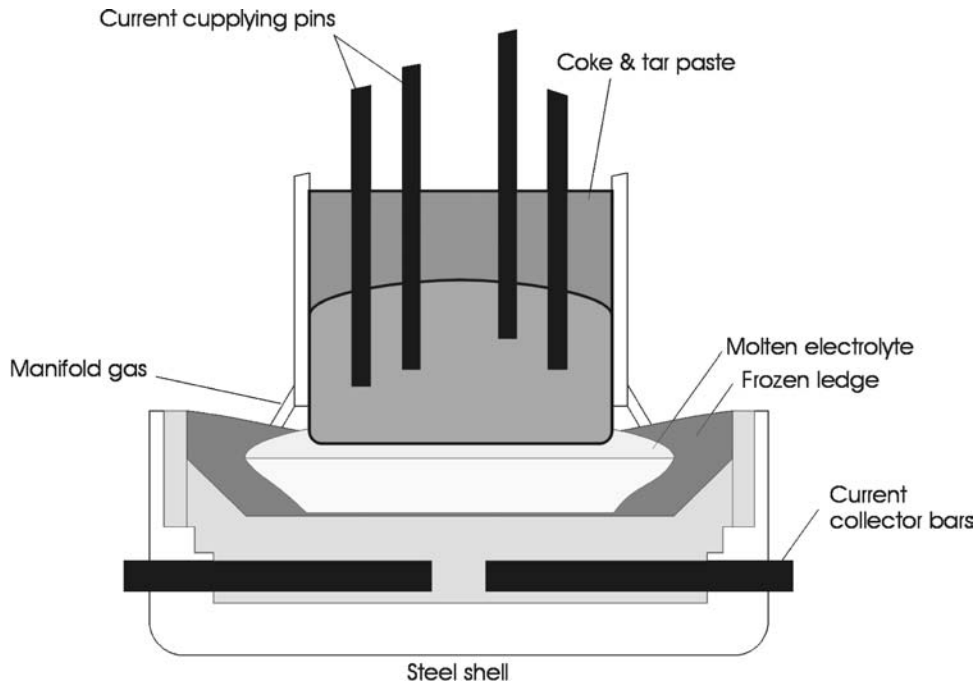
Figure 22   Aluminum electrolyzing cell with self-baking anode.

- — Development of the vibrocompacting method for producing prebaked anode blocks. This took the place of the pressing–extruding method previously used.
- — Environmental problems associated with the operation of the Söderberg electrode because of the emission of carcinogenic compounds during the self-baking process which was difficult to capture. In the prebaked system, the volatile organic matter is burned in the baking furnace and the heat generated improves the thermal efficiency of the process.
- In the prebaked cell, the anodes are mounted on a beam that can be adjusted vertically. Alumina feeders are placed in the center aisle between two rows of anodes. The cell is covered to permit collection of off-gases for processing in a gas-purifying system where volatile fluorides and other impurities are adsorbed on Al_2O_3. In the Söderberg cell, the massive anode gradually descends at the rate of its consumption, and Al_2O_3 is added on the sides after breaking the crust.

6.2   Cell Arrangement

Since each cell takes a large current at low voltage, a number of cells (120−168) are arranged in series. The line voltage may be up to 800 V, and cell amperages 34,000−130,000 A. Figure 24 shows the electrical connections to one type of electrolytic cells.

Figure 23   Carl Wilhem Söderberg (1876–1955) invented the self-baking anode for the aluminum electrolytic cell.

6.3   Alumina Feeding

- In prebaked electrode cells, alumina feeding is more frequent and in smaller quantities through the center break-and-feed system which is composed of a volumetric alumina metering section and a mechanism for punching a hole in the crust. There are usually 3 to 5 point feeders fitted to each cell, each capable of delivering 1–3 kg of alumina. Because each feeder is actuated frequently, the holes often remain open, thus providing an escape vent for the anode gas. A general view showing the alumina supply is shown in Fig. 25.
- In Söderberg cells, feeding is done on the sides of the electrode.

6.4   Discharge of Metal

When a sufficient amount of metal has accumulated in the bottom of the cell, it must be removed by tapping, ladling, or siphoning. This may be done every day, every second day or third day, depending upon the design of the cell. The molten aluminum is usually stored into a holding furnace so that the metal from many cells is blended. Practically, all commercial aluminum is remelted or heated in holding furnaces to ensure uniformity of compositions and the removal of nonmetallic impurities.
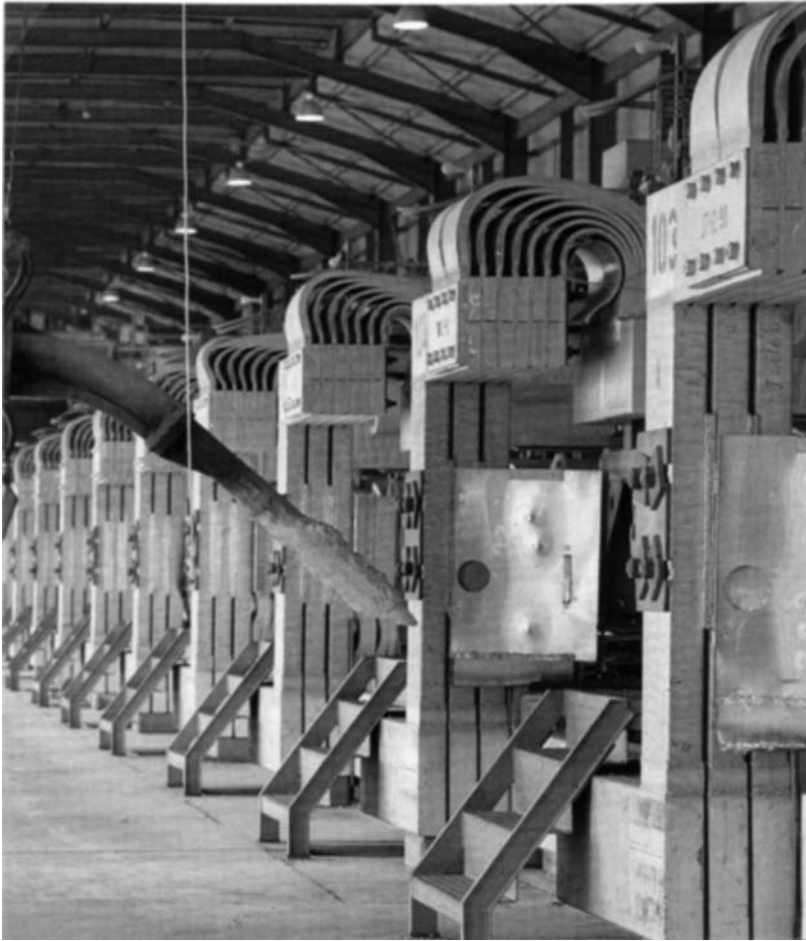
Figure 24 Electrical connections to one type of electrolytic cells.

6.5   Energy Requirement

The voltage required for the deposition of aluminum is 4.64 V About 38% of cell power (voltage) is lost due to the electrical resistance of the electrolyte. Reducing the anode–cathode distance is one way to decrease this loss. But this is limited by two factors:

- Increased reoxidation of reduced species.
- Increased possibility of short-circuiting due to magnetic turbulence (see below).

In addition to the heating effect due to electrolyte resistance, a large amount of heat is generated due to the anodic reaction (oxidation of carbon). Loss of heat by radiation results in the formation of a frozen crust on the fused salt and a frozen ledge on the sides of the cell. The formation of the ledge is desirable because it protects the sidewalls of the cell against attack by the molten salt. The thickness of
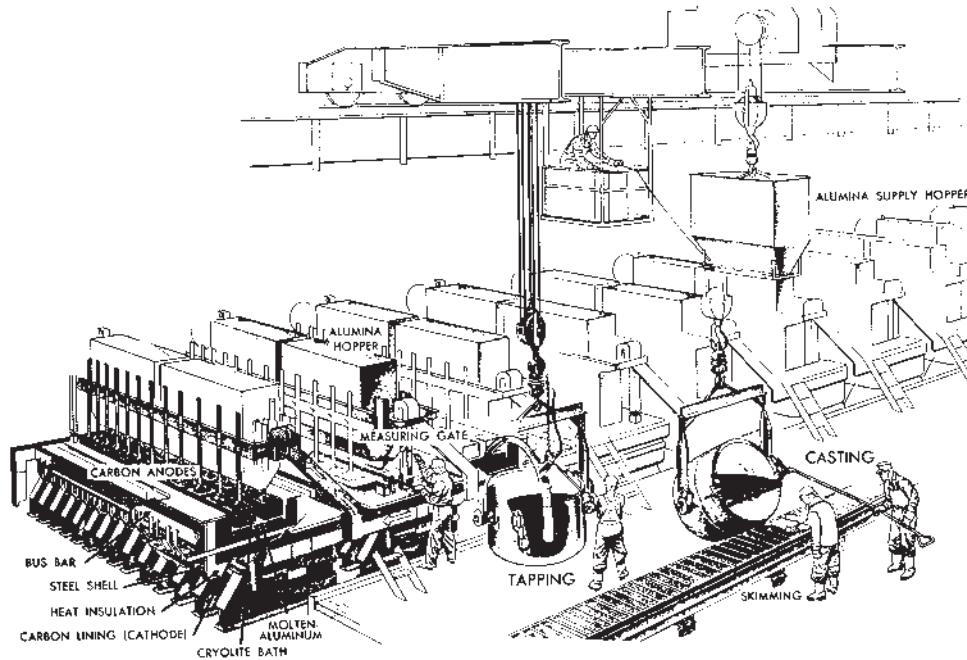
Figure 25 General view showing the alumina supply hoppers in cells using prebaked anodes.

the ledge is a function of cell temperature: when the temperature increases, the thickness decreases and vice versa.

6.6 Cell Control

In a prebaked cell, there are usually 24 anodes, each is at a different stage of consumption and, therefore, has unique electrical properties. While oxidation of the anodes is offset by metal accumulation, individual anodes must still be adjusted so that each operates at the same amperage. Since new and partially consumed anodes have different electrical resistances and different surface areas, the anode–cathode distance must differ when the anodes are set to draw the same current. Precisely controlling such a system, so that each anode carries an equal share of the total current and is positioned slightly above the undulating metal pad while avoiding shorting, takes careful instrumentation and computer control.

6.7 Fluid Dynamics

- The large electric currents used in modern cells generate strong magnetic fields within the cell which exert forces producing movement of liquid conductors (molten aluminum pad) which may lead to electrical short-circuiting.
- Gas bubbles produce significant stirring in the electrolyte and are the dominant force in bath movement (Fig. 26).
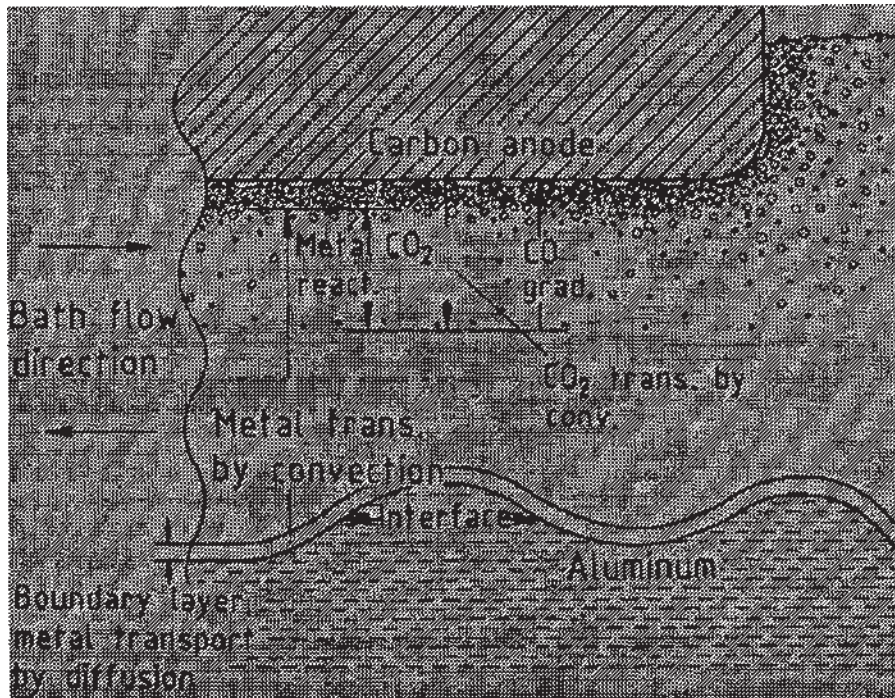
Figure 26 Gas generation at the anode surface.

- Natural convection due to temperature gradients or composition differences in the electrolyte are insignificant compared with the movement induced by magnetic forces and gas bubbles.

Today, cells over 250,000 A have been designed and operated with high efficiency by controlling magnetic turbulence to produce a more quiescent metal pad, thus permitting a reduction in the anode–cathode distance.

6.8 Anodes

Although there are two fundamental designs of anodes, they are both formulated from similar materials and undergo the same reactions within the cell. Petroleum coke of highest quality and a binding pitch (coal tar or petroleum pitch) are used for fabricating the anode. On pyrolysis, the pitch rigidly bonds the particulate coke. Anode quality affects both the energy efficiency and productivity of cells.

- About 10% of cell power is used to overcome the electrical resistance of the anodes.
- Dusting is caused by selective oxidation of the binder coke at the anode–bath interface, which releases aggregate particles into the electrolyte. Anode carbon that enters the electrolyte as dust is unavailable for electrolysis and adversely affects the current efficiency.

For environmental reasons, there has been an increasing tendency to move to the prebaked anodes. A comparison between the two anodes is given in Table 9.

Table 9 Comparison Between Prebaked and Self-baking Anodes (Söderberg)

	Prebaked anodes	Söderberg
Electrical resistivity	Low	High
Density	High	Low
Pitch content, %	14–18	25–35
Anode consumption	Low	High
Power efficiency	High	Low
Anode effect	Rare	Frequent
Manufacturing cost	High	Low
Labor	High	Low
Al_2O_3 feeding	Central (easy)	Sidewise (difficult)
Collection of fumes	Easy	Difficult
Environmental problems	Low	High

6.8.1 Prebaked Anodes

These are formed by blending sized petroleum coke aggregate, crushed spent anodes, and coal tar pitch, molding this into blocks (with preformed electrical connection sockets) by pressing or vibration, and firing to $\approx 1100°C$ in oil or gas-fired furnaces. Blending of sized fractions is important so that a compact solid can be obtained. A typical block is 70 cm wide $\times$ 125 cm long $\times$ 50 cm high. Electrical contact is made through aluminum or copper rods welded or bolted to steel stubs fixed in the anode socket by poured cast iron. The electrical resistivity of prebaked anode ranges from 0.005 to 0.006 Ω cm. Anode current density ranges from 6000 to 13,000 A/m^2.

Anode-baking furnaces are high-temperature heat exchangers applying indirect heating by combustion gases from fired fossil fuel (Fig. 27). The formed anode blocks are located in four to eight pits in each heat-exchanger section, with granular coke protecting against oxidation by flue gases and providing heat transfer from the flue walls. A furnace includes 28–48 sections in series located in two parallel rows with crossover connections at each end. Fans located at the outlet scrubber end provide a draft through a joint exhaust pipe surrounding the furnace. Ambient air is, thus, pulled through sections on cooling, to recover heat before entering the peak-fired and preheating sections. Exhaust manifolds connect each firing zone to the joint exhaust pipe.

On completion of the heat treatment, one section is disconnected at the cooling end while a new section is connected at the preheat end by moving the exhaust manifold to the next section in front. In this way, the firing zone is moved around the furnace. There are two main types of design:

- Two furnace concepts. The horizontal-flue (or open-top furnace), characterized by gas flowing through separate rows of flue-walls without coming into contact with the anode loads.
- The vertical-flue (or closed-top furnace). Gas flows are mixed in each section within the space between the lid and the pit top.

The pits are flanked on either side by combustion flues. Although the pits are separated by ceramic wall headers, the flues are interconnected. Before passing to the
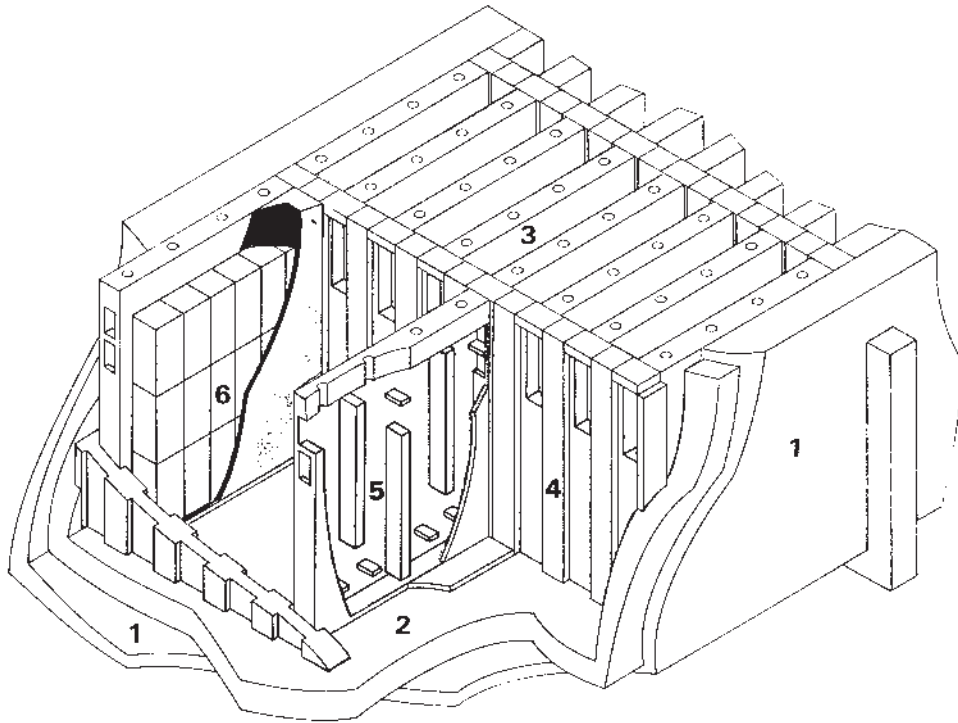
Figure 27 Anode-baking furnace: (1) concrete casting, (2) insulation, (3) section, (4) head wall, (5) flue channel, and (6) anode blocks and packing coke.

next flue, combustion products are forced to the bottom of the flue by ceramic baffles. An oil or gas air mixture is injected at the fuel input of the pit under fire and the combustion products pass through the next three pit flues before exiting by way of the waste gas manifold. Air for fuel combustion is preheated as it is drawn upstream of the fire through the three flues, containing previously baked anodes. The fuel injected represents about half of the heat necessary for baking; the other half comes from volatile matter in the pitch which theoretically should provide more than enough heat to bake anodes.

Baked anodes are cooled by drawing air through the flues, producing preheated air for a regenerative burner system. The combustion products from the pit under fire operate in a recuperative manner by preheating the anodes. The waste gas manifold ties all of these rows of pits together by an underground flue at opposite sides of the pit assembly. When the anodes in a given pit reach the finishing temperature, the fire and manifold are moved to the next pit and the process is repeated. Maintaining uniform temperature distribution so that all anodes in a pit attain the desired temperature is an important requirement to avoid anode dusting.

While nearly all large aluminum producers have their own local facilities for manufacturing prebaked anodes, some depend on purchased anodes. The anode manufacturing complex in Rotterdam is the world's largest facility of this type, with an average capacity of 300,000 tons/year of prebaked anodes. The world's leading
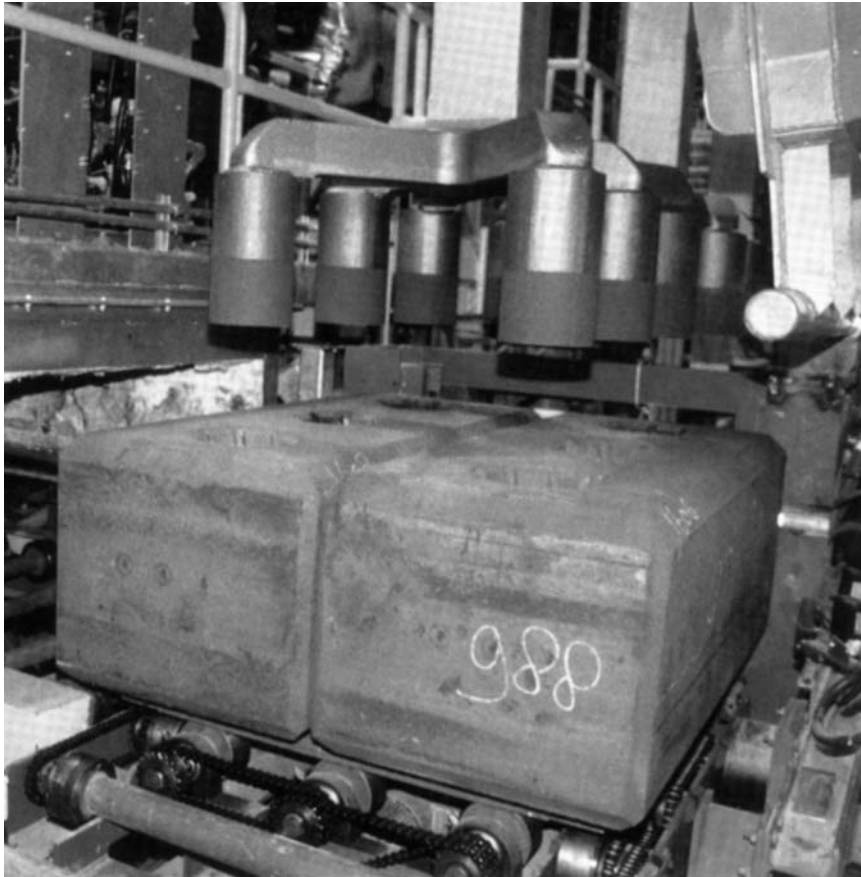
Figure 28 Rodding of anode blocks.

firm in the supply of anode paste plants for aluminum smelters is Buss AG, Pratteln, Switzerland.

Anode rodding (Fig. 28) serves to connect the anode blocks with the anode rods that feed electric current to the cells. The pins are first coated with graphite. The coating is then allowed to dry. The six stub holes in the anodes are heated to eliminate any moisture being present in the holes prior to casting (Fig. 29). Finally, the anode blocks can be aluminum-sprayed before being ready for use. For proper functioning, the anode rod requires a reliable steel–aluminum transition joint, connecting the steel stub assembly with the aluminum anode rod. Consumed anodes are removed from the cell for replacement.

6.8.2 Söderberg Anodes

The Söderberg anode uses a premixed paste of petroleum coke and coal tar pitch which is added at the top of a rectangular steel casing that is typically 6–8 m long, 2 m wide, and 1 m high. Heat from the electrolyte and from the electric current bakes the carbonaceous mix as it passes through the casing to replace

Figure 29 Fixing the rods in the anode blocks by pouring molten cast iron.

anode being consumed at the bottom surface. The baked portion extends past the casing and into the molten electrolyte.

The anode briquettes are made from petroleum coke and hard pitch. The coke is ground, blended, mixed with crushed pitch at 150°C, briquetted, then dumped into the top of the anode shell. The temperature at the top of the pot is high enough to melt the briquettes. The temperature in the lower part of the anode shell is high enough to distill the volatile matter in the briquettes and bake the anode into a single solid block. Electrical contact to the anode is achieved by steel pins (about 8 cm diameter), introduced through the pitch—carbon mixture, which is slowly converted to a carbonized mass when moving downwards to hotter zones as a consequence of the anode consumption. There are two designs depending on the method of making the electrical contact:

- *Horizontal Stud Electrodes.* These spikes are pulled and reset to a higher level as they approach the lower surface.
- *Vertical Stud Electrodes.* These are the more common electrodes as they are less labor-intensive and cause fewer operating disturbances. The pins are simply readjusted to compensate for the consumption of the carbon.

Although Söderberg anodes save the capital, labor, and energy required to manufacture prebaked anodes, their use is declining due to the following factors:

- Environmental problems associated with its operation.

- Higher electrical resistivity as compared to the prebaked electrode which causes lower power efficiency.
- Development of improved methods for fabricating the prebaked electrode, e.g., the vibrocompacting method.

6.9 Cathodes

The pool of aluminum in the bottom of the pot is the cathode. Contact is made by iron bars buried in the carbon lining. The carbon cathode must have adequate strength, good electrical conductivity, and be so fashioned that it will remain in place and carry the current to the metallic aluminum for 2–6 years. Overheating and local stresses cause it to crack and disintegrate. Broken pieces of lining may then float in the bath and cause partial short circuits between the anodes and the metal. Breaks in the lining may permit the molten aluminum metal to attack the steel shell. In this case, it must be removed from the line for repair. The carbon lining is somewhat porous and absorbs nearly its own weight of the fused electrolyte. Borides and carbides of titanium and zirconium are being considered as future cathode material that could increase the life of the cell.

Anthracite and to some extent graphite and metallurgical coke are the major constituent in the cell cathode blocks. The anthracite is calcined at 1200°C, crushed and sized, mixed with coal tar pitch, molded into blocks and baked. These, along with carbonaceous seam mix, form the potlining. High purity is not as important as in the anodes since the dissolution of impurities is very slow. Consumption of cathode carbon amounts to 0.02–0.04 kg carbon/kg of aluminum produced. Two types of cathode linings are employed.

- Ramming a hot mixture of pulverized coke with tar and pitch binders into the steel shell, using a suitable cast-iron form to give the cavity the desired shape. The entire pot is baked in a furnace at about 600–800°C.
- Preformed and prebaked carbon blocks are used to build up the lining, the blocks being cemented together with a mixture of tar, pitch, and ground coke.

7 PURIFICATION AND REFINING

Molten aluminum tapped from the cell is transported to holding furnaces where recycled aluminum scrap is melted and alloyed. Oil and gas-fired reverberatory furnaces are commonly used to melt ingots for casting. To produce sound ingot, molten aluminum must be treated to remove nonmetallic inclusions, dissolved hydrogen, sodium, calcium, and other impurities. This purification is accomplished by a variety of systems where molten metal is filtered and/or fluxed during transfer between the holding furnace and the casting station.

In the rotating gas disperser method, argon–chlorine mixture is introduced through the shaft and exits beneath the disperser. The rotating action of the disperser forms tiny bubbles of gas that move upward through the molten metal, creating an efficient cleansing action. Sodium and calcium levels are readily controlled in the 1–2 ppm range. Hydrogen reduction from 0.5 ml/100 g down to 0.1 ml/100 g are standard for most alloys at common metal flow rates. Some of the alloyed metal is shipped in molten form on special trucks to customers, e.g., in Germany.

Table 10 Electrolytic Refining Processes for Aluminum

	Hoopes process (Alcoa)	Gadeau process (Pechiney)	Aluminum industries A-G (Neuhausen)
Electrolyte composition, %			
AlF_3	30–38	23	48
NaF	25–30	17	18
BaF_2	30–38	–	18
$BaCl_2$	–	60	–
CaF_2	–	–	16
Density, g/cm³	2.5 at 1000°C	2.7 at 740°C	2.5 at 740°C
Electrical resistivity, Ω/cm³	0.3 at 1000°C	0.75–0.85 at 740°C	1.1 at 740°C
Anode layer			
Al, %	67	70	
Cu, %	33	30	
Density, g/cm³	3.14 at 740°C	3.05 at 740°C	
Cathode layer	Pure aluminum	Pure aluminum	Pure aluminum
Density, g/cm³	2.29 at 1000°C	2.3 at 740°C	2.3 at 740°C
Cell characteristics			
Temperature, °C	900–1000	740	740
Voltage	5–7	6–7.5	5.5–5.8
Amperage	20,000	25,000	14,000
Current efficiency, %	–	96–98	93
Energy requirements, kWh/kg	–	21	18
Graphite usage, kg/kg metal produced	–	0.035	0.04–0.06
Purity of product, % Al	99.90–99.98	99.99+	99.99+

Aluminum is the only metal refined electrolytically in a fused salt on an industrial scale. Aluminum produced by the Hall–Héroult process is normally 99.5–99.8% Al. The major impurities are silicon and iron. Higher-purity metal (99.99%) is mainly needed for electrolytic condensers. Two processes are used: electrolytic refining and the segregation process.

7.1 Electrolytic Refining

All electrolytic processes used (Table 10) are based on transferring aluminum from a lower anodic molten layer of impure aluminum, via an intermediate layer of molten salts, to a top cathodic layer of pure aluminum. The anode layer of impure aluminum is alloyed with 25–33% copper to make the density greater than that of the electrolyte at the working temperature of the bath which varies from 740°C to 1000°C. The electrolyte is a molten mixture of sodium fluoride and aluminum fluoride with other additions to give the desired density.

Because of the small difference in density between the three layers, the rapid increase or decrease in the density of the electrolyte with falling or rising temperature, the changes in anode alloy composition with time, and the possibility of circulation in the layers owing to the heavy currents in the associated magnetic fields, there was always the danger of mixing occurring, and in consequence, the contamination of the pure metal layer by the copper anode alloy. Operating the cell, therefore, requires great attention.

The cell is a mild-steel box lined with refractory brick (Fig. 30). The hearth consists of carbon blocks, cemented with carbon jointing paste, in which steel bars are embedded. The sidewalls are lined with magnesite bricks, backed by firebrick. Connection is made to the cathode layer by graphite blocks suspended from overhead. The graphite blocks are usually jacketed to prevent oxidation by casting a 1 cm thick layer of pure aluminum around them. The anode layer is 25–35 cm deep, the electrolyte 15–25 cm thick, and the pure aluminum cathode layer 15–20 cm thick.

The aluminum to be refined is fed at intervals to the copper alloy through a wide graphite tube reaching down to the anode layer. Slow pouring ensures thorough mixing with the copper layer and avoids the floating of the impure aluminum upward through the electrolyte. The cell may be tapped daily by removing some metal in graphite ladles from the cathode layer. All operations must be carried out with a minimum of disturbance to the fluid layers.
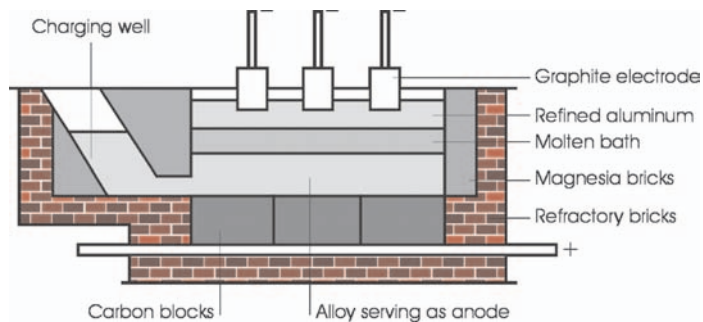
Figure 30   Cell for the electrorefining of aluminum in a fused salt.

Although little energy is theoretically required for the refining process, electrical power consumption is usually high – about 14,000–18,000 kWh/ton. The amperage of the three-layer cell is generally 30–60 kA, which is substantially lower than that of the Hall–Héroult process cell.

7.2 Segregation Process

Although the principle of purification utilizing the segregation phenomenon during solidification was well known, the process was first industrialized by Pechiney in 1975. From 1981 to 1983, some Japanese companies independently industrialized the process, and the production capacity increased to about twice of that of the electrolytic process. The reason for this increase is that high-purity aluminum produced by the segregation process became usable for aluminum electrolytic condensers through improvements of electrolytic condensers in 1982. Refined purity is generally 99.98–99.99%, depending on the purity of the raw aluminum. The segregation process has several advantages compared with the electrolytic process, including lower plant and equipment cost, lower energy consumption and more flexible produc tion rate.

There are two types of segregation processes:

- The fractional crystallization process, in which purified aluminum crystals are formed to be separated from the liquid aluminum and are compacted to remelt and fractionally crystallize. In the Pechiney process, precipitated high-purity crystals are tamped in the bottom of the vessel to squeeze out molten impure metal between the crystals and fractional remelting and the growth of crystals occurs by appropriate heating.
- The unidirectional solidification or zone melting process, in which solidification progresses in the direction inverse to the heat flow. In actual solidification, a high-impurity layer is formed in the liquid in the immediate vicinity of the liquid–solid interface. The impurity-concentrated layer is kept as thin as possible by mechanical stirring or by ultrasonic vibration of the liquid, and by heating the liquid while cooling the solid in order to get high purity and productivity. This process is suitable for mass production. To get higher purity, the refining process is repeated two or more times.

7.3 Ultrahigh-purity Aluminum

Ultrahigh-purity aluminum (meaning >99.999%) is produced by a combination of the electrolytic process and the segregation process or the zone melting process. In recent years, 99.9999% ultrahigh-purity aluminum has been produced. This material has uranium and thorium impurities of less than 1 ppb. The reduced content of radioactive elements is necessary to prevent errors associated with α-particle emission when aluminum is used as a conductor for ultralarge-scale integrated circuits.

8 ENVIRONMENTAL ASPECTS

The aluminum industry was one of the most polluting metal industries until recently when excellent measures were taken to render it a nonpolluting industry. Petroleum coke contains small amounts of sulfur; about 20% of this sulfur is emitted during the manufacture of the electrodes and the remaining during reduction. Organics

identified in the waste stream include benzo(α)anthracene, benzofluoranthene, benzol(α)pyrene, and phenanthrene; some of these are carcinogenic.

The amount of fluorine compounds emitted from a cell is at about 30 kg fluorine/ton aluminum produced. Reminding that the world production of aluminum is about 80 million tons, then fluorine emission from this source is about 2.4 million tons/year. Present technology eliminates or recovers about a half of this amount. Fluorine compounds originate from:

- Volatilization of bath components, e.g., AlF_3.
- Reaction with air humidity and the formation of HF.
- Irregular operation during the anode effect.

The carbon oxides, diluted with ≈ 100 parts of air and containing fluoride values, are passed through dry or wet scrubbing systems where the fluorides are collected and returned to the cell. Carbon oxides are discharged to the atmosphere.

- *Dry Scrubbing*. The hydrogen fluoride is chemisorbed directly onto the crystalline alumina from the dry hot exhaust gas stream, which mixes and reacts with the alumina. The reacted alumina, as well as the particulate fluorides and any other particulates, are removed from the exhaust gas stream by bag filtration. The alumina collected is then fed to the cells (Fig. 31).
- *Wet Scrubbing*. Wet scrubbers are used to remove sulfur dioxide downstream of the dry scrubber. Wet scrubbers are also used to clean flue gases from the anode-baking furnace (Fig. 32).
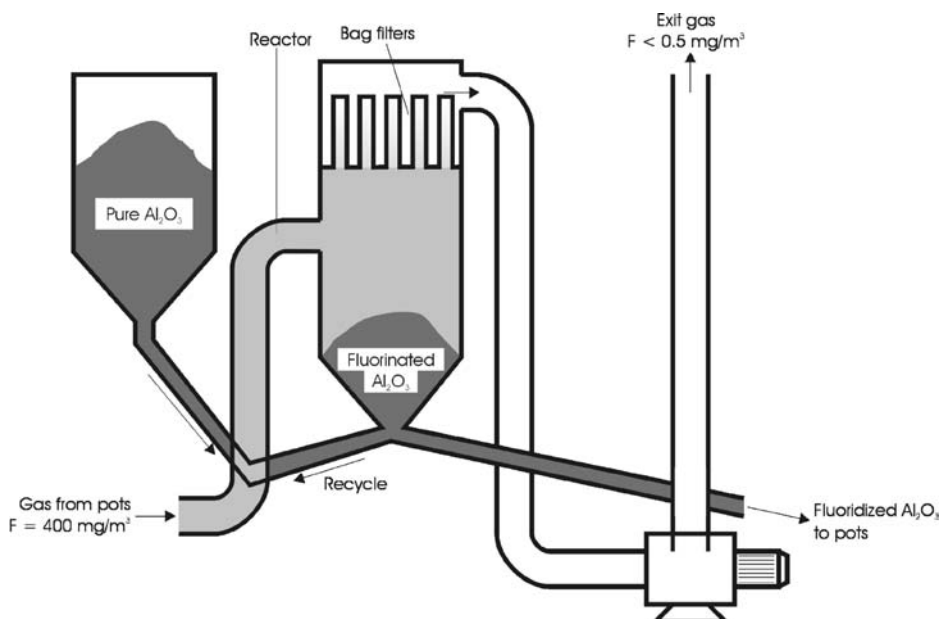
Figure 31   Absorption of fluorine compounds from gases leaving electrolytic cells by Al_2O_3 entering the cells.

Figure 32 Fume treatment system for controlling anode-baking furnace emissions.

8.1 Spent Potlining

After a certain period of operation, the graphite which acts as cathodes deteriorates and has to be replaced. This is done by interrupting the operation, draining the electrolyte, allowing to cool, then replacing the blocks. The material removed is referred to as *spent potlining* and represents a large volume of solid waste. This material is hazardous because of its cyanide content that is mostly water-soluble (Table 11). The cyanide forms as a result of air leaks in the cell whose nitrogen content reacts with carbon at the high temperature of the operating cell. The reaction is known to be catalyzed by sodium salts in the same way as in the iron blast furnace.

Table 11 Typical Analysis of Spent Pot Lining

	Percentage
F	6–10
Na	12–19
Ca	1–2
CN^-	0–1

Present disposal/management practices include temporary storage in curbed, roofed, protected areas to minimize leachate generation, and burial in landfills lined with impervious materials to prevent contamination of the surrounding environment. These procedures are costly, because of the large volumes involved. Because of its carbon and fluoride content, current research is directed towards recovering or utilizing these valuable components. Potential routes for spent potlining include:

- Recovery and reuse of the fluoride and carbon constituents.
- Use as a flux material in the steel industry.
- Use as a supplemental fuel by cement manufacturers.
- Use as fuel in fluidized bed boilers.
- Destruction of cyanide in solution at high temperature and pressure.
- Crushing, grinding, mixing with $CaSO_4$ then injecting into a furnace to destroy the cyanides and bond the fluorides as CaF_2.

9   DROSS TREATMENT

Dross is a major by-product of all processes involving molten aluminum. It is formed at the surface of the molten metal as it reacts with the furnace atmosphere. Up to 5% of the molten metal is turned into dross. This dross may contain as much as 75% of the free aluminum in the form of very small droplets entrapped in aluminum oxides and depending upon the metal being melted ingot or scrap, it can also contain lead, cadmium, or chromium. Processes have been developed to solve this problem.

The aluminum in the dross is usually recovered in oil or gas-fired rotary furnaces, various salts being added to increase the amount of aluminum recovered. This treatment produces a secondary dross containing alumina, salts, impurities, and a small amount of aluminum. Dross produced by melting pure aluminum is known as "white dross", and the dross containing contaminants such as salt fluxes and other metals is known as "black dross". Treatment in plasma furnace or electric arc furnace has also been suggested.

10   RECYCLE OF ALUMINUM

Aluminum scrap recovery has become an important low energy source of metal that requires only 5% as much energy to recycle as it requires to produce it from bauxite. A major recycling effort is aimed at beverage containers. New scrap, generated by plants making wrought and cast products, and old scrap recovered from obsolete products discarded by consumers, is returned to plants that cast ingot to produce mill products. Effective recycling of old scrap also lessens the consumption of bauxite, petroleum coke, and aluminum fluoride.

11   OTHER ROUTES FOR ALUMINUM

Although the Hall–Héroult process has gained industrial dominance, it has several disadvantages. The most serious are the high capital investment required and the high consumption of costly electrical power. There are also the costs of the alumina plant and of the carbon anode plant. Many of the aluminum-producing countries must import alumina or bauxite. The supply of petroleum coke is limited. These deficiencies have motivated researchers to find alternate processes.

11.1 Nonconsumable Anodes

Instead of reacting the oxygen generated at the anode with the anode material (carbon) to form CO and CO_2, it was conceived to use an inert anode to eliminate carbon as a raw material and recover oxygen as a by-product. The only known materials with any chance of success as nonconsumable anodes are metallic oxides, e.g., NiO and SnO_2. The advantages of inert anodes would be:

- Low labor intensity.
- Accurate and equidistant anode–cathode separation.
- Elimination of baking ovens.
- Production of a moderately valuable by-product, oxygen.
- Freedom from petroleum coke.

Tests, however, did not go beyond laboratory scale.

11.2 Composite Anodes

Anodes made of a mixture of 85% Al_2O_3 and 15% C, and baked at 900°C, were tested in a molten chloride–fluoride bath at 700°C as a possible substitute for the present technology. The advantages are: smaller anode–cathode distance and low energy consumption of about 9 kWh/kg Al, as compared to 12 kWh/kg in the present technology. However, the disadvantage is the larger weights and volumes of the anodes baked and handled.

11.3 Carbothermic Reduction

Crude aluminum or aluminum alloys may be produced by carbothermic reduction of Al_2O_3 or aluminous ores and SiO_2 at about 2000°C, followed by refining. Such process may have the following advantages: low capital and operating costs since reduction reactors usually have high productivity. Fluorides are not required. Bauxite can be reduced directly and local clays can be used. The carbon may be derived from coal. Total electrical energy is low if credit is taken for by-products. Carbon monoxide may be used to generate power.

To date, no such process has been commercially successful due to low yields, the high temperatures involved, and inefficient refining techniques. Only in Russia, aluminosilicate ores are smelted in electric arc furnaces to produce 60% Al–Si alloys.

SUGGESTED READINGS

1. W. B. Frank, et al., Aluminum, in *Handbook of Extractive Metallurgy*, (F. Habashi, ed.), Vol. 2, WILEY-VCH, Weinheim, Germany, 1997, pp. 1039–1127.
2. F. Habashi, *Principles of Extractive Metallurgy*, Vol. 4, Amalgam & Electrometallurgy, Sainte Foy, Quebec, Métallurgie Extractive Quebec, 1998, pp. 201–240 and 287–291. Distributed by Laval University Bookstore "Zone".
3. F. Habashi, *A Textbook of Hydrometallurgy*, 2nd Edn, Sainte Foy, Quebec, Métallurgie Extractive Quebec, 1999, pp. 101–195, 241–254, and 600–605. Distributed by Laval University Bookstore "Zone".
4. F. Habashi, *Kinetics of Metallurgical Processes*, Sainte Foy, Quebec, Métallurgie Extractive Quebec, 1999, pp. 187–190, 238–240, and 260–266. Distributed by Laval University Bookstore "Zone".
5. F. Habashi, *A History of Metallurgy*, Sainte Foy, Quebec, Métallurgie Extractive Quebec, 1994.

2

Smelting of Aluminum

MICHAEL M. GASIK

Helsinki University of Technology, Espoo, Finland

MICHAEL I. GASIK

National Metallurgical Academy of Ukraine, Dnipropetrovsk, Ukraine

1   INTRODUCTION

In this chapter, the basics of the aluminum reduction from oxides and other aluminum compounds are considered. The most important and practically tested processes are discussed together with their thermodynamics. The advantages and disadvantages of these processes are compared and new possibilities in aluminum production are presented. Because aluminum electrolysis is a well-documented process described in many handbooks since its invention, more attention is paid to the modern situation of the electrolysis, technological barriers and development opportunities.

In the second part of the chapter, direct reduction of aluminum-based compounds is discussed. This electrothermal process is cost effective, robust and widely used for aluminum alloys and master alloys with different elements (silicon, iron, manganese, calcium, chromium, zirconium, etc.).

Finally, electrode requirements, manufacturing and application are shown with respect to electrolysis (anodes in the Hall–Héroult process), electrometallurgy (self-baked and pre-baked) as well as for novel techniques (inert anodes). Energy issues of aluminum industry are also considered in this respect.

These three areas themselves are very wide and detailed description of the major topics in any of them would require its own book (or books) to be written. More specific details about the process of anode mass preparation or aluminum

oxycarbides formation could be found in numerous articles and technical papers to which the reader is encouraged to refer for more details (only a few of them are listed in the references).

The objective of this chapter is to present a general outline of the basic issues of the aluminum reduction, smelting, and electrolysis, alloy formation as well as very important issues about the electrodes used in these processes from the engineering point of view.

2 PHYSICAL CHEMISTRY OF ALUMINUM REDUCTION

Aluminum exists in nature in the various mineralogical forms of alumina, Al_2O_3, either "pure" (bauxites, corundum) or bonded with silica, alkalis and other elements in alumina-silicates. Dissociation of alumina at high temperatures produces sub-oxides like AlO, Al_2O_2, and Al_2O in the gas phase, although there are data about the existence of AlO and possibly "Al_2O" in the solid state as well [1]. Figure 1 shows that all major aluminum oxides are rather stable at high temperatures (negative free energy of formation). This means that the reduction of aluminum from its oxides is difficult to realize.

It is not also possible to reduce aluminum (Al^{3+}) to its metallic state in an aqueous solution. Figure 2 demonstrates that at any reasonable pH (0–14) at room temperature, either hydrated alumina ($Al_2O_3 \cdot H_2O$) or aluminate-ion like $Al_3(OH)_4^{5+}$ will exist. At very high reducing potentials, aluminum may be theoretically reduced
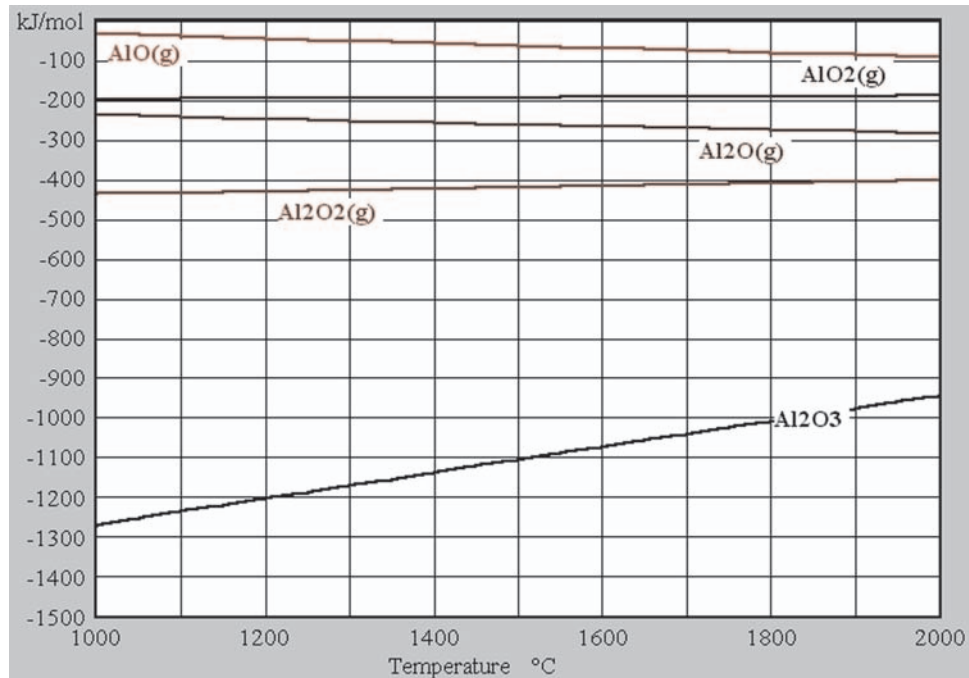
Figure 1 Ellingham-type diagram of aluminum oxides (Gibbs energy of formation, kJ/mol vs. temperature).
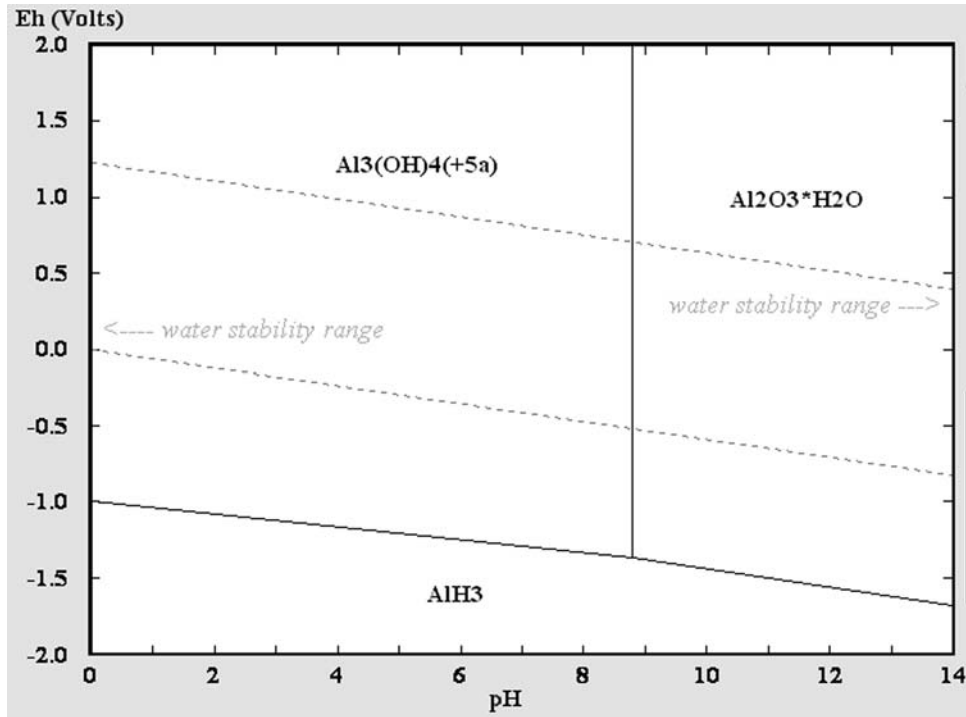
Figure 2 Electrochemical phase stability areas vs. potential (V) and pH. Dashed lines are water stability ranges. Only stable compounds and ions are shown for 25°C and for molarity of 1 mol Al per 1 H_2O.

to hydride AlH_3 but is already out of range of stability of water, so the aqueous media electrolysis cannot take place here (hydrogen generation due to water electrolysis will overcome aluminum reduction).

Alumina is known to form many forms, phases, and hydrates. The preparation of alumina from bauxites requires, thus, knowledge of reactions and processes associated with the decomposition of alumina hydrates. Holm and Lovi [2] have studied dehydration processes of European and South African bauxites using DSC together with thermosonimetry (TS). This method is based on the registration of sound signals generated by the transformations in solids [3,4]. Bauxites from South Africa contain mainly gibbsite $Al(OH)_3$, but European bauxites contain mostly boehmite γ-AlO(OH) or diaspore α-AlO(OH). It is known that dehydration of gibbsite in bauxites is likely to take place in two steps:

$$Al_2O_3 \cdot 3H_2O \rightarrow Al_2O_3 \cdot H_2O + 2H_2O \quad (at \sim 320°C)$$

$$Al_2O_3 \cdot H_2O \rightarrow Al_2O_3 + H_2O \quad (at \sim 550°C)$$
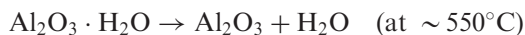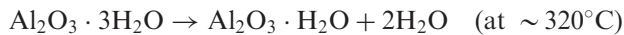
as well as directly on reaction:

$$Al_2O_3 \cdot 3H_2O \rightarrow Al_2O_3 + 3H_2O \quad (at\ 240-300°C)$$
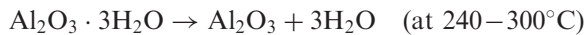
Dehydration of boehmite and diaspore takes place as a rule by a single reaction:

$$Al_2O_3 \cdot H_2O \rightarrow Al_2O_3 + H_2O \quad (\text{at } 520-560°C)$$
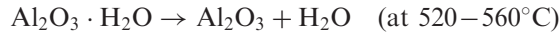
Using thermosonimetry (TS), it has been possible to determine the peculiarities of the dehydration, which cannot be seen by other methods. As an example, it was shown [2,5] that the first above-mentioned reaction has been fixed by DSC as a single-stage process, but using TS gives two distinct areas of reactions at $310-320°C$ and $322-325°C$. It has been explained by applying a two-stage mechanism as:

$$Al_2O_3 \cdot 3H_2O \rightarrow Al_2O_3 \cdot 2H_2O + H_2O$$

$$Al_2O_3 \cdot 2H_2O \rightarrow Al_2O_3 \cdot H_2O + H_2O$$

Iron hydroxides that are also present in bauxites, dehydrate at relatively low temperatures. Goethite (FeOOH) transforms into hematite (Fe_2O_3), with an endothermic effect resulting at $340-350°C$. Hydrohematite ($Fe_2O_3 \cdot H_2O$) decomposes at $50-150°C$. Hydrogoethite ($FeOOH \cdot H_2O$) has two endothermic stages: at $50-150°C$ and $340-350°C$.

Dehydration of bauxite from Guinea (West Africa) has been studied by DTA [5]. It has been shown that nearly 70% of constitutional water removes from bauxite at $280-330°C$, nearly 25% water removes at $530-570°C$ and full dehydration takes place only at $800-900°C$.

Many researchers have reported that the temperatures of phase transformations in Al_2O_3–SiO_2–H_2O system are strongly dependent on heating rate and water pressure [6]. It is being considered that gibbsite transforms into boehmite (at ~ 453 K) or into χ-alumina ($523-573$ K), and then into κ-Al_2O_3 at $773-1073$ K (Fig. 3). Bayerite α-$Al(OH)_3$ also transforms into boehmite at ~ 453 K, but also may decompose into γ- and η-aluminas at $503-533$ K. γ-Al_2O_3 transforms in its turn into
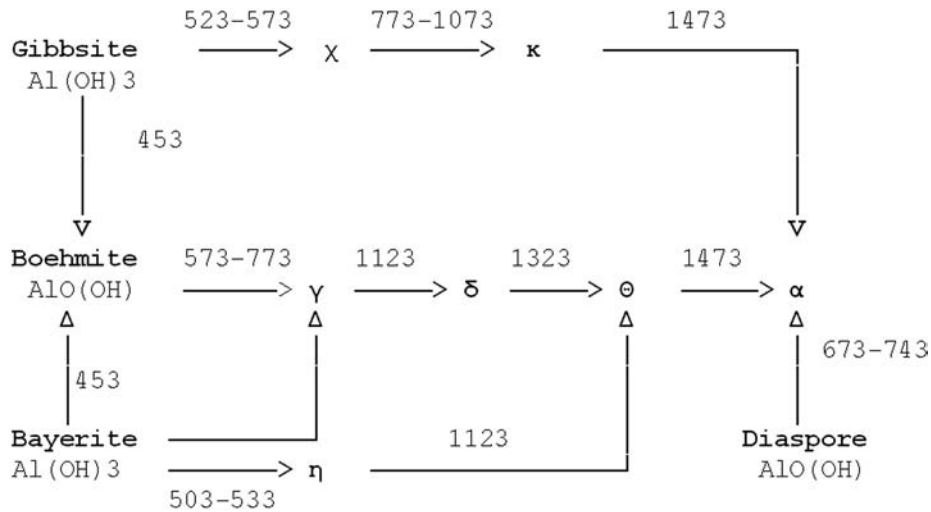
Figure 3 Temperatures (K) of phase transformations of aluminas and alumina hydrates (gathered from data of different sources).

the δ-phase (>1123 K), and then into Θ-Al_2O_3 ($\sim$1323 K). Alternatively, η-alumina also transforms into Θ-phase directly at $\sim$1123 K. All these phases finally give alpha-phase, but at different temperatures: $\sim$1473 K (Θ-, κ-phases), and 663–813 K (diaspore). Mizuno and Saito [6] have shown by TG and DTA that thermal decomposition of aluminum nitrate and sulfate mixtures with fumed silica gives different products at various temperatures of heat treatment.

Gasik and Sale [7] have used simultaneous thermal analysis (TG, DTA, and DTG) to determine the thermal decomposition of fine alumina hydrate made of aluminum nitrate precursor. Figure 4 represents the results of STA run in air at a heating rate of 10 K/min. It is seen that water removes out of the specimen in two steps, which have their DTG-peaks at 393–423 K and 693–723 K, respectively. The results were compared with the ones obtained by other studies [6,8–10], and these processes were considered to be diffusion controlled. The first of them represents a transformation of $Al(OH)_3$ to $AlO(OH)$, and the second stage is possibly related with the formation of γ-Al_2O_3 from $AlO(OH)$. The formal kinetic analysis for both the stages shows that these reactions are likely to be described by Avrami–Erofeev's equation [11,12]. This is in an agreement with the hypotheses about nucleation, growth of nuclei, and retardation due to overlap of growing nuclei. The formation of different products (aluminas and alumina hydrates) depends strongly on water partial pressure. The authors [13] have reported about such gibbsite transformation, which may be treated as a continuous sequence of structures − from boehmite to α-Al_2O_3. It can be described by the formula, $Al_2\square O_{3-v/2}(OH)_v\blacktriangle_{1-v/2}$, where $\square$ means cationic and $\blacktriangle$ means anionic vacancies, respectively. They have found that the scheme
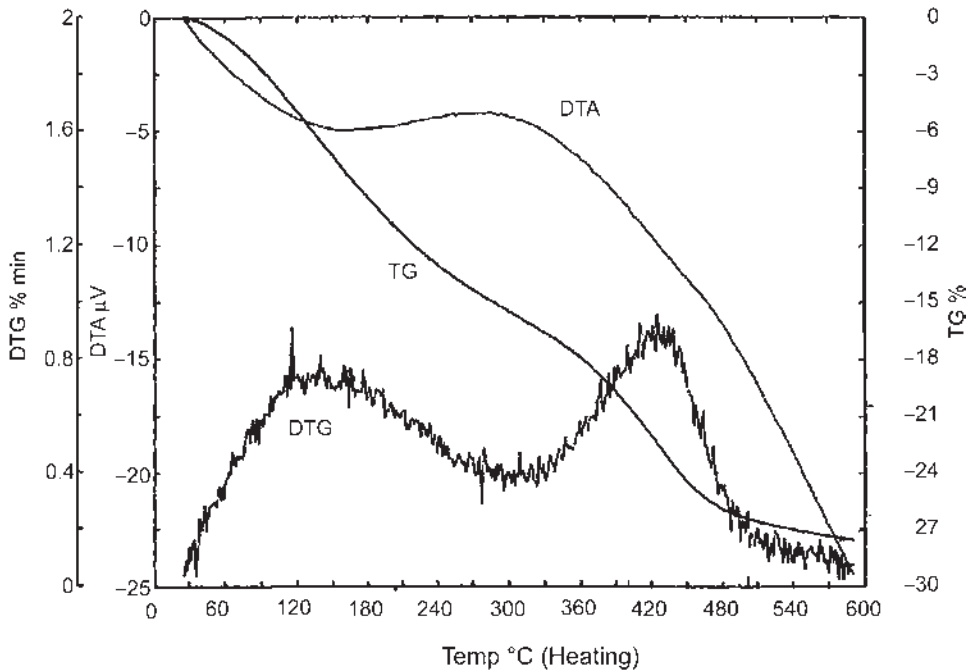
Figure 4   Simultaneous thermal analysis (STA) of alumina hydrate (after Gasik and Sale [7]).

as in Fig. 3 depends on the initial crystallinity of the gibbsite specimen. With TG/DTA and following x-ray analysis, it was shown that high-amorphous gibbsite dehydrates into χ, κ-, and α-aluminas, eliminating the stage of boehmite. However, in the core of a well-crystallized sample, boehmite forms at 573 K and further transforms to γ-, Θ-, and finally, to α-Al$_2$O$_3$ [13]. It was also observed that the temperature range and temperature of the DTA peak is proportional to the degree of crystallinity.

The above data of the alumina transformations are very important to understand not only for the bauxite processing, but for the aluminum smelting reduction as well. Alumina processed by different methods may have different phases and, thus, reaction ability. This could significantly affect the reduction reaction kinetics [5,11]. No full thermodynamic data are still known for all alumina phases. Using the data available, some phase stability diagrams may be, however, constructed (Fig. 5). This picture shows phase stability areas in the Al–O system heated up in argon (normal pressure 1 bar). Here, one may recognize that at $\sim 1650°$C, oxygen partial pressure under alumina will be $\sim 10^{-20}$ bar.

The reduction of alumina directly to pure aluminum is actually only possible by electrolysis (like the Hall–Héroult process). Using carbon leads to the formation of carbide and oxycarbides, which affect thermodynamics and kinetics of the process, as shown below. These compounds do not form in the presence of other elements like silicon and iron.
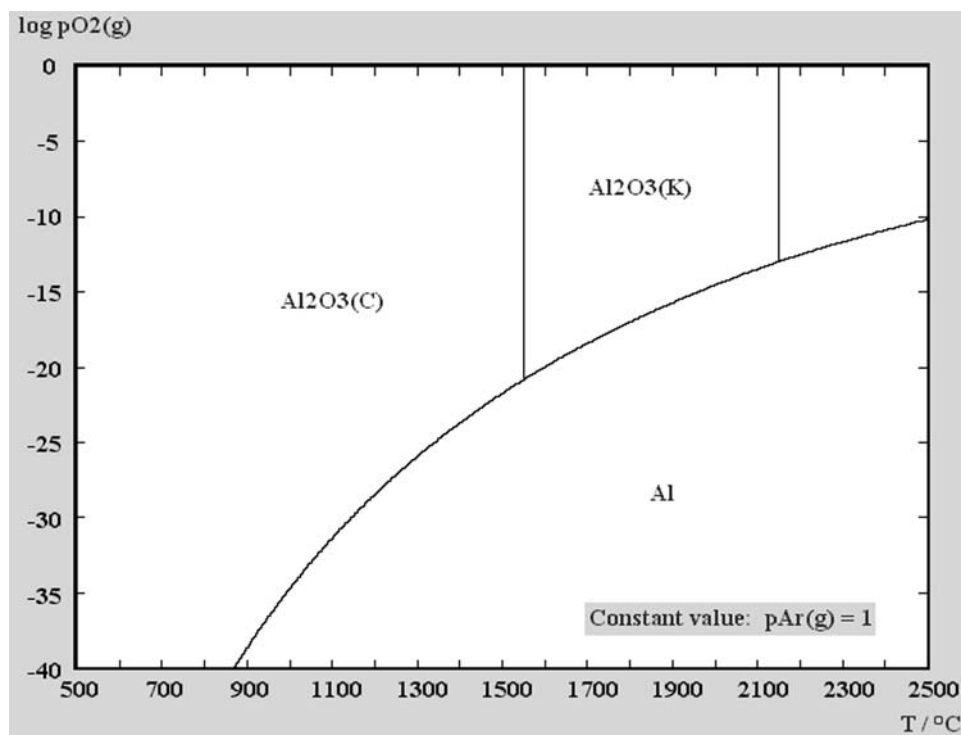
Figure 5 Phase stability diagram for the Al–O system in argon (1 bar).

It is known that during co-reduction of several elements, their chemical potentials and, therefore, activities may be significantly lowered and, thus, thermodynamic conditions for the reduction may be improved.

3 THE HALL–HÉROULT PROCESS

3.1 The Process Basics

The smelting process that is still used today was discovered almost simultaneously but independently in the United States and France. Working in Ohio, Charles Martin Hall made the same discovery that metallurgist Paul Lois Toussaint Héroult made in a laboratory in Gentilly: both men dissolved alumina in molten cryolite and then extracted the aluminum by electrolysis. In the Hall–Héroult process, alumina (Al_2O_3) is dissolved in an electrolyte that consists of molten fluoride salts kept at about 950–970°C. The alumina itself is separated from bauxite, a natural ore, by the Bayer process. The process takes place in electrolytic cells (or "pots"), where carbon cathodes form the bottom of the pot and act as the negative electrode. Anodes (positive electrodes) are held at the top of the pot and are consumed during the process when they react with the oxygen coming from the alumina. When direct current is passed through this melt, the alumina is decomposed into molten aluminum, deposited at the cathode, and oxygen, which reacts with the carbon anode to form CO_2.

The cells today are of two types: those with pre-baked anodes and those with baked-in-place anodes (Söderberg). All the pot-lines built since the early 1970s use the pre-baked anodes, where the anodes, manufactured from a mixture of petroleum coke and coal tar pitch (acting as a binder), are "pre-baked" in separate electrode plants (see Sec. 6 for more details about electrode production). In the Söderberg technology, the carbonaceous mixture is fed directly into the top part of the pot, where "self-baking" anodes are produced using the heat released by the electrolytic process.

The electrolyte used is based on cryolite (Na_3AlF_6), which is the best solvent for alumina. To improve the performance of the cells, various other compounds are added including aluminum fluoride (AlF_3) and calcium fluoride (CaF_2), to lower the electrolyte's freezing point. The electrolyte ensures that a physical separation is maintained between the liquid aluminum (at the cathode) and the carbon dioxide/carbon monoxide (at the anode).

Although the mechanism of electrolysis is still imperfectly understood, most investigators agree that cryolite ionizes to form hexafluoroaluminate (AlF_6^{3-}), which dissociates to form tetrafluoroaluminate (AlF_4^-), sodium (Na^+), and fluoride (F^-) ions:

$$Na_3AlF_6 \rightarrow 3Na^+ + AlF_6^{3-}$$

$$AlF_6^{3-} \rightarrow AlF_4^- + 2F^-$$
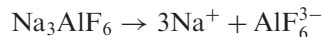
Alumina dissolves by forming oxyfluoride ions $Al_2OF_{2n}^{4-2n}$, at low concentrations and oxyfluoride ions $AlOF_n^{1-n}$, at higher alumina concentrations:

$$Al_2O_3 + 4AlF_6^{3-} \rightarrow 3Al_2OF_6^{2-} + 6F^-$$

$$Al_2O_3 + AlF_6^{3-} \rightarrow 3AlOF_2^-$$
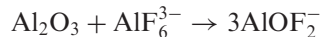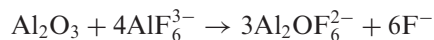
Cells are generally operated with 2–6% Al_2O_3 in the electrolyte. Measurements indicate that Na^+ ions carry most of the current and, most probably, a charge transfer occurs at the cathode interface. Here, hexafluoroaluminate ions are discharged, forming aluminum and F^- ions to neutralize the charge of the current carrying Na^+ ions:

$$12Na^+ + 4AlF_6^{3-} + 12e \rightarrow 12(Na^+ + F^-) + 4Al + 12F^-$$
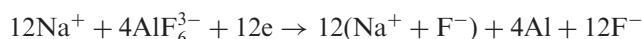
Oxyfluoride ions discharge on the anode, forming carbon dioxide and AlF_6^{3-} ions:

$$6AlOF_2^- + 3C + 24F^- \rightarrow 3CO_2 + 6AlF_6^{3-} + 12e$$
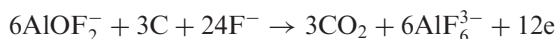
In operation, cryolite freezes on the sidewalls of the cells forming a "ledge" which protects the sidelining from severe attack by aluminum and molten cryolite. It also freezes over the top of the bath and forms a "crust" to support a top layer of alumina thermal insulation. Alumina is fed to the bath through holes punched in the crust. The carbon dioxide exits through holes in the crust and is collected under the hoods. The carbon dioxide and air leaking in is now ducted to dry scrubbers which remove fluorides from the gas stream. Fresh alumina contacting the gases removes the hydrogen fluoride and the evaporated fluoride particulate. This alumina, fed to the cells, returns the fluoride to the cells. The hydrogen fluoride comes from the residual hydrocarbons in the anodes and trace water in the alumina and air humidity reacting with the fluoride bath.

Control of alumina concentration in the cells is accomplished by a slight underfeeding. When the alumina reaches a critical level, the cell goes on anode effect, caused by a limiting rate of diffusion of alumina to the anode surfaces. The cell voltage then rises and some fluorocarbons are generated. The ratio of sodium fluoride to aluminum fluoride in the cryolite bath changes over time and corrective additions are added based on laboratory analyses.

Molten aluminum is deposited at the bottom of the pot and is siphoned off periodically, taken to a holding furnace, often but not always blended to an alloy specification, cleaned and then generally cast. A typical aluminum smelter consists of around 300 pots. These will produce some 125,000 t Al annually. However, some of the latest generation of smelters are in the 350,000–400,000 t range. The smelting process is continuous. A smelter cannot easily be stopped and restarted. If production is interrupted by a power supply failure of more than four hours, the metal in the pots will solidify, often requiring an expensive rebuilding process.

A modern alumina smelting cell consists of a rectangular steel shell typically 9–12 m long, 3–4 m wide, and 1–1.2 m high. It is lined with refractory insulation that surrounds an inner lining of baked carbon. Thermal insulation is adjusted to provide sufficient heat loss to freeze a protective coating of electrolyte on the inner walls but not on the bottom, which must remain substantially bare for electrical contact to the molten aluminum cathode. Steel (collector) bars are joined to the carbon cathode at the bottom to conduct electrical current from the cell.

Although the principles of electrolysis process have remained unchanged for more than 100 years, the performance of the electrolytic cells has been vastly improved. Today's best cells operate with energy efficiencies about 50%, which is
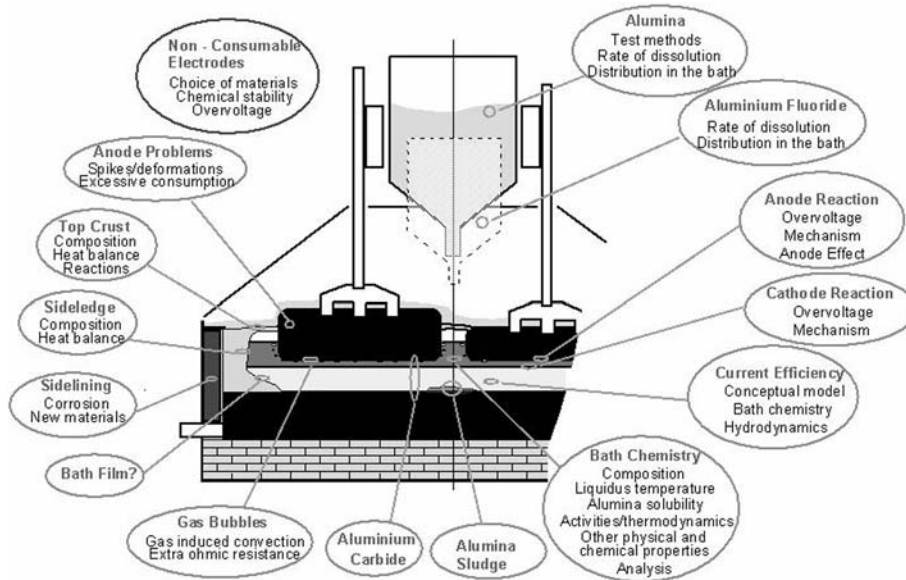
Figure 6 The schematic of the electrolysis cell and the main research directions in the process improvement (based on the data from SINTEF, Trondheim, Norway).

above average for electrolytic processes in molten salts. One of the keys to further improvements in energy efficiency and productivity is a better understanding of the chemical and physical conditions prevailing in the cells. In Fig. 6, there is a schematic of the process and the main research activities for the process improvement (SINTEF, Norway).

3.2 Energy and Environment Issues of the Electrolysis

The energy consumption for aluminum reduction is estimated to be 15–16 kWh/kg of aluminum at current levels of 60–70 kA [14,15] without taking into account the energy necessary for alumina fabrication from bauxites. The lowest energy consumption that can be achieved today is about 13 kWh/kg of aluminum for a line of modern, high-amperage (up to 325 kA) reduction cells [15,16]. It was recently estimated that at an average annual production of 20 million tons of aluminum in ~ 200 plants worldwide, the average energy consumption was about 14.7 kWh/kg [17]. Despite a substantial decrease (~ 2.5 times) in energy consumption for electrolysis over the last 100 years, there is still enough to do — to reduce alumina electrochemically, only 6.365 kWh/kg is needed [16].

The theoretical reversible potential for the reduction of alumina by carbon is 1.22 V and various additional over voltages including polarization, raise this value up to 1.76 V [18]. If other energy sinks are added (heating up the electrolyte, the anode, etc.), then the voltage would be around 2.02 V. Additional losses occupy another ~ 2.18 V: electrolyte drop 1.35 V, anode and cathode voltage drops 0.35 V, bubble-induced voltage drop 0.25 V external voltage drop ~ 0.15 V [18]. Production cells normally have current (electrical) efficiencies ranging from 85% to 95%. Large, modern reduction cells operate with current efficiencies from 94% to 96% [15].

Some part of the anode voltage drop compensates by carbon consumption for alumina reduction reaction, making carbon use efficiency about 80% [18].

Major energy savings are achieved through the recycling of scrap aluminum, which requires only about 5% of the energy required to produce primary aluminum (see chapter 4). On the other hand, secondary (recycled) aluminum cannot sometimes be used due to its high level of contaminants. Further increase of current over 350–400 kA does not seem to be a realistic approach either, so alternative solutions for both electrolysis process and other reduction method are being intensively studied worldwide [15,16].

Emissions from aluminum reduction processes are primarily gaseous hydrogen fluoride (HF) and particulate fluorides, alumina, carbon monoxide, carbon dioxide (CO_2), volatile organics, and sulfur dioxide (SO_2) from the reduction cells. The source of fluoride emissions from reduction cells is the fluoride electrolyte. The dissociation of the molten cryolite is the source of the perfluorinated carbons (PFCs) — tetrafluoromethane (CF_4) and hexafluoroethane (C_2F_6), which are formed during anode effects. The factors related to the formation of PFCs are not currently well understood, but they can be formed either by the direct reaction of the fluorine with the carbon anode or electrochemically [18,30].

In the past, the main environmental concern of the industry was emissions of perfluorocarbons from the reduction cells. The industry has made significant achievements in reducing these emissions, and the focus of concern has shifted to carbon dioxide (CO_2) emissions. It is the main component of the anode gas during the electrolysis process. It is also generated during the production of the electricity used in electrolysis, and is, thus, strongly dependent on the source of electrical power.

The total process air emissions (excluding power plant combustion emissions) for smelting and ancillary processes for representative industry processes are estimated to be 1.56 kg/kg of aluminum, of which about 90% (1.4 kg) is CO [15]. Total CO releases from the processes of bauxite mining through the production of molten aluminum have been estimated at 11.7 kg/kg of aluminum [15,18]. Combustion emissions account for approximately 70% of this total.

A variety of control devices has been used to abate the emissions from reduction cells and anode baking furnaces. To control gaseous and particulate fluorides and particulate emissions, one or more types of wet scrubbers (spray tower and chambers, quench towers, floating beds, packed beds, Venturi tubes) have been applied to all types of reduction cells and to anode baking furnaces. In addition, particulate control methods such as wet and dry electrostatic precipitators (ESPs), multiple cyclones, and dry alumina scrubbers (fluid bed, injected, and coated filter types) are used now on all the cell types and with the anode baking furnaces.

3.3 Developments and Opportunities in the Electrolysis Process

The general targets of aluminum electrolysis have been outlined by the Aluminum Association as follows [15]:

- Improve the performance of the Hall–Héroult cell to:
  - achieve an average cell efficiency of 97% on an annual basis;
  - reduce the energy intensity to 13 kWh/kg Al (near to mid term) and to 11 kWh/kg Al (long term);

- reduce the capital cost of aluminum production to \$2500 per annual metric ton of capacity;
- cost effectively minimize the generation of perfluorocarbons (PFCs).
- Reduce/eliminate CO emissions during smelting.
- Reduce the cost of aluminum reduction by 25% using alternative technologies.
- Develop new uses for wastes and byproducts from aluminum processes.

As one may see, besides traditional ways of the electrolysis process and cell construction improvement, new technologies and material solutions are also being sought. The achievement of these targets greatly depends on the ability to overcome existing technology barriers, Table 1 [15].

One of the key factors in achieving a Hall–Héroult cell average energy use of 13 kWh/kg (identified as a mid-term goal) is to optimize materials for the management of the internal control of the cell. Even if the energy requirements of the cell are lowered by reducing the anode–cathode space (or by other means), some method of conserving the energy in the cell is still needed. However, cell insulation is degraded by the corrosion of the sidewall and of the bottom which creates a material-related R&D need. Related needs are a better understanding of the best design, the development of low-cost materials, and the development of wetted cathodes [15,18].

The analysis of the cell bath [17] shows that excess energy is not only related to natural heat losses, but also to the higher electrode distance (bath depth) and ohmic losses at cathode side (between carbon blocks and steel current collectors). The distribution of the current in cathode is not uniform; in particular, for cathodes with side current leads: $\sim 80\%$ of the whole current is released through 1/3 of the cathode block, 15% at the next 1/3 and only 5% at the last third part [17]. The higher is the current, the larger is the gradient of the current and, thus, magnetic–hydrodynamic effect (horizontal current density up to 3 A/cm^2), leading to a curvature of the molten aluminum surface. To avoid shortcuts, the electrode distance should be increased — the higher the current is, the larger is the curvature and, thus, the ohmic losses (proportional to the square of the current).

Table 1 Technology Barriers in the Aluminum Production (The Aluminum Association)

Environmental	• Hydrocarbons capture ineffective
Technological	• Inability to raise thermal efficiency of reduction
	• Lack of robust understanding of bath chemistry
	• Too high voltage losses
	• Lack of basic material knowledge on dimensionally stable anodes
	• Need for improved cathode/wetted cathode research
	• Lack of new and feasible cell design concepts
	• Low efficiency of petroleum coke calcination
Economical	• High costs of equipment
	• Older cell improvement and reconstruction are not cost effective
	• No economical methods to remove impurities

Despite the current non-uniformity problems are known and respective solutions have been patented over 30 years ago, the costs of the reconstruction of existing electrolyzers were considered too high to be accepted. One of the reasons for high losses (15–16% of total losses) is poor electrical contact between steel leads and carbon blocks in cathode. Traditionally, this contact is made by the casting of pig iron between the steel and carbon [17,19]. Pig iron melt wets perfectly carbon steel but does not wet carbon surface, resulting in mostly mechanical contact only. Such contacts have very weak strength, low resistance to corrosion and destruction and high electrical resistance (real contact surface is only 5–10% of their visible or geometric surface). In metal/metal contacts, such a surface may achieve 30% of total possible contact surface.

To solve the problem, it was suggested to make special contact jams (cubic blocks) of 100–200 mm, where a hole of 30 mm diameter and of the same depth is drilled first [19]. In this hole, a special alloy is arc welded. The alloy wets perfectly porous carbon block and, thus, creates contact surface higher than 100% due to the porosity of graphite material (25–28%). The alloy infiltrates the block up to 15 mm depth, decreasing the contact resistance to a few micro-ohms. The tests conducted with industrial electrolyzer have demonstrated that, at working temperatures, the resistance of new contacts is 40–50% lower than traditional ones (pig iron), and stays nearly constant with time (normally, the contact resistance for pig iron joints increases up to 3% annually). New construction of the contacts has allowed decreasing energy consumption of 0.2 kWh/kg of aluminum [17].

It was also measured that new contacts have given 3–5% higher output of liquid aluminum and have decreased heat losses of $\sim 3\%$. Calculation shows that energy savings of 0.6–0.7 kWh/kg of aluminum are possible with new current lead design [17].

Shortly, the following issues have been identified for primary aluminum production [15,18]:

- conduct R&D of anode and cathode technologies (materials, their processing, experimental design for material selection, wetted cathode, inert anodes);
- R&D in alternative cell concepts (and processes);
- alternative carbon sources for new anode materials (higher ash and sulfur);
- improved control of electrical regime of the cell;
- robust bath chemistry;
- low-resistance external conductors (leads);
- higher cell efficiency with power modulations;
- new refractories and linings for the cell.

The Aluminum Association has stated that one of the most critical needs of primary aluminum production is for a variety of near-, mid-, and long-term R&D activities on anode and cathode technologies [15]. These activities should focus on the development of new materials (e.g. for inert anodes/cathodes), processing methods for these materials, experimental design for material selection, and wetted cathode technology. Using a systems approach to design dimensionally stable cells and optimize material use for the internal control of the cell could yield cell efficiency improvements within the next 10 years. In addition, the performance of signal analyses on the cell voltage in pot-lines could yield information that could be used to understand better and control the cell [15,18].

4 ALTERNATIVE PROCESSING OF ALUMINUM

Another of the most critical needs is for long-term R&D on alternative reduction and refining processes. This includes entirely new ideas for producing aluminum — e.g. direct reduction type processes or processes that totally eliminate electrolytic reduction [15]. A coordinated effort covering both refining and reduction is recommended to optimize the cost/benefit of using current or alternative raw materials. Cost-benefit analyses of kaolin and other alternative sources of alumina could contribute toward the ultimate goal of reducing production costs, as shown in Sec. 4.1.

Besides electrolytic process (reduction of alumina dissolved in cryolite), many alternative processes are being studied worldwide. Although some of the processes have been tested decades ago, there is still very little industrial implementation of these methods. The main obstacles for alternative process realization are the lack of combined units (integrating reduction and refining), lack of system approach, and for metal quality improvement, absence of new methods of low-temperature electrolysis of solid aluminum, existing needs to maximize the use of chemical energy, besides, only electrical [15].

The methods may be described as chemical (reduction by carbon or other agents) and electrolytic (electrolysis of compounds other than alumina). Some of these methods are discussed below (methods names are shown in the alphabetical order).

4.1 ALCAN Process

In the ALCAN process, bauxite and coke mixture are being heated in an electric arc furnace at temperatures above $2000°C$. Because pure aluminum cannot be thermodynamically obtained by carbothermal reduction, the product is in a form of aluminum-rich alloy with $\sim 30\%$ Fe, $\sim 10\%$ Si, $\sim 5\%$ C, and Ti (primary alloy). The alloy is being put into a reactor with aluminum trichloride ($AlCl_3$) vapor preheated to $1300°C$. The chloride vapor reacts with aluminum in primary alloy to produce aluminum monochloride, AlCl:

$$2[Al]_{alloy} + AlCl_{3\{gas\}} \rightarrow 3AlCl_{\{gas\}}$$
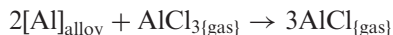
In the next reactor, aluminum monochloride vapor is contacted with aluminum droplet at $700°C$; when the above reaction is reversed, aluminum is extracted into a liquid melt and aluminum trichloride is recycled.

ALCAN process has claimed substantial capital and operational benefits in comparison with the Hall–Héroult process. The main reason, which limits the application of this process, is the very aggressive stress corrosion associated with the handling of gaseous $AlCl_3$.

4.2 Alcoa Process

In this process, alumina manufactured by the Bayer process, reacts with chlorine in the presence of carbon (coke) at $700–900°C$. The products of this reaction are gaseous $AlCl_3$, CO, and CO_2:

$$Al_2O_3 + 2C + 3Cl_2 \rightarrow 2AlCl_{3\{gas\}} + CO + CO_2$$
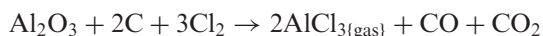
The gas mixture is directed into a condenser, where aluminum trichloride is separated at about $\sim 70°C$. Solid salt is added into a chloride bath, forming an

electrolyte solution of $\sim 5\%$ AlCl$_3$, $\sim 45\%$ LiCl, and $\sim 50\%$ NaCl. The electrolysis is carried out at $700°$C in a special double-sided (bipolar) electrode cell with 2.7 V and anode current density 0.8–2.3 A/cm^2. Chlorine is collected at the anode side and recycled into the beginning of the process (chlorination reaction). Liquid aluminum is collected at the bottom of the cell.

The advantages of the Alcoa process are higher conductivity of the electrolyte (~ 4 Sm/cm vs. ~ 2.5–3.0 Sm/cm in the Hall–Héroult process), and lower distance between the electrodes. Thus, energy consumption in the Alcoa process (8.9 kWh/kg Al) is significantly lower and efficiency is higher than in the Hall–Héroult process. The problems of Alcoa process are mainly related to chlorine handling, corrosion, and environmental issues (side formation of chlorated hydrocarbons, residual deposits, etc.).

4.3 Carbon Reduction Process

The use of carbon reduction to produce aluminum from alumina or directly from bauxites may look the most straightforward approach to be realized. As shown above, theoretical reduction reactions like:

$$Al_2O_3 + 3C \rightarrow 2Al + 3CO$$
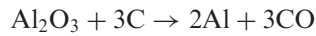
seem to be practically attractive. However, the reduction of alumina by carbon proceeds through the formation of gaseous sub-oxides (Al$_2$O, AlO, Al$_2$O$_2$), carbides (Al$_4$C$_3$) and oxycarbides (Al$_2$OC, Al$_4$O$_4$C) [14]. It was experimentally determined that alumina is being reduced by carbon in the following sequence:

$$Al_2O_3 \rightarrow Al_4O_4C \rightarrow Al_2OC \rightarrow Al_4C_3 \rightarrow Al$$

The carbide, Al$_4$C$_3$, is stable up to very high temperatures (2200–2500°C), higher than melting point of the alumina, so carbide formation in the furnace means "solidification" of the bath and the ceasing of the reaction by simple kinetic reasons.

Porada and Gasik [5] have presented the following sequence of carbothermal reduction of alumina with the stoichiometric mixture (Al$_2$O$_3$+3C):

$$Al_2O_3 + 3C \rightarrow \tfrac{2}{3}(AlO \cdot Al_2O_3) + \tfrac{1}{3}CO + \tfrac{8}{3}C \quad (\sim 1650°C)$$
$$\rightarrow \tfrac{1}{2}Al_4O_4C + CO + \tfrac{3}{2}C \quad (\sim 1800°C)$$
$$\rightarrow Al_2OC + 2CO \quad (\sim 1850°C)$$
$$\rightarrow 0.3Al_2O_{3liq} + 0.3Al_4C_3 \quad (\sim 2000°C)$$
$$\rightarrow 1.47Al_{(liq)} + 0.04Al_4C_3 + 0.11Al_{(gas)} + 0.13Al_2O_{(gas)} + 2.87CO$$
$$(\sim 2050°C)$$
$$\rightarrow 2Al_{(gas)} + 3CO \quad (\sim 2500°C)$$
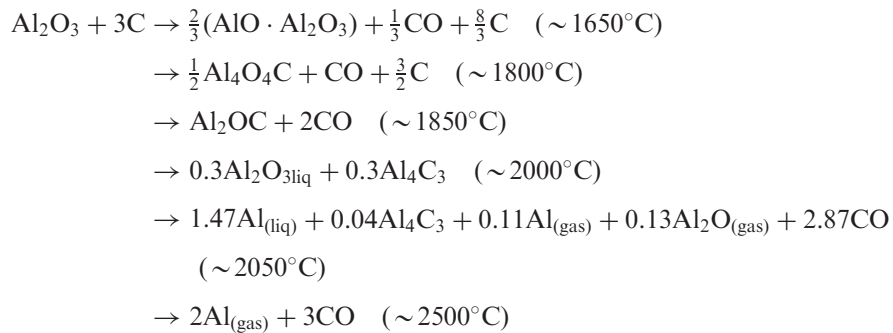
The quaternary equilibrium point, Al$_2$O$_3$–Al$_4$C$_3$–C–Al$_{(liq)}$, in the Al–O–C system is achieved at about 2380–2390°C [5]. Therefore, aluminum may be reduced by carbon from alumina, if necessary conditions for carbide and oxycarbides decomposition will be achieved. As shown below, this is practically possible in the case of low aluminum activity (aluminum solutions with other elements).
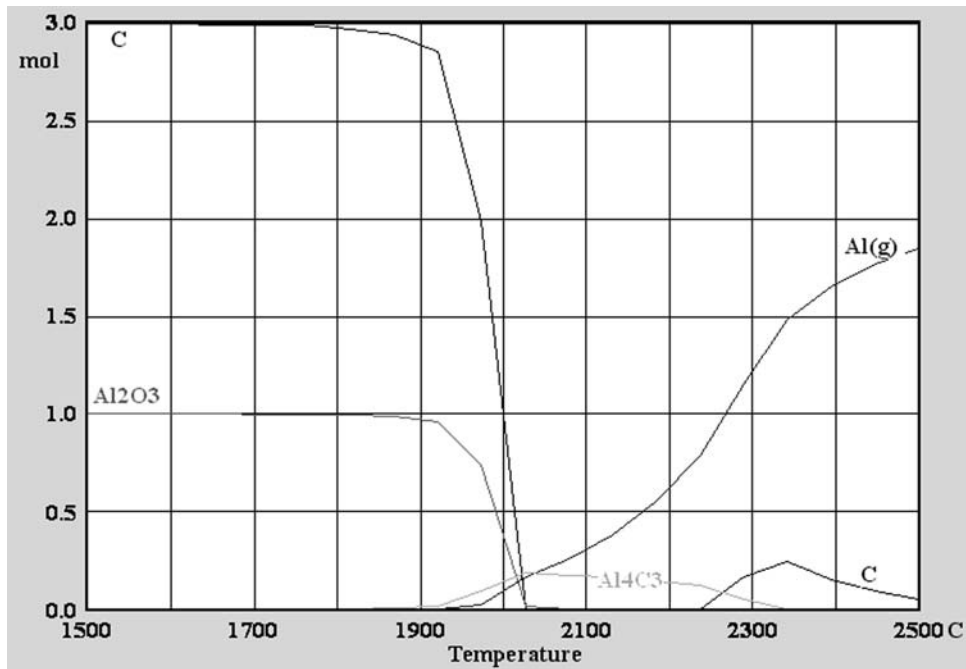
Figure 7 Thermodynamic equilibrium of the stoichiometric mixture (Al_2O_3+3C) reduction at 1 bar pressure in argon for different temperatures.

Theoretical calculations show that during the reduction of the stoichiometric mixture (Al_2O_3+3C) at 1 bar pressure in argon, some amount of gaseous aluminum may be produced (Fig. 7) but only when over 2050°C and only when accompanied by the carbide. If the pressure is lowered, the temperature may be lowered too. Figure 8 shows a similar picture for 1900°C, where a pressure of ~0.4 bar is needed for aluminum to reduce. However, the theoretical yield of aluminum may be reached only at very low pressures, yet accompanied by the carbide. Carbide formation makes the reduction process practically unsuitable to realize using "traditional" electrical or shaft furnaces. However, there are no theoretical limitations about alumina to be reduced by carbon in some kind of "novel", non-equilibrium processes (sub-micron particles with low diffusion paths, plasma, microwave or laser-assisted processes, etc.). This area is yet very little studied.

4.4 Elliot-Mitt Process

This process uses a carbothermal reduction of alumina when the resulting aluminum dissolves in molten tin to prevent the formation of carbide and oxycarbides. To assist the reaction, higher temperatures (~1800°C) and lower pressures (~0.1 atm) are desired. Due to the presence of tin, aluminum activity is low enough to make necessary thermodynamic prerequisites for its reduction.

However, the practical realization of the process has resulted in the alloy of ~9% Al only. Although separation of aluminum from tin is possible by several methods, the product cannot yet satisfy requirements for pure aluminum.
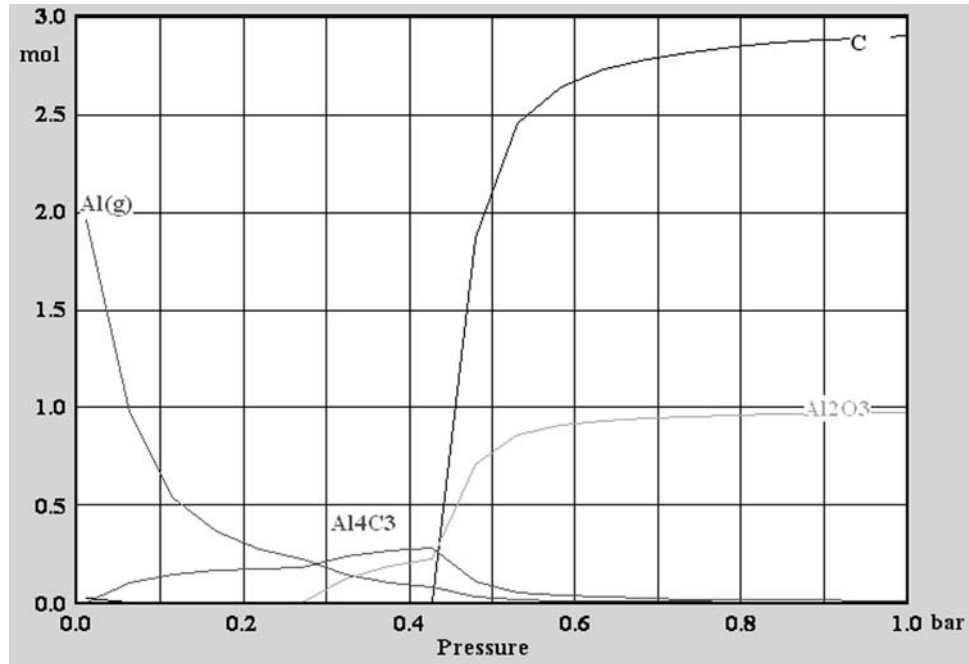
Figure 8 Thermodynamic equilibrium of the stoichiometric mixture (Al_2O_3+3C) reduction at 1900°C in argon for different pressures.

4.5 Monochloride Process

The monochloride process chemically unites the Alcoa and the ALCAN processes. Here, alumina or bauxite is directly being reduced by coke in the presence of $AlCl_3$ vapor (not with chlorine, like in Alcoa process). Thus, both reactions of these processes are taking place simultaneously:

$$Al_2O_3 + 3C + AlCl_{3\{gas\}} \rightarrow 3AlCl_{\{gas\}} + 3CO$$

This reaction, however, requires much higher temperatures (1750–1850°C) than those of Alcoa process. Additives (silicon, iron) present in bauxites also react in these conditions to form lower chlorides ($SiCl_2$, $FeCl_2$). The mixture of these chlorides is submitted into a reactor, where liquid lead is sprayed from the top to rinse the gases. Lower chlorides then dismutate to elements and higher chlorides:

$$3AlCl_{\{gas\}} \rightarrow 2Al + AlCl_3$$

$$2SiCl_{2\{gas\}} \rightarrow Si + SiCl_4$$
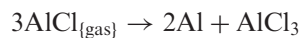
Lead-metal liquid drops are collected at the bottom of the reactor and aluminum is separated from lead by soaking the mixture at 700°C. Higher chlorides are recycled back to the process. The liquid aluminum obtained contains some silicon and lead.

This process has been successfully tested in laboratory conditions, but so far has not yet been realized in large-scale plants. It is expected that gross economic

benefits of this process may be significant in comparison with the chain "Bayer+ Hall–Héroult", if a cost-efficient plant can be constructed.

4.6 Reynolds Process

This process also uses carbon reduction but as a multi-stage procedure, similar to those processes that occur in a blast furnace. Alumina is being reduced by carbon (coke) first to oxycarbides (Al_4O_4C) and then to carbide Al_4C_3. Adding more alumina results in slag formation that reacts with carbide and oxycarbides to form aluminum metal. At these high temperatures, aluminum vaporizes and reacts with residual alumina to form sub-oxides:

$$4Al_{\{gas\}} + Al_2O_{3(liq)} \rightarrow 3Al_2O_{\{gas\}}$$
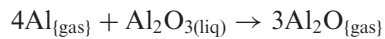
Sub-oxides may then dismutate to produce aluminum and slag at lower temperatures. After several following stages, metal melt is being cleaned from carbon by alumina-oxycarbide slag and, finally, by gas.

The capital costs of Reynolds process were reported to be about 2/3 of the Hall–Heroult plant, but its practical realization has faced some difficulties, mostly with the construction of the furnaces and problems with material stability at these operational conditions.

4.7 Toth Aluminum Process

The shortage of high-quality bauxite supply has initiated studies of the processes that may use alumina–silicate raw materials instead of pure alumina. Toth Aluminum process uses, in the beginning, the same alumina reaction with chlorine in the presence of carbon (coke) at 700–900°C like in the Alcoa process:

$$Al_2O_3 + 2C + 3Cl_2 \rightarrow 2AlCl_{3\{gas\}} + CO + CO_2$$
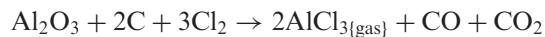
Reduction of liquid aluminum trichloride is carried out by manganese at 300°C and at 15 atm (200 psi) in a special pressure reactor to make aluminum powder:

$$3Mn + 8AlCl_{3(liq)} \rightarrow 3(MnCl_2 \cdot 2AlCl_3)_{(liq)} + 2Al$$

and liquid salt melt is separated into aluminum trichloride (recyclable into the process) and manganese dichloride. The latter is oxidized to Mn_2O_3, which is used to recover manganese metal by a carbothermal reduction. There are some variations in the Toth Aluminum process, but the main limitation is its complicity and costs, related to the recovery of manganese and the low yield of aluminum. Even stoichiometric reaction needs 2 mol of alumina to produce 1 mol of aluminum. Fabrication of metal manganese is not possible by a carbon reduction method, so only high-carbon ferromanganese (75% Mn, 7% C) can be received with the recovery of 65–70% of manganese only [14]. Ferromanganese itself is not a very good reducing agent in comparison with pure manganese. According to the existing literature data, this process is not recently cost effective to be realized, taking into account that manganese mixes well with aluminum and the resulting alloy has little applications.

4.8 Electrolysis of Other Aluminum Compounds

Among alumina, other aluminum compounds are being considered for electrolysis process. One possible candidate is aluminum nitride, AlN. If aluminum nitride will be available commercially in industrial quantities, it may be a good alternative for electrolysis because the reaction:

$$2AlN \rightarrow 2Al + N_2$$
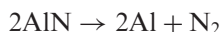
will need only 0.72 V at 1000°C. Nitrogen would prevent anode from oxidation and this process would be an energy-saving one. The problem, however, still exists in finding a suitable electrolyte or solvent which may dissolve substantial amounts of AlN at medium temperatures.

Another potential candidate is aluminum sulfide, Al_2S_3, which may be made by the reduction–sulfidation reaction:

$$2Al_2O_3 + 3S_{2\{gas\}} + 6C \rightarrow 2Al_2S_3 + 6CO$$
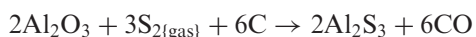
This reaction is thermodynamically possible at over 1200°C, but process temperature should be over 1400°C to keep aluminum sulfide in the liquid state. The electrolysis of aluminum sulfide requires about 1.2 V, but released sulfur may dissolve in the electrolyte and form aluminum sub-sulfides. In the experiments, the yield was found to be about 70%, with the energy consumption close to that in the electrolysis of $AlCl_3$. The process may be competitive to chloride electrolysis if an efficient method of sulfide production will be developed and sulfur utilization enhanced.

4.9 Reduction of Alumina Using Hydrogen

This case is unlikely to be used for aluminum production, but it has an interest with respect to the reaction of alumina with molecular or atom hydrogen and the dissolution of hydrogen in liquid aluminum alloys. In work [20], a brief thermo-dynamic assessment has been made to demonstrate the possibility of forming gaseous aluminum by the reaction of H_2 or H with alumina.

The phase stability areas in the Al–O–H system are shown in Fig. 9. If alumina is heated up at 1 bar hydrogen at ~ 1300°C, the reduction may proceed when water partial pressure is less than 10^{-8} (i.e. water vapor concentration is less than $\sim 10^{-5}$ mg/l H_2). Such humidity of hydrogen is very difficult to maintain at high temperatures continuously. If alumina is heated up to 1500°C under pressurized hydrogen (a few bars), than humidity of hydrogen may be substantially lower, Fig. 10.

The formation of metallic aluminum has been experimentally confirmed [20], where alumina was reduced by hydrogen dissolved in aluminum. The authors [20] have indeed recognized the difficulty of the reduction due to very low water traces to be continuously and efficiently removed. This method may be successfully applied but possibly in the non-equilibrium conditions (plasma, etc.), where rather favorable reduction may occur in limited volumes and proceed with higher rates [20,16].

Generally, for the above-mentioned alternative processes, the Aluminum Association has identified the following needs [15]:

- long-term R&D on alternative reduction and refining processes (closed loop, use of chemical energy);
- R&D for electrolysis of solid aluminum;
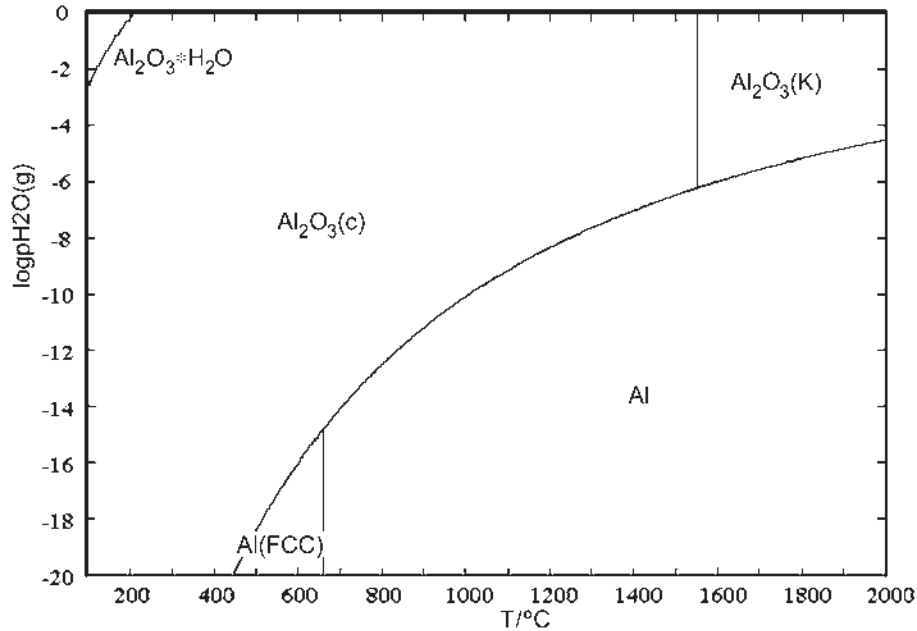
Figure 9 The Al–O–H phase stability diagram at 1 bar H_2.

- alternative raw materials (alumina sources like kaolin, kyanite);
- elimination of intermediate refining steps;
- development of durable equipment for the handling of molten salts and aggressive reagents.

Below is shown how some of these problems may be solved using aluminum alloy direct production instead of making pure aluminum first.

5 DIRECT REDUCTION OF ALUMINUM ALLOYS

A significant part of all aluminum alloy production is occupied by aluminum–silicon cast alloys (silumines), aluminum-based engineering alloys (cast and wrought – with magnesium, copper, zinc and other elements) as well as various master alloys (Al–Me). The majority of these alloys have been manufactured by mixing electrolytic aluminum with other elements. Taking into account the high costs of pure aluminum, silicon, magnesium, etc., the costs of these alloys were rather high.

5.1 Aluminum–Silicon Alloys

As shown in the next chapter, there is no practical or theoretical reason why pure elements have to be manufactured first (if they will be mixed together later at the next stage). Lack of high-quality bauxites for aluminum electrolysis has initiated more active research in the use of other raw materials (kaolin, kyanite, disten-sillimanite) for aluminum alloys. If these materials are used for alloy smelting instead of making pure aluminum first, there is no need for the separation of alumina either in the removal of other compounds [21,14]. In the 1960 s for the first time in
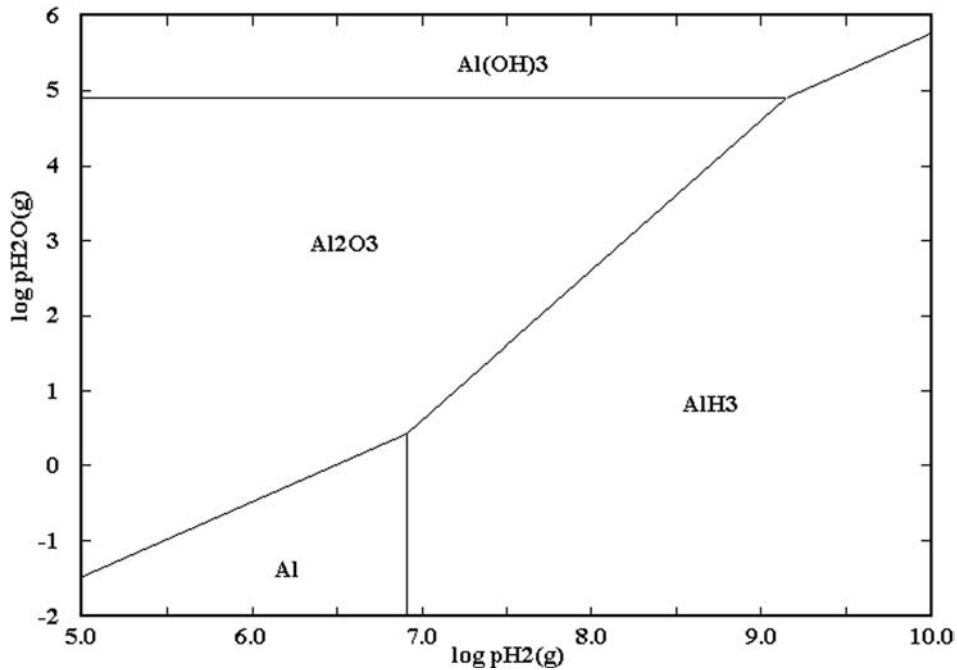
Figure 10 The Al–O–H phase stability diagram at 1500°C.

the world, such a plant has been built in Zaporizhzhya, Ukraine. Here, the new process of direct reduction of alumina–silica raw materials by coke into the so-called "primary alloy" (58–60% Al, 37–39% Si, 1.4–1.6% Fe, 0.5–0.6% Ti, 0.8–1.0% Ca) was commercially realized [21,22].

Here, the raw materials used were enriched kaolin (37–40% Al_2O_3, 46–47% SiO_2, 0.3–0.7% Fe_2O_3, 0.25–0.35% TiO_2, 0.15–0.50% CaO). This kaolin ($\sim$40%) was mixed with technical alumina (4–25%), gas-grade coal, oil coke (altogether 30–40%) and binder to form briquettes. Later, kaolin was substituted by enriched disten-sillimanite $Al_2O_3 \cdot SiO_2$ to reduce the kaolin share to 8–25% only [14,21,23]. This mineral concentrate is a waste product from titanium–zirconium ore dressing plant [24] and has significantly higher melting temperature with respect to kaolin. This ensures better reduction process. The smelting is performed in open-hearth ore reduction furnaces of 16.5 MVA power with self-baked electrodes of diameter 1200 mm.

The costs of the Al–12% Si alloy made by the direct reduction of primary alloy and then its mixing with aluminum are 20–25% lower than for the mixing of primary aluminum with silicon. The productivity of the process is higher by 35–40% and investment required is lower by 30–33% [14,24].

How does the direct reduction proceed? Reduction of alumina by carbon was considered above. Silica is usually reduced to SiC, unless special conditions [21] are created for the silicon formation (CO removal). Mullite $3Al_2O_3 \cdot 2SiO_2$ can be reduced by carbon:

$$3Al_2O_3 \cdot 2SiO_2 + 13C \rightarrow 6Al + 2Si + 13CO$$
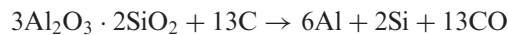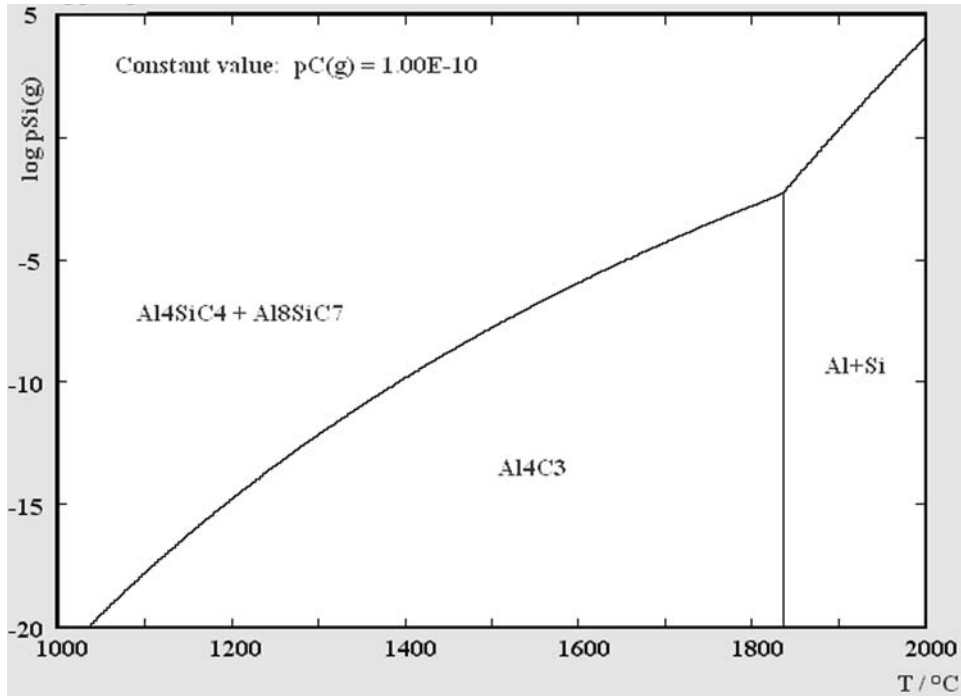
Figure 11 Phase stability areas in the Al–Si–C–O system at higher carbon potential $(\log(P_C) = -10)$.

but this reaction would be thermodynamically impossible until 2000°C, if the mutual dissolution of aluminum and silicon is not taken into account (i.e. if aluminum and silicon are considered as pure phases with unit activity). For the reaction balance, it can be seen that 3 mol of aluminum are formed per 1 mol of silicon — a solution with $\sim$70% Al would result. Even then, the direct reduction of mullite to aluminum–silicon melt would be only slightly favorable.

As soon as carbon is present with the Al–Si–O system, the reduction of alumina proceeds through oxycarbides and carbide [14]. As a result of the reduction, simple carbides (SiC, Al_4C_3) and complex carbides ($Al_4SiC_4 = Al_4C_3 \cdot SiC$, $Al_8SiC_7 = 2Al_4C_3 \cdot SiC$) are formed. These carbides are very refractory (melt over 2350°C) and they, with oxycarbides, retard the reduction process [24]. Depending on the silicon concentration, equilibrium carbide may appear in different composition: Al_4C_3 and Al_2OC (at low silicon), Al_8SiC_7 (at higher silicon), SiC (more than 30% Si). If after oxygen removal the carbon potential is still high (e.g. $\log(P_C) = -10$), the metal alloy may be only produced above 1850°C (Fig. 11). If carbon was already consumed (i.e. not much excess carbon was left, $\log(P_C) = -20$), liquid Al–Si alloy may be received already at 1200°C (Fig. 12).

The Al–Si system is known to have a non-ideal behavior. The activity coefficients of aluminum and silicon may be expressed as a function of their molar fraction [21]:

$$\log(\gamma_{Al}) = -0.4187X_{Al}^3 - 0.2486X_{Al}^2 + 1.535X_{Al} - 0.887$$
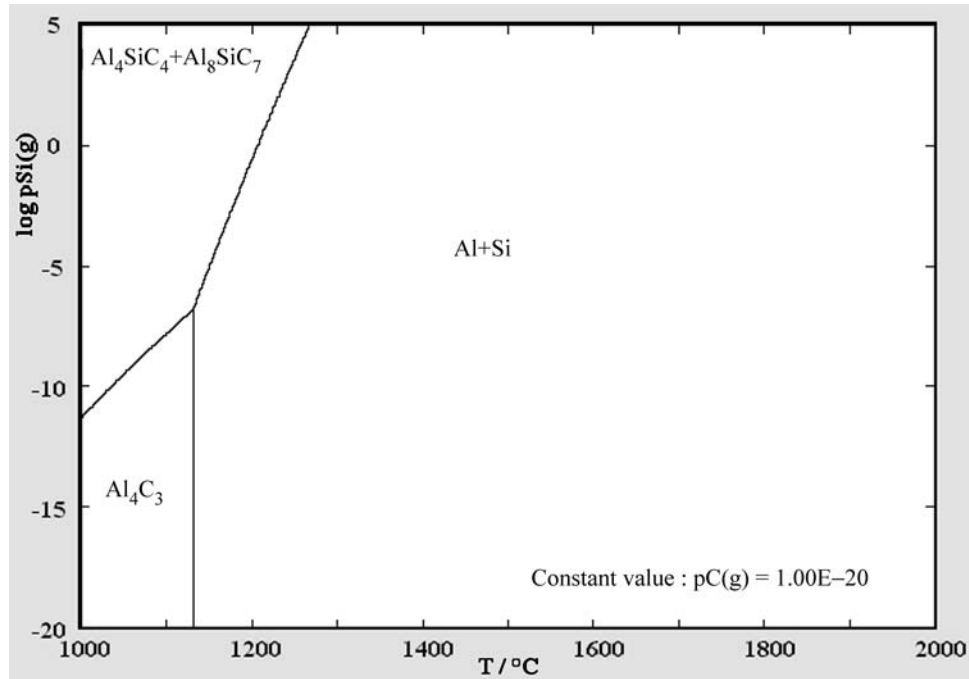
Figure 12 Phase stability areas in the Al–Si–C–O system at lower carbon potential $(\log(P_C) = -20)$.

$$\log(\gamma_{Si}) = -0.3531X_{Si}^3 - 0.2984X_{Si}^2 + 1.5352X_{Si} - 0.8881$$

where X_i is the molar fraction of the element i, γ_i is the activity coefficient in the activity definition as $a_i = \gamma_i X_i$ [14,21,25].

When the charge composition is optimized and electrical and temperature parameters are being held within the operation window [14,21,23,24], liquid Al–Si alloy can be made by the direct reduction of alumina–silica raw materials in the electrical furnace with self-baked electrodes. Besides the alloy, slag of variable composition is also formed (30–60% Al_2O_3, 35–50% SiC, <2% CaO, <20% SiO_2, <5% oxycarbides). To avoid contamination of the primary alloy with these inclusions, special fluxes are being added into the ladle before melt pouring from the furnace. Resulting unpurified alloy may contain up to 2–3% of these inclusions.

Why is direct reduction of primary alloy more advantageous? There are several important reasons for this [14]:

- higher specific power (kWh/m^3) of ore reduction furnaces vs. electrolytic cells,
- no DC is required, conventional AC from the grid is used (no AC/DC conversion losses),
- no special alumina production stage is needed (hydrometallurgical Bayer process is bypassed),
- different alumina and alumina–silica raw materials may be used, which cannot be used for alumina extraction,

- no cryolite and other fluorine salts are required,
- low-cost carbon reducing agents (coke, pitch) can be used,
- higher recovery of aluminum and silicon,
- lower impurities (non-ferrous metals) when used for steel killing, and
- low capital investment required.

The second stage of the commercial Al–Si alloys making is in the mixing of molten primary alloy ($\sim 35\%$ Si) with molten aluminum for the electrolytic process in the neighborhood [24] of about 750–800°C. Protecting and refining fluxes can be added to the metal mix and its temperature is maintained by built-in gas burners [21].

As a result of mixing, silicon content in the alloy is reduced to a desired value (7–22%) and liquidus temperature of the alloy is decreased. Some metallic impurities (Fe, Ti, Ca, etc.) react with aluminum and silicon forming intermetallides of higher melting temperature. These intermetallides must be removed from the alloy, because they usually weaken properties of silumine castings [24]. This is achieved by adding manganese (to reach the ratio of Mn/Fe ~ 1.5–2.0) and solid Al–Si alloy of similar composition (to decrease temperature to 595–600°C). Manganese reacts with Fe and Si and releases aluminum from intermetallides. The melt is then being vacuum filtered through silica sand, where all high-temperature intermetallides remain. These tails have nominal composition of 72–74% Al, 17–20% Si, 1.5–2.5% Fe, 1.2–1.5% Ti, 2–4% Mn and are recycled in a separate furnace (900°C, with flux of 60% $NaCl+Na_3AlF_6$).

During filtration, a part of the melt oxidizes and turns into slag ($\sim 80\%$ Al_2O_3, $\sim 2.5\%$ Al_4C_3, $\sim 3\%$ SiC, 2–3% SiO_2, $\sim 2\%$ CaO, and $\sim 10\%$ alkali halides). Slag may contain up to 30% of metal phase as inclusions and droplets [24]. All tails and slags of the process are being nowadays used as raw secondary materials in the adjacent metallurgical processes [14,24].

5.2 Aluminum-Based Master Alloys

The authors [21,26] have developed a technology for the complex processing of practically all kinds of alumina–silica raw materials (poor bauxites, kyanite, kaoline, disten-sillimanites, etc.) with high level of impurities (iron, titanium, etc.). It consists of several stages of smelting. The first stage includes smelting of raw material with some iron (steel) chips and coke. Here, alumina is not actually reduced, but silicon reduces to form nearly standard grade ferrosilicon (18–30% Si) alloy. Alumina remains in the slag which turns into "electrocorundum" — eventually pure alumina (96–98% Al_2O_3) with low silica, titania, calcia and iron oxides. This electrocorundum is a commercial product, which is used for the abrasive and refractory manufacturing [5,14], hydrochemical extraction of alumina [21] or for synthetic slag smelting (45% Al_2O_3, 53% CaO) used in steel making. Electrocorundum is also used instead of kaoline for aluminum alloys smelting, as shown above.

If quartzite and coke is added to the charge, in an electrical furnace, they do form primary aluminum alloy ($\sim 36\%$ Si). If ferroalloys are added (Fe, FeMn, FeCr, FeSiCr, FeSiMn), complex composition master alloys may be easily produced [26]. Alternatively, some alloying elements (Zr, Ti) may already be present in the raw alumina–silica materials as oxides and their reduction (up to 10% in the final alloy) can be realized during electrothermal reduction during primary alloy smelting.

In iron and steel making, a significant part of aluminum used for steel killing is a cheaper secondary metal (recycled). Because of recycling problems, it may contain 5–12% of non-ferrous metals ($< 4.5\%$ Cu, $< 2\%$ Zn, $< 0.4\%$ Pb, $< 0.3\%$ Sn), which are undesirable in steel. In any case, the use of primary or secondary aluminum in steel making has become problematic due to its low density and high affinity to oxygen. During the alloying of the molten steel bath, the aluminum use efficiency is 20–30% only and the resulting fine alumina particles contaminate steel ingots [26]. This is why the use of ferroaluminum (10–30% Al) with lower aluminum activity, higher density (6.9–7.1 g/cm^3), and higher recovery (50–60%) is much more advantageous.

Recently, many kinds of ferroalloys with aluminum (Al–Fe, Al–Fe–Mn–Si, Al–Si–Ca, Al–Cr–Fe, Al–Zr–Fe) are being produced with this method (Table 2). These alloys do require less energy for their smelting (3.9 kWh/kg for FeAl, 5.2 kWh/kg for FeAlSiMn), so, their application in steel making is more cost effective than pure aluminum:

	Primary Al	FeAl alloy
Aluminum content (%)	95–98	8–30
Energy required (kWh/kg)	13–16	3–4
Al utilization at steel killing (%)	20–30	50–90
Costs for use in steel making (relative units per kg Al)	55–70	10–40

It is well established that aluminum production for steel making and foundry is much more efficient by making a direct reduction in the form of ferroalloys of aluminum-rich primary alloys.

6 ELECTRODES FOR ALUMINUM PRODUCTION

6.1 Self-Backing Electrodes (Söderberg Type)

Self-baking anodes for aluminum production seem to be used for the first time in 1923 in Norway. At that time, it was found that the main problem was in the high potential losses ~ 1.5 V between contact plate and the anode [27,28]. New technical solutions have resulted in two principal kinds of current leads — side leads and top leads, the latter being the most used recently.

Self-baking anodes are usually composed of a metal shell (steel) and a carbon-based mass, which form a carbon block during baking [27]. The shell is usually welded of steel sheets of 2–4 mm in the form of a cylinder or an ellipse. The diameter of the shell is usually 400–2000 mm or 2800×650 or 3150×650 mm. Inside the shell are special ribs (fins) for better adhesion of the anodic mass to the shell. This type of construction is practically used in all kinds of electrometallurgy furnaces.

All kinds of self-baked anodes (anodic mass) are being fabricated using the following scheme [27,28]. Solid carbonaceous raw materials (coke, anthracite, etc.) are being crushed, heat treated and classified. Secondary raw materials (recyclable electrodes, graphite pieces, etc.) are also crushed and sieved. Binder (pitch, oils and other additions) is mixed and delivered for dosing with solid and secondary materials. Mixed charge (e.g. thermoanthracite −20% of 4–10 mm, 24% of 10–20 mm, 16% of crushed (< 4 mm), then 40% coke < 4 mm, and additionally (over 100%) 24% of the coal pitch as binder) is used to form briquettes of ~ 1.5 kg

Table 2 Typical Compositions of Some Commercial Aluminum-contained Master Alloys

Master alloy	Leading element(s) (%)	Other elements (%)
Ferroaluminum	8–30 Al	< 4 Si, < 4 C, < 0.06 P, < 0.06 S, balance Fe
Manganoaluminum	12–16 Al, 40–80 Mn	< 2.5 Si, < 1.5 C, < 0.03 S, < 0.3 P, balance Fe
Ferrosiliconaluminum	10–15 Al, 50–55 Si	Balance Fe and minor impurities
FAMS (Fe–Al–Mn–Si alloy)	10–14 Al, 22–26 Mn, 12–14 Si	< 1.5 Ti, < 0.7 Zr, < 2.5 C, < 0.08 P, balance Fe
FAMSC (Fe–Al–Mn–Si–Ca alloy)	4–10 Al, 17–22 Mn, 45–50 Si, 7–12 Ca	Balance Fe and minor impurities
Ferrochromealuminum	>20 Al, >50 Cr	< 1 Si, < 0.06 C, < 0.02 S, < 0.035 P, < 0.06 Cu (Si content may be increased if necessary)
Ferrozirconiumaluminum	17–20 Zr, 20–28 Al	2–3 Si, < 0.08 P, < 0.08 C, < 0.01 S, balance Fe
Aluminumchrome (complex alloys)	26–38 Al, 20–30 Cr, 25–32 Mo	1–18 W, < 3 Si, balance Fe

in the special machines. Ready briquettes may be promptly used in the electrolysis or put in a warehouse. The properties of the mass are shown in Table 3.

The briquettes are loaded into the shell and are being first sintered and then melted under the heat flow from the furnace. Normally, in the electrode, there are four main zones [27]:

Zone I: The mass is being heated and sintered at temperatures up to 65–70°C.
Zone II: In this zone, the binder (pitch) is liquid, temperatures 70–350°C.
Zone III: Here, the baking of coke starts at temperatures 350–550°C.
Zone IV: At higher temperatures, the mass is already baked and forms a solid electrode. Here, four sub-zones may be also detected [27].

To produce 1 t of aluminum, about 400–500 kg of anode mass are needed (370–420 kg are consumed for electrochemical oxidation, for foam and side oxidation by CO_2 30–80 kg, and for oxidation of secondary anode 10–15 kg). This means that electrode mass production should be roughly half of the primary aluminum production. Electrode mass should have the following parameters [27]:

- low specific resistance (< 75 μΩ m);
- low porosity (< 30%);
- relatively high strength (>30 MPa);
- low oxidation rate.

Table 3 Typical Requirements and Properties for Anodic Mass

Property	Mass type		
	A	B	C
Volatiles (%)	13–18	13–19	13–18
Ash (%)	< 8	< 9	< 7
Resistance (μΩ m)	90	150	87
Tensile strength (MPa)	1.47	1.47	1.47
Fluidity coefficient	1.5–2.5	2.2–3.0	2.0–2.7

Because self-baking anodes emit significant amount of hydrocarbon, etc. emissions, their use in aluminum smelters has been gradually decreased. New cells, nowadays, use pre-baked anodes.

6.2 Pre-Baked Electrodes

Pre-baked electrodes may be of two kinds: graphitized (graphite) and carbonaceous (coal). Carbon electrodes have higher specific resistance (~ 40 μΩ m) than graphite electrodes (~ 10 μΩ m), but bear higher compression loads (~ 36 MPa vs. ~ 25 MPa for graphite) [27,14]. Previously, these electrodes were seldom being used in aluminum electrolysis because of their higher costs and design peculiarities, allowing them little freedom to be fitted into the existing electrolysis cells. In the 1970s however, the use of pre-baked electrodes was favored also due to the growing environmental concerns [29].

For graphitized electrodes, thermoanthracite, petroleum coke (low ash), pitch coke, and coal pitch are used [14,27]. All raw materials are annealed at 1200−1300°C to reduce moisture (< 0.5%), volatiles, decrease resistance and increase density. Raw materials are screened, mixed and pressed (extruded) into green bodies which are being baked at 1200°C (green electrodes). These electrodes are being treated in special graphitizing furnaces by direct heating with electrical current at 2700−2900°C for a long time (almost one month). Ready electrodes are extracted from the furnaces and are mechanically treated to achieve the standard dimensions (Fig. 13). Coal electrodes are manufactured under the same scheme as for the graphitized electrodes, but their baking is completed at 1200−1300°C after 300−400 h, i.e. without graphitization. The quality of coal (carbon) electrodes is usually lower than that of the graphite ones [27].

Anode properties in the cell influence the behavior by three different mechanisms: air reactivity (oxidation kinetics), carbon dioxide reactivity ($CO_2 +$ $C \rightarrow 2CO$ reaction) and physical losses due to thermal shocks, fatigue and possible mechanical damages [29]. Pre-baked anodes have less porosity (typical density range of 1.53–1.60 g/cm^3) and, thus, gas permeability (typically 10 ± 5 Darcy units − higher for lower density anodes). Nevertheless, the reaction of anode with CO_2 (~ 960°C) is the dominant reaction for the carbon consumption in the bath. At higher levels, the temperature drops but the air penetration is more probable. The critical area is, thus, the alumina crust − air interface, where both carbon oxidation and burning due to air and carbon dioxide ($\sim 450−800$°C) are possible [29].
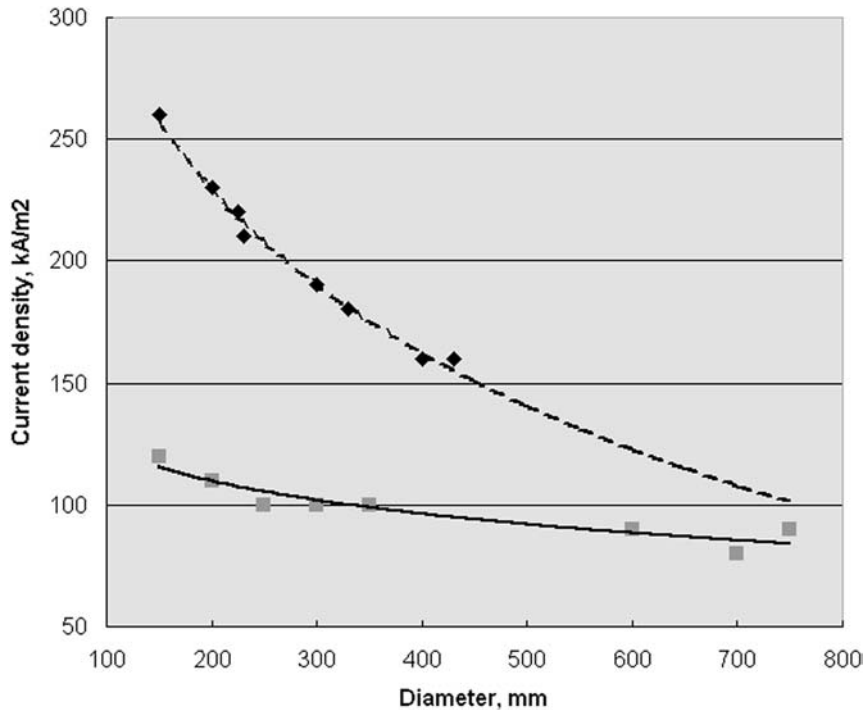
Figure 13 Practically achievable standard dimensions and current densities in the pre-baked graphitized (♦) and carbon (■) electrodes (after M. I. Gasik).

Thermal conductivity of pre-baked anodes depends mostly on the baking temperature — the higher the temperature (i.e. the larger the graphitization), the higher the thermal conductivity, which will result in: 3 ± 0.3 W/m K at 1000°C or 5–6.5 W/m K at 1350°C. Higher baking temperature affects also the mechanical properties of the electrode and its reactivity with CO_2. Higher temperatures decrease it.

Normally, the pre-baked electrode has four main zones [27,29]:

Zone I: The electrode is being heated at temperatures 200–400°C (at the edges).

Zone II: Carbon oxidation starts, temperatures 450–600°C.

Zone III: Electrode fully immersed in the alumina crust, temperatures 700–850°C.

Zone IV: Electrode is at its working position, intensive reaction with carbon dioxide and with alumina, temperature ∼960°C.

In zones I and II, temperature distribution is usually non-uniform. The excess carbon consumption for pre-baked anodes is similar to self-baked anodes, normally at 20–150 kg C/t Al [29]. Pre-baked anodes have smaller dimensions than self-baked ones because there is the difficultly to manufacture high-quality homogeneous graphite or carbon blocks of more than 1500 mm diameter [27]. There are also some activities in coated electrodes — many coating compositions (Al + SiC, Al + TiO₂,

H_3BO_3, SiC, Si) have been tested. Usually, a coating of ~ 0.8 mm thickness is stable until very high temperatures ($< 1850°C$), and its application decreases anode consumption at 15–20%. However, the costs of coating are also 15–25% higher, so the use of coating pre-baked anodes has to be carefully considered from the point of economy [27].

6.3   Inert Anodes

For the past years, all primary aluminum has been produced using an anode typically made of carbon. From an operating perspective, these types of anodes and cathodes are far from ideal. The carbon in the anodes is continuously consumed during the conversion process, forcing them, in the case of pre-baked cells, to be replaced about every two to three weeks. In addition, the undulating surface of the molten aluminum in the cell creates a large gap between the anode and the cathode to prevent short circuiting of the cell [17,19]. This gap makes the cell consume much more electricity than is needed to convert the alumina. The production of aluminum also creates fluoride and carbon dioxide emissions, as shown above [15,18,30].

The combination of an inert (non-reacting) oxygen-evolving anode, which would not be consumed during the electrolytic process, and an aluminum-wettable cathode, whose surface would remain stable (unlike the sloshing of molten aluminum) would allow for a more energy-efficient, productive, and environmental friendly cell operation. A narrowed anode–cathode gap requires less electricity (voltage drop) to produce aluminum. Use of a non-carbon anode will also reduce CO_2 and PFC (CF_4 and C_2F_6) emissions. In addition, the life expectancy of an inert anode (approximately 5 years) will significantly reduce the down time associated with anode replacement.

If such anodes can be combined with chemically stable cathodes and cell-lining materials, the cells can be made much more compact and probably also more energy efficient than the traditional cell.

The recent developments in inert anodes have been outlined in [18,31–33] in more detail. Basically, what the inert anode must be [31]:

- Physically stable at application temperatures (950–1000°C).
- Corrosion resistant (to both electrolyte and environment).
- Electrically conductive and electrochemically stable.
- Resistant to thermal shock.
- Mechanically robust.
- Easy to assemble and deploy.
- Recyclable.
- Cheap.

The efforts to date have been concentrated on the three major types of materials for inert anodes: ceramics, composites and metals [31]. The advantages and limits of these materials are shown in Table 4. For example, a composite of copper with nickel ferrite ($NiFe_2O_4$) and nickel oxide (NiO) was successfully tested (Fig. 14). Here, the cathode was also made of TiB_2–Al_2O_3 composite [31,34]. However, despite a successful test operation, the industrial application of these anodes was found to be difficult due to problems with aluminum purity and catastrophic failure of anodes due to thermal shock [31].

Table 4 Material Comparison for Inert Anodes

Material class	Advantages	Disadvantages
Ceramics	Does not easily oxidize (or not at all)	Low electrical conductivity
	High melting point	Reacts with electrolyte
	Electrochemically stable	Low thermal shock resistance
		Low mechanical properties
Metal-ceramic composites (cermets)	Higher stability and conductivity as for ceramics	Reacts with electrolyte
	Composition may be tailored	Questionable long-term stability
		Electromigration problem
		Higher manufacturing costs
		Metal contamination
Metals and alloys	High electrical conductivity	Difficult to find proper inert metal
	Good mechanical properties	Could be oxidized
		Potential bath contamination in an abnormal operation

All oxide candidate materials are, in fact, nobler than alumina and do have a limited noticeable solubility in the electrolyte [32]. This means that these oxides will be continuously reduced (i.e. anodes corroded) and will contaminate aluminum. The economy of the inert anode application was analyzed in [33]. To minimize the wear and corrosion rate to 2 mm/year, 1.7% of silver has been suggested to be added to the anode composition, as shown in Alcoa patent [33]. The cost of the inert anodes in a typical smelter operation was found to be compatible with that of carbon anodes. If additional advantages of the inert anodes will be taken into account (such as greenhouse gas emissions, PFCs, etc.), and the reasonable investment made for retrofitting the existing cells, inert anodes of the above composition may have an operating cost improvement over existing technology [33].

Application problems were also found for metal-based inert anodes. One of the more or less successful approaches includes Cu–Al bronze (7–15% Al) and it was developed by MIT [31]. During the operation, aluminum dissolved in copper, reacts with oxygen to form a protective thin alumina layer. This layer prevents copper from oxidation and reaction with aluminum, yet being too thin to conduct electricity within acceptable voltage drops. If the coating is broken or dissolved in the electrolyte, a new portion of aluminum will oxidize and form new coating. The wide range of current densities (0.25–2.5 A/cm^2) have been reported with very low level of copper in reduced aluminum ($< 0.1\%$) [31]. On the other hand, if the anode has failed, the whole aluminum bath was immediately contaminated with a large amount of copper. Besides copper, nickel-based superalloy with an optimal composition was reported to have a good oxidation resistance and high stability in the electrolyte at 960°C [31,32].
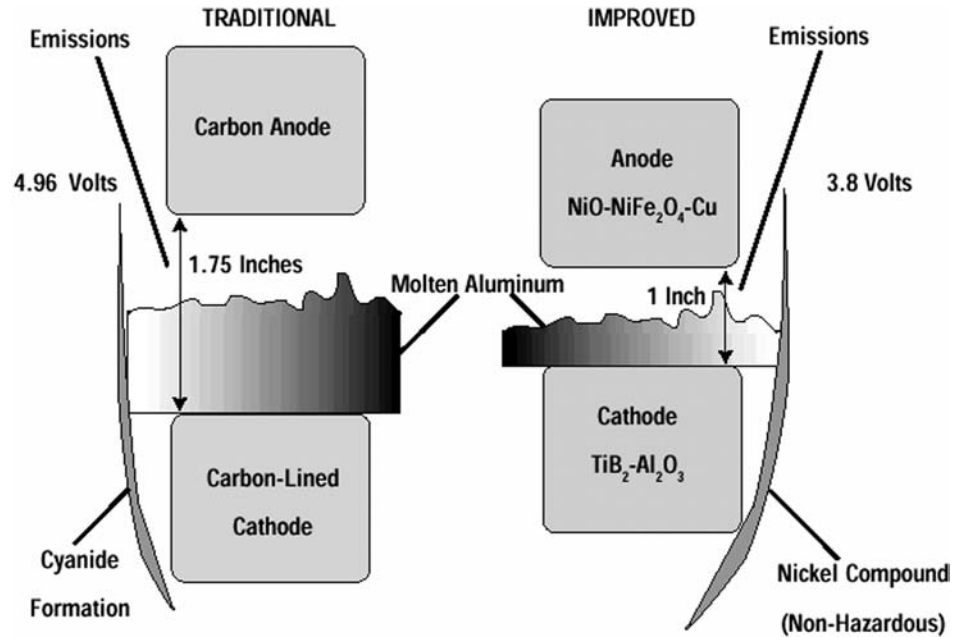
Figure 14 The comparison of traditional (carbon) and inert anodes operation in the electrolysis. (From Ref. 34.)

Other candidates for anode and cathode materials are also being studied, e.g. as shown in Fig. 15 [35]. Whereas, titanium diboride (TiB_2) was nominated as the best cathode material, it is expensive and far from optimal [32,33].

In general, despite its importance, the choice for the best "inert" anode material is still open. The metal seems to provide the best parameters of electrolysis but makes a huge problem with aluminum contamination, which cannot be recently solved in a cost-effective way [32,33]. One of the issues, which should be considered, is new advanced cell design, which may be more suitable for specific anode and cathode types than for conventional or retrofitted cell [33].

7   CONCLUSIONS

In this chapter, the following main areas of primary aluminum processing have been briefed: thermodynamics of alumina(s) and its reduction, basics of modern electrolysis (the Hall–Héroult process), alternative methods of aluminum extraction and electrode issues (self-baked, pre-baked, inert). In addition, attention has been paid to the direct reduction of alumina-contained raw materials into aluminum-rich primary alloys and special master alloys.

Each of these areas is a subject of intensive studies and developments of over more than 100 years and the limited space of this book would not be enough to list the critical references only. The authors have made an attempt to give a concise yet detailed outline for the process background, their technical realization, environmental, energy and economic issues. Many of these questions are outside of this
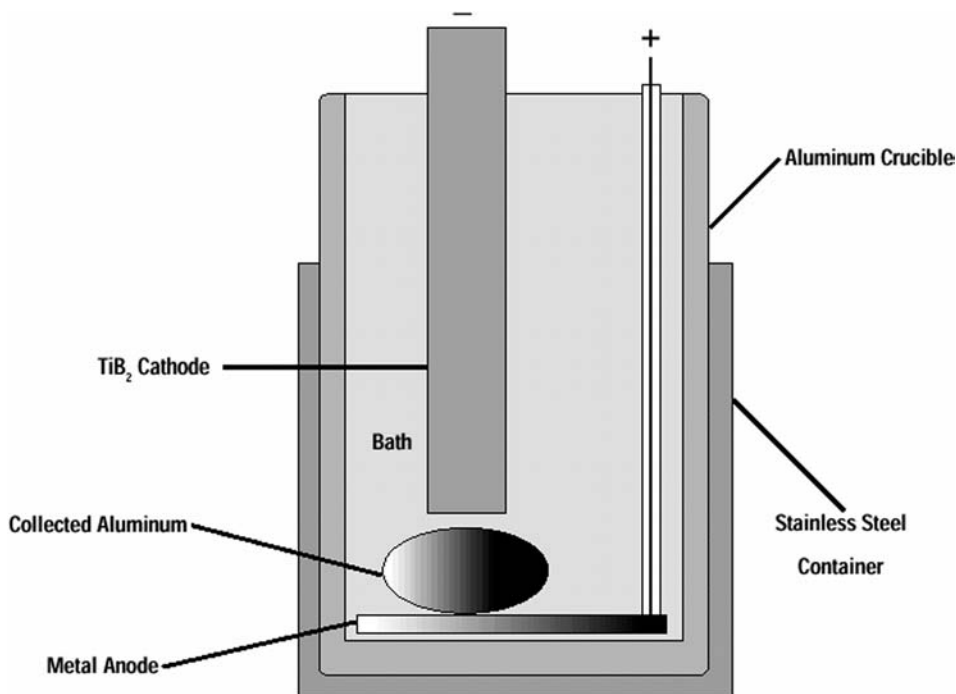
Figure 15 The experimental device for analysis of performance of inert anode. (From Ref. 35.)

chapter, but readers will find more details in the references below as well as in other chapters of this book.

The authors would like to thank all the contributors, who have taken part in the research in the last few years — particularly, in the area of direct reduction and electrode baking process development.

REFERENCES

1. M. Kh. Karapet'yants and M. L. Karapet'yans, *Thermodynamic Constants of Inorganic and Organic Compounds*, Ann Arbor–Humprey Science Publishers, London, 1970, 461 pp.

2. J. A. Holm and K. Lovi, "Investigation of Some European and South African Bauxites by Thermosonimetry," Chichester 1982, 1, pp. 306–312.

3. K. Lønvik, "Thermosonimetry," Thermochim. Acta 1987, 110, pp. 253–264.

4. G. M. Clark, "Some Applications of Thermosonimetry of Organic Crystals," Anal. Proc. 1986, 23(11), pp. 393–394.

5. A. N. Porada and M. I. Gasik, "Electrothermy of Inorganic Materials," Metallurgy Publ., Moscow, 1990, 232 pp. (in Russian).

6. M. Mizuno and H. Saito, "Preparation of Highly Pure Fine Mullite Powder," J. Amer. Ceram. Soc. 1989, 72(3), pp. 377–382.

7. M. M. Gasik and F. R. Sale, "Decomposition of Precipitated Aluminium Hydroxide Studied by Simultaneous Thermal Analysis," in: Proc. 10th ICTA Congress, Hatfield, UK, 1992, p. 25.

8. F. W. Dynys, M. Ljundberg, and J. W. Halloran, "Micro-Structural Transformations in Alumina Gels", in: Mater. Res. Soc. Symp. Proc. (C. J. Brinker, D. E. Clark, and D. R. Ulrich, eds.), Vol. 32, North-Holland, NY, 1984, pp. 321–326.

9. H. Nagai, S. Hokazono, and A. Kato, "Synthesis of Aluminium Hydroxide by a Homogeneous Precipitation Method. I. Effect of Additives on the Morphology of Aluminium Hydroxide," Brit. Ceram. Trans. J. 1991, 90, pp. 44–48.

10. D. E. Clark and J. J. Lannutti, "Phase Transformations in Sol–Gel Derived Aluminas," in: *Ultrastructure Processing of Ceramics*, *Glass*, *and Composites* (L. L. Hench and D. R. Ulrich, eds.), 1986, pp. 126–141.

11. M. Gasik, "Reduction Processes Kinetics and Experimental Methods," in: Proc. Seminar on "Solid State Red-ox Reactions", POHTO Institute, Oulu, Finland, 2001, 40 pp. (available on CD-ROM).

12. M. E. Brown, D. Dollimore, and A. K. Galway, "Reactions in the Solid State," in: *Comprehensive Chemical Kinetics* (C. H. Bamford and C. F. Tipper, eds.), Vol. 22, Elsevier, Amsterdam, 1980, 340 pp.

13. S. K. Mehta, A. Kalsotra, and M. Murat, "A New Approach to Phase Transformations in Gibbsite: The Role of the Crystallinity," Thermochim. Acta 1992, 205, pp. 191–203.

14. M. I. Gasik, N. P. Lyakishev, and B. I. Emlin, *Theory and Technology of Ferroalloys Production*, Metallurgy Publ., Moscow, 1988, 784 pp. (in Russian).

15. The Aluminum Industry Technology Roadmap, The Aluminum Association, 1997, 56 pp.

16. A. R. Burkin, ed., *Production of Aluminium and Alumina*, John Wiley and Sons Publ., London, 1987, 241 pp.

17. B. E. Paton and V. J. Lakomsky, "Welded Electrical Contacts between Metals and Carbon Materials in Electrometallurgy Equipment," Electrometallurgy 2001, 12, pp. 3–8 (in Russian).

18. H. Kvande and W. Haupin, "Inert Anodes For Al Smelters: Energy Balances and Environmental Impact," JOM 2000, 5, pp. 29–33.

19. V. J. Lakomsky, "Improvement of Current Distribution along a Cathode Block of an Aluminium Cell," Aluminium 1996, 72(11), pp. 822–826.

20. O. Braaten, A. Kjekshus, and H. Kvande, "The Possible Reduction of Alumina to Aluminum by Hydrogen," JOM 2000, 2, pp. 47–53.

21. M. I. Gasik, B. I. Emlin, N. S. Klimkovich, and S. I. Khitrik, *Electrical Smelting of Alumina–Silicates*, Metallurgy Publ., Moscow, 1971, 304 pp. (in Russian).

22. R. I. Ragulina and B. I. Emlin, *Electrothermy of Silicon and Silumine*, Metallurgy Publ., Moscow, 1972, 240 pp. (in Russian).

23. US Patent No. 3,892,558, Briquette for Producing Aluminum–Silicon, 1975.

24. I. M. Bastryga and M. I. Gasik, Production of Silumine by Electrothermy, in: Proc. Intern. Sci. Tech. Conf. Problems of Electrometallurgy, Nikopol, Ukraine, 1999, pp. 266–269.

25. P. N. Ostrik, M. M. Gasik, and V. D. Pirog, *Metallurgy of Powder and Sponge Master Alloys*, Kiev, Tekhnika Publ., Ukraine, 1992, 128 pp.

26. L. N. Gasik, V. S. Ignatyev, and M. I. Gasik, *The Structure and Quality of Commercial Ferroalloys and Master Alloys*, Tehnika Publ., Kiev, Ukraine, 1975, 152 pp. (in Russian).

27. M. I. Gasik, *Electrodes for Ore Reduction Furnaces*, 2nd Edn., Metallurgy Publ., Moscow, 1984, 248 pp. (in Russian).

28. M. A. Korobov and A. A. Dmitriev, *Self-Baking Anodes for Aluminum Electrolysis*, Metallurgy Publ., Moscow, 1972, 208 pp. (in Russian).

29. W. K. Fischer and R. Perruchoud, "Determining Pre-Baked Anode Properties for Aluminum Production," JOM 1987, 11, pp. 43–45.

30. Inventory of U.S. Greenhouse Gas Emissions And Sinks: 1990–1993, Report EPA 230-R-94–014, U.S. Environmental Protect in Agency, Office of Policy, Planning and Evaluation, Washington, DC, 1994, 27 pp.
31. D. R. Sadoway, "Inert Anodes for the Hall–Heroult Cell: The Ultimate Materials Challenge," JOM, 2000, 5, pp. 34–35.
32. J. Thonstad and E. Olsen, "Cell Operation and Metal Purity Challenges for the Use of Inert Anodes," JOM 2000, 5, pp. 36–38.
33. J. Keniry, "The Economics of Inert Anodes and Wettable Cathodes for Aluminum Reduction Cells," JOM 2000, 5, pp. 43–47.
34. Advanced Anodes and Cathodes Utilized in Energy-Efficient Aluminum Production Cells, Project Fact Sheet, U.S. Department of Energy, 1999, 2 pp.
35. Inert Metal Anode Life in Low Temperature Aluminum Reduction Process, Project Fact Sheet, U.S. Department of Energy, 1999, 2 pp.

3

Creation of Master Alloys for Aluminum

MICHAEL M. GASIK

Helsinki University of Technology, Espoo, Finland

VLADISLAV I. MAZUR

National Metallurgical Academy of Ukraine, Dnipropetrovsk, Ukraine

1 INTRODUCTION

This chapter deals with master alloys and other technologies, primarily directed at the grain refinement of aluminum and its alloys. Less discussed, but also very important, are issues of the microstructure formation and resulting properties of cast aluminum alloys as a result of applications of master alloys, modifiers, and special processing methods. Close control of the cast structure is a major requirement in the production of high-quality aluminum alloys. The most recently used method to provide a fine and uniform as-cast grain structure is to add nucleating agents to the melt to control crystal formation during solidification [1–4].

There exist many books, articles, and technical papers on master alloy manufacture, properties, and applications, to which the reader is encouraged to refer for more details (some, but not all, of them are listed in the references). In this book, we would like to present a more general outline of the basic issues (what these master alloys are, why they are needed, how to apply them) from the engineering point of view. The goal was also to point out the advances in aluminum alloy processing, which are often not taken into account by alloy manufacturers and users.

Section 2 of this chapter briefly describes the need for master alloys and their basic types and forms. Section 3 and part of Sec. 4 consider the physical metallurgy

of master alloys and their application practice. The rest of Sec. 4, as well as Sec. 5, describes novel techniques for grain refinement, microstructure, and property control in general. It is assumed that the reader is familiar with the general practice of aluminum alloy manufacture as well as phase equilibrium diagrams (these subjects are covered in other chapters of this book and are thus not included here explicitly).

2   MASTER ALLOYS FOR ALUMINUM

2.1   Principles of Master Alloys

One of the most important goals of aluminum alloy production by casting is the refinement of grain size, because coarse grain size immediately reflects in lower property levels. In the case of fixed smelting and casting conditions, this goal is usually achieved by different modification methods. These methods are conventionally based on the introduction of a small amount of special master alloys into a slightly overheated melt. During crystallization, the nature and appearance of the master alloy(s) play the main role in the grain size refinement [5]:

- element solubility in the aluminum solid solution, binary, or multi-component compound(s) according to the stable and metastable phase stability diagrams,
- surface properties (surface tension, adhesion between phases, adsorption, etc.),
- intermetallide morphology,
- their particle size distribution,
- their distribution uniformity,
- processing parameters (heating/cooling rates, temperature, time).

The first issue remains the main one for the development and selection of master alloys for particular grain refinement or modifying effects. It is based on the stable phase diagrams, assessed theoretically and experimentally, as well as on the chemical thermodynamics of the alloy systems [6–9]. The basic idea of "master alloys" has its grounds in the fact that many alloying elements are very active (Sr, Ti, Mg) or are just unsuitable (Na, B) to be entered in a pure form into aluminum alloys. Also, the production of purified metals like strontium or sodium with the purpose of dissolving only a small amount of addition (0.01–0.2%) is obviously not a cost-effective method. Many active metals can be manufactured much more easily in their alloy form (Al–Ti, Al–Sr, Al–Mg) due to lower activity of the leading elements. The dissolution of the master alloy in the melt also proceeds under a different mechanism than that of pure metals.

The master alloys used most often for aluminum contain titanium and other transition metals. These master alloys have been extensively studied for many years and are available commercially from different manufacturers [1–5,10]. It is necessary to note that the chemical composition of the master alloys is only one, but not the most important, parameter to characterize their efficiency as grain refiners.

Often, master alloys are compared on the basis of the intermetallide particles that already exist in the master alloy [5] and act as seeding nuclei during

crystallization of the melt, influencing, in this way, the grain growth. For example, for Al–Ti–B master alloys, the best modification effect was obtained with spheroid particles of $TiAl_3$ of 300 μm and $TiB_2 < 3$ μm [2,5].

In the case of Al–Si alloys, so-called primary silicon crystals as well as silicon-rich phases remain in the liquid even at high temperatures (800–1200°C). In this case, the morphology and the size distribution of silicon crystals will determine the final grain size of the alloy, and thus, modifiers should affect first the silicon rather than aluminum crystals. Many master alloys (like Al–Sr) are used mainly for the purpose of modifying, and not primarily grain refinement. On the other hand, some additions, like sodium, do not eventually affect the crystallization process by forming some intermetallides – rather, they change surface tension, adsorption, and other similar properties locally, and in this way prevent or assist the crystallization of one or another phase type, its polytype, or compound [8].

There are also master alloys with combined effect, acting as both grain refiners and modifiers. For example, an Al–10% Sr–2% B alloy (StroBor[®]) is used for this purpose [4]. Unlike traditional grain refiners, this master alloy has no titanium – rather, it takes advantage of the fact that both primary and secondary Al–Si foundry alloys already contain some titanium, either as a deliberate addition or as residual titanium in the scrap. Major intermetallic phases present in this master alloy include $SrAl_4$ and SrB_6. Upon addition to the foundry melt, these phases go into solution, and boron reacts with titanium present in the melt to form intrinsic TiB_2, which then serves as nuclei for grain refinement during solidification. This mechanism is different from that of the Al–Ti–B master alloy, which already has TiB_2 particles before alloying.

Laboratory tests were carried out on alloy 356 (7.0% Si–0.21% Fe–0.33% Mg–0.10% Ti) for specimens solidified at a rate of 0.8°C/s [4]. In this example, both master alloys were added to the melt at equivalent boron levels of 0.004% B, which, in the case of the StroBor alloy, also results in a 0.020% Sr addition. Both master alloys were added to an electric resistance furnace and held for a period of 3 hr. The furnace was not stirred during the holding period in order to assess the degree of fade in grain refiner performance with time. As shown by the grain sizes observed, both master alloys are effective grain refiners and produced essentially the same grain size. Following a 0.2% addition by weight of StroBor, the eutectic phase in the abovementioned test was fully modified and exhibited a finely "differentiated" fibrous structure.

Such master alloys [4,11] may have different compositions (7.5% Sr–1.3% B, 10% Sr–1% B, 10% Sr–2% B) and may also contain additional titanium (7.5% Sr–1.7% Ti–1.4% B). The selection of the master alloy is thus dependent upon the main problem to be solved (grain refinement, modification, properties adjustment, etc.).

2.2 Classification of Master Alloys

There exist a variety of master alloys for aluminum and aluminum alloys, depending on the principal (leading) element. Most master alloys have a special color code (Table 1) and classification numbers provided by the Aluminum Association (AA) and the European Standardization Committee (CEN). This assists in easier identification of the master alloys. Some of the master alloys with

Table 1 Basic Compositions of the Most Common Master Alloys for Aluminum

Principal element	Amount (wt. %)	Committee europeen de normalisation (CEN)	The aluminum association (AA)	Color codes[a]
B	3	90500	2203	Yellow (1 s)
	4	90502	2204	Yellow (2 s)
	5	90504	2217	Yellow (3 s)
	8	—	—	Yellow and black
Be	1 or 2.5	—	—	—
Bi	5	—	2005	Yellow and red
	3	98300	2003	Purple and yellow
	8	—	2016	Purple and yellow
Ca	10	92000	2009	White and orange
Cr	10	92402	2910	Purple (1 s)
	20	92404, 92405	2920	Purple (2 s)
Co	10	92700	2006	Orange and light blue
Cu	33	92900, 92901	2132	Orange (2 s)
	50	92902, 92903	—	Orange (3 s)
	54	—	2154	Orange (3 s)
Fe	10	92600, 92601	—	Black and brown
	20	92602	2820	Black and orange
Mg	10	91200	—	White and black
	20	91202	—	White and purple
	25	—	2010	White and purple
	50	91204	2011	White (2 s) and purple (2 s)
Mn	10	92500, 92501	—	Brown and white
	25	—	2425	Brown
Ni	10	92800	2500	Gray
	20	92802	2501	Gray (2 s)
Sb	10	95100	—	White and yellow
Si	20 or 25	91400, 91401	—	White
	36	91402	2302	White

Alloy	Composition			Color
Sr	50	2350	91403	White (2 s)
	3.5	2012	93800	Light blue
	5	–	–	Blue and yellow
	10	2007	93804	Light blue (2 s)
	12	–	–	Light blue (3 s)
	15	–	–	Light blue (2 s) and orange (1 s)
Si–Sr	10% Sr, 14% Si	2700	–	Light blue and white
Ti	6	2006	92202	Red
	10	2210	92204, 92205	Red and black
V	5	2605	–	Black
	10	–	–	Black
Zr	10	2600	94002, 94003	Dark blue (3 s)
	15	2615	94004	Dark blue and red
Zr–V	3% Zr, 2% V	2632	–	Dark blue and black
Ti + B	1.7 + 1.4	–	–	–
	3 + 0.2	2220	–	Green and purple
	3 + 1	2214	–	Green and brown
	5 + 0.1	2201	–	Green and red
	5 + 0.2	2207	–	Green and black
	5 + 0.6	2202	–	Green and yellow
	5 + 1	2252	–	Green
	10 + 1	2211	–	Green and white
Ti + C	3 + 0.15	–	–	Red and purple
	3 + 0.3	–	–	Red and orange
	5 + 0.18	–	–	Red and brown
	6 + 0.04	–	–	Red and white

[a] nS = number of stripes.

Note: Color coding of master alloys in the United States and in Europe may be different, and some sources have reported different colors used for the same elements, for instance: chromium – light blue (CEN), orange, purple (AA); iron – green (CEN), black (AA); manganese – silver gray (CEN), brown (AA); zinc – white (CEN), not specified (AA). Table 1 may serve as a guideline only, and thus, users are advised to consult the original standards and the color system used to avoid possible confusion.

uncommon content of the leading element(s) have no AA or CEN coding, but may have color coding, or vice versa. For example, Al–10% Sr–1% Ti–0.02% B master alloy (AA code H2017) has one light blue and one red stripe, and it is not listed in Table 1. Some plants produce their own kinds of master alloys in different tailored forms, often with a very high content of alloying elements (75–99%), and assign a special color code to them. This coding is sometimes not officially registered, but helps factory users to distinguish these alloys from other master alloy types. It is necessary to point out that the color coding of master alloys in the United States and in Europe may be different (Table 1), so every manufacturer has to ensure that master alloys are properly coded to avoid possible confusion.

Master alloys are available in a variety of forms: waffle ingot, notched ingot, slab ingot, sheared ingot, button, splatter (flake), broken ingot, coiled rod, Korrect^{TM} bar, and cut rod, depending on the leading element, manufacturing procedure, and the purpose of the master alloy itself (Table 2).

Besides "standard," or better said, the most common, master alloys (Table 1), there are many experimentally developed master alloy types for a variety of applications. The majority of them seem to have no international standards or clear domestic regulations (e.g. factory developments for their own needs). For example, several master alloys with rare earth (RE) metals have been developed and tested in China (Table 3) for Al–Si and Al–Cu cast alloy grain refinement etc. [12]. Most compositions have about 10% RE (near the eutectic composition of the Al–RE

Table 2 Typical Commercially Available Product Forms of Master Alloys

Master alloy product form (Common name)	Typical nominal specifications	
	Metric	English/American
Pigs, piglets	0.45–0.50 kg	∼1 lb
Waffles (in plates)	6.5–7.7 kg	14–17 lb
One kilogram notched ingots	∼1 kg	∼2.2 lb
Slab ingots	∼18 kg	∼40 lb
Pyramid-shaped buttons	0.23 kg	8 oz
Buttons	0.14–0.23 kg	5–8 oz
Sheared ingots[a] (cut cast bars)	0.34–3.63 kg	0.75–8 lb
Broken ingots (irregular pieces)	6.5–100 mm	0.25–4 in.
Flakes	0.03–0.3 kg	1–10 oz
Korrect bars[b]	∼11.2 g/linear cm	∼1.0 oz/linear in.
Rods (cut)	0.15, 0.33, 0.50, or 1.0 m	5.75, 12, 18, or 36 in.
Rods (in coils)	For 9.5 mm diameter: 180, 272, 454, 1360 kg	For 3/8 in. diameter: 400, 600, 1000, 3000 lb

[a] The cut bar is taken from a continuous cast trapezoidal bar with a cross-sectional area of 1.78 in. Each inch weighs ∼2.75 oz (0.078 kg). Products that are available in this form are, e.g., Al–Ti–B, Al–Ti–C, Al–Ti, and Al–Sr master alloys [4].
[b] Korrect bar is the trademark name of KB Alloys, Inc., for a 0.9 in. (23 mm) extruded bar. Products available in this shape are, e.g., Al–5% Be and Al–10% Sr.

Table 3 Rare Earth (RE) Master Alloys, Typical Types

Composition[a]	Typical applications	Main effects
Al–(5–22%) Si–(Me)–RE1	Al–Si cast alloys for pistons and similar applications	Grain refinement, minimizing gases and voids, improving wear resistance and strength, decreasing thermal expansion
Al–(4–8%) Cu–RE1	Al–Cu heat-resistant cast alloys	Decreasing cracking, enhancing strength and heat resistance (350–400°C)
Al–RE2	Soldering Al alloys	Improving solderability

[a] Me = Cu, Mg, Mn, Fe; RE1 = La, Ce, Y; RE2 = Ce, Mischmetal.

systems) to keep lower liquidus temperature (640–700°C). Higher-order master alloys (Al–Mg–Si–Fe–RE etc.) are also being used for deformed aluminum alloys (wire, cable, home products, foil, engineering and construction of Al–Mg and Al–Zn alloys). Due to many application fields, the comparison of RE influence on the alloys' properties has to be done on a case-by-case basis [12].

The microstructure of the master alloy plays an important role in the final modification effect. For instance, for "conventional" Al–Ti–B alloys, the average grain size of intermetallic compounds may vary significantly in different master alloy specimens (Table 4). Eventually, the grain refinement effect from master alloys of similar compositions will be different for alloys with 500 μm and 5 μm of TiAl$_3$ grain size [5]. For the examples of Table 4, the noticeable grain refinement effect at the level of 0.02–0.03% Ti was achieved by adding Al–2% Ti rod, and the best one − by Al–Ti–B master alloy (soaking time 4–5 min at 730–740°C). Rods and Al–Ti–B piglets have kept their modifying activity up to 60 min of soaking, but Al–1.8% Ti piglets − up to 80 min [5]. Similar results have been obtained for Al–Ti–B master alloys in both rod and piglet form [3] − the best grain size of

Table 4 Example of the Composition and Structure of the Al–Ti–B Master Alloys (all Master Alloys are Obtained from Different Manufacturers)

Alloy number	Composition (%)	Form of the master alloy	Microstructure parameters (TiAl$_3$ Phase)	
			Morphology	Average length (μm)
1	1.8 Ti	Piglet	Needles	300–800
2	4.5 Ti	Waffle plate	Needles	100–200
3	4 Ti	Centrifugal cast	Needles	40–60
4	5 Ti	Granule	Disperse needles	5–8
5	2 Ti	Rod	Disperse needles	6–10
6	4.5 Ti, 1 B	Piglet	Blocks	10–15

170–200 µm at the 0.0075% Ti level was received after 20 min (rod) and 160 min (piglets) soaking time. The effect above was mainly caused by smaller grain size of the rod master alloys than that of piglets. Smaller crystalline alloys have high solubility kinetics in aluminum, so their modifying effect may be reached almost immediately after introduction in the melt [3,5]. However, this does not automatically guarantee that rod-type master alloys are always better than piglets.

A comparison was made for short soaking times (up to 5 min) and lower titanium contents (<0.08%) for different master alloy types (Table 4). Here, the Al–Ti–B master alloy was superior to all the Al–Ti alloys studied [5]. The Al–Ti rod has had the least modifying effect, far less than that of the Al–Ti–B piglets, despite the same grain size of the $TiAl_3$ intermetallides. One of the possible reasons for this is the weaker stability of deformed intermetallides in comparison with refractory TiB_2 particles present in the Al–Ti–B piglets.

This example suggests that master alloy product data sheets should also have such parameters included as intermetallide mean grain size, microstructure, and the manufacturing method. This will assist in the most proper selection of the master alloys for specific aluminum alloy processing.

3 PHYSICAL METALLURGY OF BASIC MASTER ALLOY SYSTEMS

3.1 Phase Equilibria

Details of the Al–M (M = master alloy leading element(s)) phase equilibria may be found in numerous handbooks on phase diagrams [6–9] and in databases as well as dedicated publications, including this book. Usually, these stable phase diagrams are used for master alloy design, because they are directly related to both master alloy processing and master alloy application.

In master alloy processing, the main issues are the liquidus surface of the system in question and the thermodynamics of the Al–M solutions. The "ideal" system should have:

- optimal liquidus temperature (e.g. 50–100°C lower than that of the basic aluminum alloy),
- low vapor pressure (in the case of magnesium and some non-metals),
- reasonable activity of the alloying element (low activity helps to prepare the master alloy, higher activity helps in the reaction of the master alloy with the molten aluminum bath), and
- suitable intermetallic compounds formed during either production or application of the master alloy.

The following common additions are used for aluminum alloys, and the engineering practice has acquired the following main addition methods for them [13]:

- Boron is added to precipitate transition metal impurities such as Ti and V, and so improve electrical properties. It is mainly added as a dilute master alloy (<1% B).
- Bismuth is added both as the pure metal and as a dilute master alloy to enhance machinability of the aluminum alloys.

- Chromium prevents grain growth in Al–Mg alloys, inhibits recrystallization in Al–Mg–Si and Al–Mg–Zn alloys, and corrects for Fe to produce a golden color in anodizing. The main addition form is the concentrated tablet or briquette, although a significant quantity is also added as dilute master alloy.
- Copper is added mainly to increase strength, and it is added in many forms (pure metal, master alloys, concentrated tablets, and powder injection).
- Iron improves the high-temperature strength of some kinds of alloys. Addition forms are similar to that for copper.
- Magnesium provides high strength with good ductility, together with excellent corrosion resistance and weldability. For Al–Si alloys, it also improves heat treatment ability. It is mostly added as the pure metal, although master alloys are also widely used.
- Manganese improves strength and also plays a role in preventing recrystallization. The main technique of addition is as concentrated tablets or briquettes, powder injection, and, to a lesser extent, master alloys.
- Lead, like bismuth, is principally used to enhance machinability. The main method of addition is as the pure metal, although concentrated tablets and master alloys are also used.
- Silicon is used in "foundry alloys" (silumines) as it gives excellent fluidity in casting. It is also used in extrusion alloys, to which it contributes high mechanical properties. The main method of addition is as pure metal, but significant amounts are added as master alloys and through powder injection.
- Strontium is added to modify the eutectic in Al–Si alloys. It is mostly added as a dilute master alloy.
- Titanium provides an important contribution to grain refinement. It is mainly added as tablets (including pure metal), but most often through master alloys.
- Zinc is used to improve strength. It is almost exclusively added as the pure metal.
- Zirconium is added to inhibit recrystallization. It is mainly added as a master alloy (up to 15% Zr), but also as a concentrated tablet.

One may see that the variety of alloying elements and respective phase equilibrium diagrams do make it very difficult to define "optimal" liquidus temperature or "reasonable" activity. The situation is more complicated when the master alloy has several alloying elements. For example, Fig. 1 shows the phase stability diagram for the Al–Ti–B system at 700°C expressed through partial pressures of boron and titanium in the gas phase (equilibrium state). It is seen that very small variations of boron or titanium "pressures" may result in changes of the intermetallic compounds and thus phase composition of this master alloy.

   The most important use of thermodynamics of the Al–M system has, for many years, been in master alloy production. Irrespective of the method used (mixing of technically pure metals, reduction from salts, co-reduction from ores and mixed raw materials, etc.), the process should use the optimal driving force
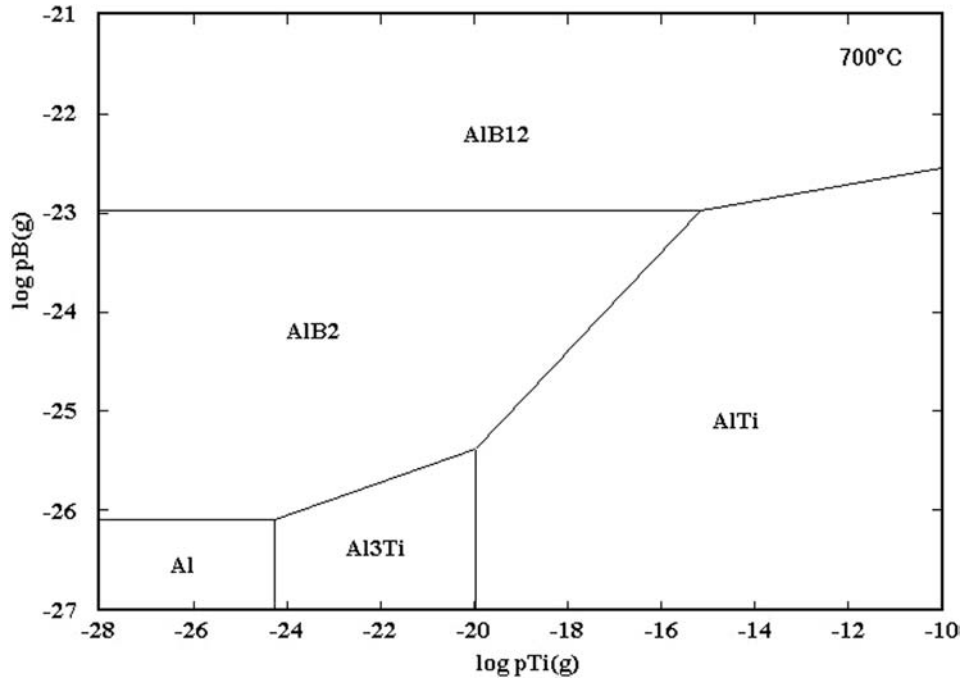
Figure 1 Calculated phase stability diagram for the Al–Ti–B system at 700°C (pure components, ideal solutions) in the coordinates of partial pressures of titanium and boron.

(minimal Gibbs energy of the system) to ensure proper recovery of useful elements with cost-effective technology. Gasik and Emlin [14,15] have described major methods of such master alloy processing (see Chapter 2 of this volume).

Whereas chemical thermodynamics assists in selection of the manufacturing process, care should be taken on the kinetics side to get purer master alloys with proper microstructure. In some cases, refractory intermetallides, which contain "harmful" elements (for certain types of aluminum alloys, they may be Fe, Ni, Mn, or Cr), could be filtered out of the melt. This is not possible for master alloys that already have "useful" intermetallides (TiB_2, Al_3Ti, etc.), so a small amount of impurities may trigger undesirable crystallization of the master alloy and affect its properties.

In general, phase equilibria and chemical thermodynamics (element activities in the melt, partial and integral properties (enthalpy, entropy)) are very important issues in answering the question: could this particular composition of the master alloy be produced within practically reasonable conditions (temperatures, gas pressures, used refractory linings, etc.)? If yes, then engineering implementation of the master alloy production could consider specific aspects of the process (charge materials, furnace type, metal and slag handling). If not, then the process may be still possible by, for example, reducing targeted element concentration (this will reduce its activity and thus improve thermodynamic prerequisites for the alloy formation) [14,15].

A special role of thermodynamics and metal physics is in understanding the mechanism of modification by certain master alloys. Analysis of electronic structure and pair interactions in the Al–Si–Fe–Mn–Mg–O system has led to the formulation of the hypothesis [36] that the most effective master alloy modifiers (X) for Al–Si castings will be those that decrease the orientations of the Si–X and O–X bonds [16,17]. The main result of the modifier is branching of the eutectic and Si-rich phases with their isotropic growth. This is explained by partial dissolution of X atoms in the lattice of the Si-rich phase, assisted by, e.g., hydrogen impurities (blocking Si–H, O–H bonds) and opposed by oxygen (formation of highly directed Si–O–Si bonds) [16,17]. If modifiers are added in excess, aluminum could form Al–X intermetallides, thus masking the effect of modification but in some cases suppressing excessive grain growth. Within this approach, there are no "good" or "bad" modifying elements, because their combined effect in the presence of other elements or impurities may be opposite to that of the addition of only one modifying element [36].

It is worth noting that there are other hypotheses explaining the modifying effects for aluminum and its alloys, and so far, there is no solid evidence to fully support one or another hypothesis.

3.2 Metastable Phases

According to modern understanding of the melt structure [5,18], the melt is not a homogeneous liquid but a colloidal solution with small particles and clusters, enriched by one or a few components of the alloys. This state can exist for a few milliseconds or hundreds of hours in a wide temperature range, may transform into another such state, and may be destroyed only after substantial overheating beyond the liquidus line [5]. If these small particles have sizes larger than the critical radius of the nucleus, they may act as seeds for melt crystallization (Table 5). On the other hand, if clusters have stoichiometry and near-order parameters too far from the phase coming up during crystallization, they cannot satisfy the orientation-dimensional principle (Dankov's principle) and thus cannot act as crystallization seeds. Instead, they may stay in the liquid until very low temperatures and trigger

Table 5 The General Crystallization Route for Alloys. T_h—Homogenization Temperature, T_{cryst}—Crystallization Temperature

Temperature	Melt composition	State of the alloy
$T \gg T_h$	Near-ordered or disordered liquid	Overheated melt
$T \geq T_h$	Atoms and clusters in dynamic equilibrium	Melt
$T_h > T > T_{cryst}$	Clusters and colloidal particles	Melt
	Prenuclei and crystallization centers	Melt
$T_{cryst} \geq T$	Overcooled liquid (Often metastable)	Solid, semi-solid or melt
$T_{cryst} \gg T$	Solidified alloy (Crystals)	Solid

the formation of phases not normally seen in the "phase equilibrium." Obviously, the mechanism of the grain refinement in this case will be different even if the aluminum alloy and the master alloy compositions remain the same.

Skripov [18] has determined the critical radius of nuclei as

$$r_{crit} = 2\sigma T_0 v/(L \, \Delta T) \tag{1}$$

where σ − melt/crystal interface surface tension, T_0 − equilibrium temperature of crystallization, v − crystal specific volume, L − latent heat of transformation, ΔT − liquid phase undercooling vs. equilibrium temperature. It was experimentally confirmed for a variety of aluminum alloys that not only ΔT values, but also T_0 values depend on the initial charge grain size and processing history (metallurgical heritage effects). For Al−12% Si alloy cooled at 3.1K/s from the same temperature, only alloys with initially coarse grains retained the equilibrium eutectic temperature (577.1°C) with $\Delta T \sim 6.1$K. If the alloy had small grain size before melting, T_0 decreased by 7.5 K with $\Delta T \sim 4$K. Alloys with different processing history but crystallized in the same conditions have shown a clear dependence on the crystallization kinetics and temperature values [5]. According to formula (1), the critical radius of the colloidal silicon particles in these alloys should be ~ 50 nm, which is very close to that of the silicon particles separated from the Al−Si melts by centrifugal sedimentation technique and experimentally observed in rapidly solidified thin films [8] (Fig. 2).

One of the most possible ways of grain size refinement by modifiers is also lowering of the liquid/crystal surface tension. With Eq. (1), it will significantly lower the critical work A_{crit} necessary for the formation of nuclei, since $A_{crit} \sim 32\sigma^3\{T_0/(L\Delta T)\}^2$ [5]. In the case of modifying of cast Al−Si alloys by sodium, the surface tension is estimated to decrease from 0.85 ± 0.003 to 0.617−0.797 N/m (i.e. by 7−30%), which may decrease the critical work for
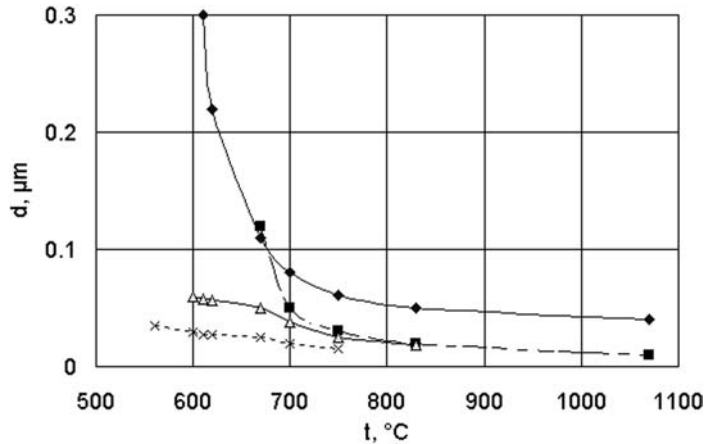
Figure 2 Primary crystal sizes (μm) found in rapidly solidified aluminum alloys quenched from different temperatures − Al−Si thin quenched solid film (◆), quenched drop (■), Al−Si modified by sodium (△), Al−Cu (×).

nucleation almost twice. It is necessary to say, however, that this surface tension is extremely difficult to measure as a function of temperature, crystal, and alloy prehistory as well as treatment time, so only occasional data have been published on that.

On the basis of Table 5, it is clear that the mechanism of the modification and, respectively, the grain refinement effect will depend not only on the modifier type and form, but also on the "point of the state" of the alloy (function of alloy's temperature). It was observed that no single reasonable hypothesis exists to describe a variety of metal solidification phenomena [16].

A. V. Mazur [19] has carried out a review of the analysis of structure-sensitive properties of Al–Me systems (Me = Zn, Cu, Mg, Si). For example, melt viscosity is considered to be one of the most sensitive properties to reflect structural changes in the melts. In pure Al, the viscosity curve has two "jumps" – at 770°C and 850°C. In the Al–Si system, the first one shifts to 790–805°C and the second one to 845–865°C (depending on the silicon concentration). The first temperature shift increases for higher silicon percentage [19]. A similar picture exists in the Al–Cu (770–800°C and 855–870°C) and Al–Mg (790–800°C and 880–890°C) systems. For the Al–Mg system, however, the second curve jump disappears in alloys with >20% Mg, and the second one for >40% Mg. The Al–Zn system seems to have no viscosity anomalies, at least in the ranges and temperatures studied. Thus, different alloy systems have different structures of liquid phase and hence crystallization mechanism and kinetics. At the modern level of metallurgical science, the effect of metastable phase formation as well as possible heritage effects should always be considered when processing structure-sensitive alloys and cast components.

4 MASTER ALLOY APPLICATION

4.1 Common Practices for Master Alloy Use

Common master alloys with a range of different Ti/B or Ti/C ratios (Table 1) are available to accommodate special conditions that may exist in the user's plant [4,11]. In selecting the proper grain refiner alloy, the user must take into consideration conditions such as the alloy to be treated, the quantity of recycle or secondary aluminum used, the desired grain size in the product, and the melting and casting practice used [4]. The method of introduction of master alloys into the melt depends on the purpose, the alloying element, and the manufacturing process. For instance, master alloys may be added in the furnace before melt release or in the mold just before the solidification starts. They may also be added in mixers, intermediate ladles, channel induction furnaces, transfer launders, semi-automatic bulk charging systems, and precise compositional adjustment for smaller foundries. Lumpy and cut forms of master alloys (Table 2) require weighing and subsequent feeding into the melt. Continuous rods may be submitted automatically from the coils, e.g., into a metal jet poured out of the furnace, into the mold, etc. There are a variety of feeding systems for the master alloys, and every factory uses the one best suited for their particular conditions.

Besides traditional forms of master alloys, i.e. ingots in particular forms, other shapes for their application have been developed. One of these forms is the master alloy briquette. It contains a controlled mixture of alloying element (75–85% or

more) in powder form, aluminum powder, and/or a sodium-free non-hygroscopic flux. The briquettes are produced by compressing the powder components between pillow-shaped indentations on two opposing rollers. Dissolution of the briquettes relies on the alloying element powder particles forming aluminides. It is recommended that the briquettes be charged after melting is completed and the bath has been skimmed at $>720°C$. Normal furnace stirring to ensure a uniform mixture is sufficient to assure complete dissolution of the briquettes. The flux aids in removal of the oxide film found on a powder particle to expose the metallic element to molten aluminum, promoting rapid dissolution.

A similar principle is used in alloying tablets, such as ALTAB^{TM}, for the precise compositional adjustment of alloy melts. These tablets are usually a mixture of alloying element (75–99%) in powder, sponge, or needle form, aluminum powder, and/or a flux. In the case of tablets or briquettes, the goal is more uniform distribution of alloying elements within the melt bath, faster dissolution of the master alloys, and more precise control of the metal properties.

The application of master alloys to Al–Si casting requires more careful consideration than for pure aluminum. Here, both grain refinement, and silicon dendrite and eutectic growth happen at the same time, so several acting mechanisms could result in uneven grain size distribution in the cast products. Authors [20] have studied the effect of Al–Ti–B master alloy additions on dendrite arm spacing (DAS) in the Al–8% Si–2% Cu alloy (DIN 226 S) rather than simply on grain size. It was independently shown [5] that elongation of all modified Al–9% Si alloys is inversely proportional to the square of DAS:

$$\delta, (\%) = 11.42 + 7.024 \times 10^{-4}/d^2 \tag{2}$$

where DAS (d) is expressed in millimeters. With different additions of titanium by Al–2% Ti master alloy, the effect of the Ti/B ratio from 1/6 to 6 has been investigated [20] for casting from $750°C$. Without boron additions, titanium at the 0.13% level eventually decreases the grain size (from 1.1 to 0.5 mm), but no further improvement is achieved at higher titanium content. Moreover, the DAS refinement is most efficient (from 35 to 20 μm) at 0.11% Ti, increasing with higher titanium content. Authors have conducted a regression analysis of these trends (Figs. 3 and 4), which indicate that the minimal grain size of 0.365 mm may be achieved at 0.17% Ti and a Ti/B ratio of about 0.22. On the contrary, the minimal DAS is achieved at 0.11–0.13% Ti and Ti/B ~ 2. Thus, the manufacturer will have to consider which property will be the most important to get after casting. It is necessary to note that there is no single opinion about the mechanism of the Al–Ti–B master alloy impact on the grain size and the DAS. Hu and Li [20] have explained it by complicated interactions of Ti, Al, Si, and B, but they did not study the effect of temperature and soaking time, which are very important kinetic parameters, or the possibility of metastable phase formation.

In another work [21], the influence of Ti, Na, and Sr additions (independently and simultaneously) to the Al–7% Si–Mg alloys was studied. The grain refinement and mechanical properties were measured for alloys poured from $720°C$. Titanium was introduced as Al–3% Ti–1% B master alloy, Na as metallic sodium in aluminum cans, and strontium as Al–9.9% Sr master alloy [21]. The titanium content was reported to be 0.18–0.19% ("with recommended range of 0.1–0.2%"),
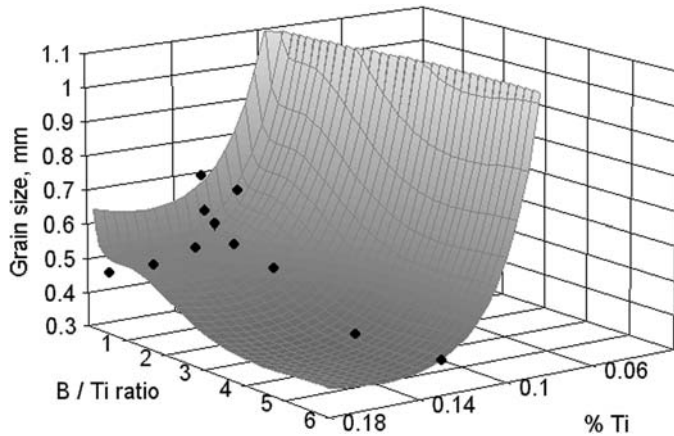
Figure 3 Estimation of the average grain size in Al–8% Si–2% Cu vs. Ti percentage and B/Ti ratio. Points are experimental data [20].

but results of the previous study (Figs. 3 and 4) show that there is a large difference between 0.1% and 0.2% Ti [5,20]. No boron content has been mentioned (from the master alloy composition it may be estimated as ∼0.05%, B/Ti ∼ 0.33 − quite far from "the optimal"). Although authors have found an effect of Ti, Na, and Sr on grain refinement, they have reported no significant increase in mechanical properties as a function of Ti, Ti–Na, or Na–Sr additions. This work illustrates typical underestimation of other casting parameters as well as the history of the alloy and initial state of the master alloys. Without such information, it is often not only impossible, but also misleading to draw any conclusions about the effect of certain kinds of master alloys or modifiers [5,19].
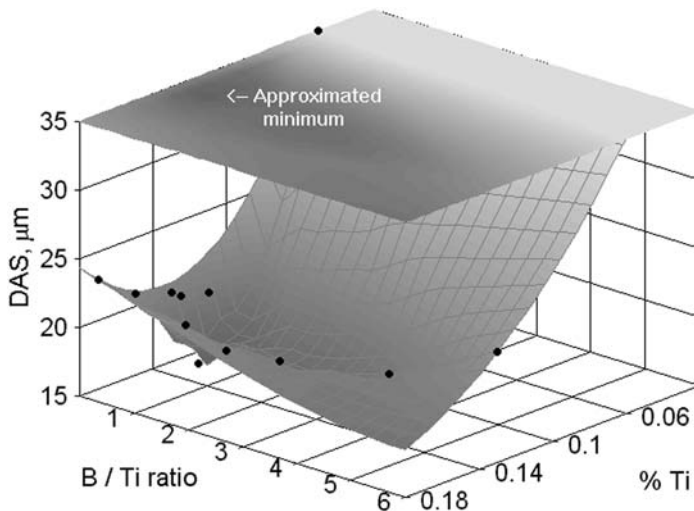
Figure 4 Estimation of the average dendrite arm spacing (DAS) in Al–8% Si–2% Cu vs. Ti percentage and B/Ti ratio. Points are experimental data [20].

As a main guideline, one has to remember that no "optimal" Ti/B ratio or Ti percentage exists for any alloy without exact reference to the processing parameters. As shown below, holding the melt at different temperatures may result in achievements quite the opposite of the expected ones.

4.2   Alternative Techniques For Master Alloy Addition

Conventional master alloys for grain refining in aluminum are inserted into the melt in their solid form (rods, waffles, lumps, etc.). For the Al–5% Ti–1% B master alloy, fine TiB_2 and $TiAl_3$ particles in the alloy are thus already present in the system, and the effect of grain refinement would obviously depend on the master alloy's "intrinsic" quality (morphology, particle size distribution, surface properties of these phases, etc.). Additional operations like thermal–mechanical treatment of the alloy to make preshaped rods or wires may also impact the efficiency of the refinement.

Solid master alloys need some time for dissolution in the molten aluminum. The kinetics of the dissolution will affect the refinement result together with the master alloy quality. To get the maximal effect, an incubation time of a few minutes is usually required for the uniform distribution of the alloying element through the melt. It was shown [22] that addition of the liquid master alloy would reduce this time period at least three times (from 15 to 5 min) and improve the refinement effect. It is necessary to note that preparation of the liquid master alloy and its dosing may present some additional difficulties in the manufacturing practices.

In some cases, there are limitations of manufacturing of the master alloy itself. For instance, Al–5% Ti–1% B alloy is often prepared by the introduction of KBF_4 and K_2TiF_6 salts into molten aluminum, followed by crystallization of the melt and rolling [1,23]. Since the idea of grain refinement by means of this master alloy application is based on the formation of fine TiB_2 particles, it was suggested [23] that such particles be produced separately, in the form of a nanopowder (2–30 nm, specific surface area 20–30 m^2/g) by plasma chemical synthesis. Such powders of TiN and TiCN were shown to significantly improve the properties of aluminum and decrease the grain size. Due to difficulties in the handling and processing of nanopowders in smelter and foundry factories, it was also suggested that special pellets of these powders be formed in a thin aluminum shell (master alloy pellets of Al–2.7% Ti). Introduction of such pellets of 8% wt. at 670–680°C with the following heating and crystallization [23] resulted in $\sim$30% finer structure of 99.5% Al in comparison with modification of the sponge master alloy Al–2% Ti. The disadvantage of this process is that complete dissolution of the pellets requires rather high overheating temperatures of the aluminum bath (1380–1400°C). As shown in the last part of this chapter, the effect of overheating and subsequent soaking may easily mask or overcome the effect of the master alloy addition, so such results should be considered with care.

It also worth noting that alternative techniques for adding master alloys are being extensively studied all over the world, but most of them are being applied for specific conditions only, and these results are unfortunately more or less ad hoc. As in the case of master alloy characterization, maximum information about experimental conditions is desired to be included with the results of the master alloy applications.

4.3   Novel Techniques for Grain Refinement and Property Monitoring

Traditional casting and foundry technology still remains a very complicated system, with many unclear links between the major processing parameters. It does not fully ensure correct information transfer from the starting charge to the final ingot stage [5]. The main question discussed above was the grain refinement process, but one may certainly ask: why do we need essentially smaller grain sizes — would it not be wiser to speak about the higher properties? In the casting of aluminum alloys, this is a very important question.

Earlier, the main goal of casting was to obtain smaller grain size, because the connection of the smaller grain size to better properties like strength and toughness is very well known in metallurgy. The practice, however, taught that a small grain size does not automatically guarantee high and stable property level. This has turned engineers and scientists to look in greater detail into the mechanism of the crystallization process and the morphology of the resultant phases.

As shown earlier, the composition and the structure of the melt (aluminum, alloys, or any other metals) is not always a homogeneous or near-ordered liquid (Table 5). Despite this fact being known for years, the utilization of the melt properties (i.e. before adding any modifiers) first for grain refinement is not extensively used in industry. The influence of heating temperature and casting temperature on the grain size of aluminum are demonstrated in Ref. [24]. When pure aluminum melt is poured from $820°C$, a rapid increase of the grain size is observed. This temperature was not influenced by the addition ($<0.6\%$) of any metal (Be, Mo, W, Ti, Fe, Cu, Mn, Zr), but when impurities added were over 1%, this temperature decreased to $810–815°C$. There is also numerous experimental evidence of anomalous electrical conductivity, oxidation kinetics, viscosity, density, and other properties [5,18,19,25]. It was suggested that molten aluminum has the following transformations in the liquid state [19]:

- FCC-1→BCC at $800–900°C$;
- BCC→FCC-2 at $1150–1250°C$;
- FCC-2→CP at $1390–1450°C$.

It is worth noting that earlier works did not find such anomalies or deviations of diffusion coefficient or electrical conductivity [26,27]. However, these anomalies have an order of magnitude of $\sim 5\%$, which is much lower than the experimental error of $10–15\%$ reported by earlier works. New experimental methods like electron and neutron beam as well as high-resolution X-ray diffraction have reasonable sensitivity, and their data leave no doubts about the complicated structure of molten aluminum and other metals, and other phase or phaselike transformations in the liquid state [8,19,28]. This information has recently been used for the development of new grain refinement and property improvement technologies. The novel techniques for grain refinement may be generally listed as:

- use of metallurgical heritage effect (variations in composition of the melting charge);
- use of thermal–temporal treatment (modification of temperature program and soaking times);
- use of external energy sources (laser, plasma, electron beam furnaces) to destroy clusters and associates and thus to impact the final microstructure of the alloy.

In this section, mostly the first (use of metallurgical heritage) and the second (temperature and time) methods will be considered. They become more and more important due to the increased amounts of recycled aluminum alloys. The charge compositions, microstructure, and alloying elements all influence the quality of the aluminum ingots. The knowledge of this influence is essential for the production of high-quality alloys.

The effect of metallurgical heritage (processing history) of the master alloys has been studied for Al–Ti and Al–Ti–B master alloys [5]. Master alloys with 1.5–5% Ti and 0–1% B have been produced by conventional casting into a pig iron mold (25–30 mm), centrifugal casting into a pig iron mold (5–6 mm), roll-casting between water-cooled rolls (1.5–2 mm), extrusion from ingot into the rod (10 mm), self-propagating high-temperature synthesis (SHS), and single and double remelting of the cast master alloy with mold casting.

The aluminum batch of 100 g was melted and heated up to 750°C. Master alloys were introduced into the melt for 0.01%, 0.005%, and 0.1% Ti respectively, stirred over 20 min and cooled in an alumina crucible at a low cooling rate (<30 K/min). It was found that for master alloys of the same composition, their manufacturing method has a strong influence on the modifying effect. For instance, SHS-made Al–2% Ti master alloy resulted in the same grain refining effect at 0.05% Ti as for Al–5% Ti–1% B master alloy at the 0.12% Ti level or cast Al–5% Ti master alloy at the 0.1% Ti level.

Similar results have been obtained for modification of the Al–6.5% Mg alloy. It was discovered that the Al–5% Ti–1% B master alloy produces practically the same results as boron-free Al–4% Ti–0.3% Mg or Al–1% Ti–0.2% Mg master alloys at the level of 0.02–0.04% Ti [5]. After heat treatment, the highest strength values of 300–305 MPa were obtained for Ti/Zr = 1.5 and Ti + Zr = 0.23% (Fig. 5) for the pressure-cast master alloys. In the case of Al–Ti and Al–Zr master alloys made of "fresh" charge, the maximal effect of 300 MPa was achieved at the level of Ti + Zr = 0.33% (higher titanium and zirconium contents substantially increase
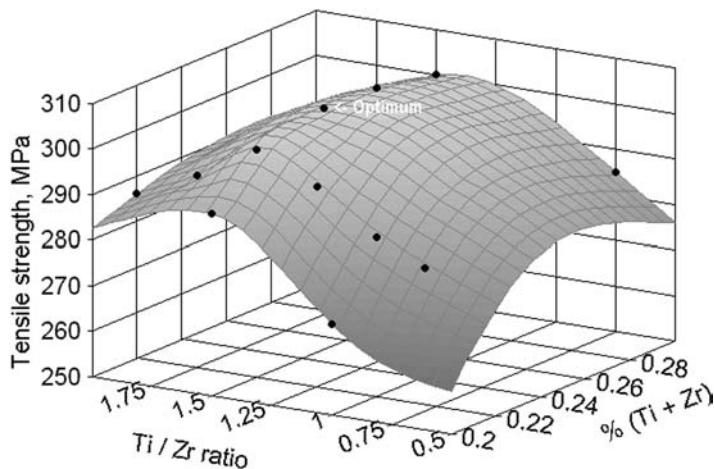
Figure 5 Strength of the Al–6.5% Mg alloy after addition of Al–Ti and Al–Zr master alloys vs. modifier concentration and Ti/Zr ratio. Points are experimental data [20].

hydrogen concentration). Thus, the use of pressure-cast master alloys allows one to get similar or higher results with 30—40% less addition of master alloys in comparison with the conventional type [5].

In many cases, the scrap of the same alloy is used during processing (remelting, recycling). After a certain period of time, returnable scrap and metal wastes have significantly higher concentration of alloying (modifying) elements and undesirable additions, which requires a respective adjustment of the modifying technique. For the case just shown of Al—6.5% Mg alloy, the use of charge with 50% own scrap and an addition of 3—10% of the same but pressure-cast alloy has required less master alloy addition. The same mechanical properties of 300 MPa strength and slightly higher elongation (25% in comparison with 20% for conventional) have been obtained with $Ti + Zr = 0.17\%$ and $Ti/Zr = 1.2$ [5]. Similar results were reported for Al—Cu alloys as well. This confirms the importance of knowledge of the maximal number of parameters of the alloy manufacturing process to select the proper modifying route.

For the Al—9% Si alloy, special experiments have been conducted [5], and correlation relationships were revealed for recommendation of the charge composition versus alloy casting parameters for the average ingot wall thickness of

- 10—20 mm $\rightarrow \sim 3\%$ of pressure-cast alloy, $\sim 0.5\%$ flux modifier, and $700 \pm 10°C$,
- 20—40 mm $\rightarrow \sim 5\%$ of pressure-cast alloy, $\sim 0.65\%$ flux, and $690 \pm 10°C$,
- 40—60 mm $\rightarrow \sim 7.5\%$ of pressure-cast alloy, $\sim 0.75\%$ flux, and $680 \pm 10°C$.

Here, the amount of pressure-cast alloy (scrap) of the same composition (i.e. Al—9% Si) should be introduced in the molten alloy, preferably in molds. These recommendations generally hold for the same "fresh" charge. If the amount of recycled own scrap is significant, the situation changes (Fig. 6). Here, one may see that flux modification (1—1.5%) is only advantageous for fresh charges. When the scrap percentage is over 40—50%, the addition of flux modifiers makes things only worse, and the best way is to keep flux under 0.5% [5]. Generally, the use of scrap with small grain size is almost as effective as the use of traditional modifiers and thus decreases master alloy expenditure by two to four times.

Maximal values of the properties (strength, elongation, etc.) are sometimes not the main goal in castings. In many cases, the stability of the properties from batch to batch and from one ingot to another is more important. For Al—7% Mg alloys, these optimal processing parameters have been found as 675—685°C pouring temperature, 0.05—0.08% Ti, and 0.7—1.0 Ti/Zr, depending on the wall thickness [5].

Many works [5,19,25,29—32] have demonstrated the influence of the thermal—temporal treatment (TTT) on the properties of aluminum and other alloys. For example, the eutectic Al—12.4% Si alloy has been prepared from pure elements at 750°C, heat-treated at different temperatures (Table 6) and then quenched from the liquid state [32]. Regimes 1 and 2 have corresponded to the preservation of the microstructural inhomogeneity in the melt, whereas overheating to >1000°C should destroy associates and clusters in the melt. The results, such as DAS and the eutectic's microhardness, clearly indicate that suitable overheating temperature and soaking time may produce ingots with similar or even better
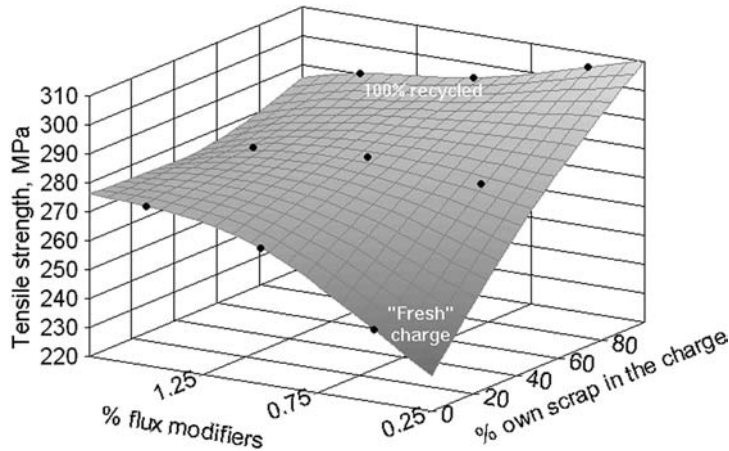
Figure 6 Tensile strength of Al–9% Si alloy vs. recycled scrap percentage and the addition of flux modifiers. Points are experimental data [20].

properties than those with added master alloys or modifiers [32]. This TTT technique is discussed in greater detail in Sec. 5.

Generally, two main guidelines may be formulated for grain refinement [5]:

- addition of specially made or collected alloys (including casting or rolling scrap) with small crystalline size to the charge − this will introduce seed nuclei of smaller size and thus have a modifying effect on the grain size;
- melt overheating to high temperatures and/or reasonable holding time at the isothermal conditions − this leads to structural changes in the liquid state (decomposition or phase changes in clusters, associates, disperse particles) and initiates the crystallization process according to the metastable phase diagram. This will provide the appearance of "unusual" metastable phases and certainly different microstructure in the final cast component.

The last, but not the least, is the influence of the processing environment. In many publications, and especially in the manufacturing practice, melting and casting are

Table 6 Variation of Experimental Processing of Al–12% Si Melts

Regime #	Temperature (°C)		Soaking time at T_{oh} (hr)	Cooling rate ($T_{oh} \rightarrow T_c$) (K/min)	DAS (μm)	Microhardness (MPa)
	Overheating (T_{oh})	Casting (T_c)				
1	800	800	5	−	40	700
2	800	700	5	4.2	35	710
3	1080	700	0.1	9	20	750
4	1090	700	0.1	9	15	900
5	1080	700	0.1[a]	486	<10	1250

[a]Additionally soaked for 3 hr at the casting temperature of 700°C and then rapidly cooled.

being done in "protective" or "inert" atmospheres. It is known that aluminum reacts with oxygen at very low partial pressures of the latter (at 750°C, the equilibrium oxygen pressure over Al is $\sim 4.4 \times 10^{-40}$ Pa). It is also clear that no furnace is capable of holding such low oxygen pressure, and thus, some small amount of aluminum oxide(s) will be always present in the system. These oxides may be rather stable even beyond 1200–1300°C, and may also act as crystallization nuclei. For example, for Al–20% Si alloy in alumina crucible, the equilibrium activity of dissolved oxygen will be ~ 25 times higher than in the same alloy, in the same atmosphere, but in a beryllia crucible [19].

5 ADVANCED TECHNIQUES FOR MICROSTRUCTURE MODIFICATIONS AND PROPERTY IMPROVEMENT FOR ALUMINUM ALLOYS

From the variety of aluminum alloys, one of the most important objects for studies on modification, grain size control, and property improvement is aluminum–silicon alloys. For the stable phase diagram, it is known that Al–Si eutectic, Al-rich (α), or Si-rich (β) phases are present in the system [6–9]. Thus, modifiers or other additions should be selected on the basis of their target phase − modification of the primary silicon (β-phase) is obviously a different process than modification of the α-Al crystals. This foresees that α-, β-, and eutectic phases do always exist in the form prescribed by the phase diagram − in other words, silicon crystal would always be the same in a hypereutectic Al–Si alloy whatever the temperature is, 700°C, 900°C, or 1100°C. It is, however, not the real case, and a variety of silicon crystals and eutectic morphologies appear in these alloys that do not fit to a simple eutectic phase diagram. It was suggested that in the Al–Si systems, some metastable phases may form during crystallization, and this is the reason for differences of the modification effect [28,33].

Authors [19,25,28,33] have studied in detail possible metastable phase transformations and equilibria in the Al–Si system for realistic processing conditions. The starting point was that the Al–Si system is very similar to the Fe–C system (metal–non-metal, eutectic equilibrium), but formation of the metastable cementite Fe_3C in the latter is well known and easily realized [33]. It was suggested that similar metastable phases do exist in the Al–Si system as well, but their stability may be of a few orders lower than that of cementite. In the 1970s the first metastable Al–Si phase (cubic X-phase, $a = 0.69891$ nm) was found in rapidly solidified thin films. Analysis of the Al–Si alloys with 6–40% Si has demonstrated such a variety of morphology of eutectic, α-, and β-phases as cannot be explained by simple eutectic crystallization [33]. There, two other metastable phases (η, ω) have been found [28,33]. In the near-eutectic region (10–12% Si), it was discovered that these phases also form metastable eutectics (η + X and α + ω), and their composition depends on the overheating temperature. Iron impurities (up to 0.3%) were observed to stabilize these eutectics. Finally, additional phases have been found, and analyzed by X-ray diffraction and microanalysis in the Al–Si hypereutectic alloys [19,25] (Table 7). These phases may stay for several minutes, weeks, or days and may be discovered by X-ray diffraction and transmission electron microscopy (TEM). For example, the tetragonal ψ-phase was found by TEM electron beam to fully decompose in 4 min, but in the Al–18.68% Si alloy with 0.05% Sr, its stability was high enough to take lattice parameter measurements [33].

Table 7 Metastable Phases in the Al–Si System. (See Ref. [33])

Phase name	Structural type and parameters (nm)	Silicon content (at. %)	Possible formula
X	Cubic, $a = 0.69891$	7.0–8.5	Al_6Si
η	Rhombohedral, $a = 0.74770$, $b = 0.7663$, $c = 0.5759$	15–18	Al_4Si
ω	Hexagonal, $a = 0.50597$, $c = 1.10225$	29–32	Al_7Si_3
æ	Tetragonal, $a = 0.3695$, $c = 0.7160$	37.5–38.5	$Al_{62}Si_{38}$
ψ	Tetragonal, $a = 0.4500$, $c = 0.8199$	42–45	Al_3Si_2
ρ	Tetragonal, $a = 0.4487$, $c = 0.5160$	51–52	AlSi

Metastable phases decompose into Al–Si eutectic with the respective neighboring phase (α or β) with various morphologies, depending on the starting metastable phase. This effect is actually responsible for the properties and their potential degradation in the Al–Si castings [19,28]. Respective "metastable Al–Si phase diagrams" for practical values of cooling rates have been constructed [19,33] (Fig. 7).

This phase diagram summarizes the discovered metastable phases, but it is necessary to note that appearance of the phases depends on the overheating
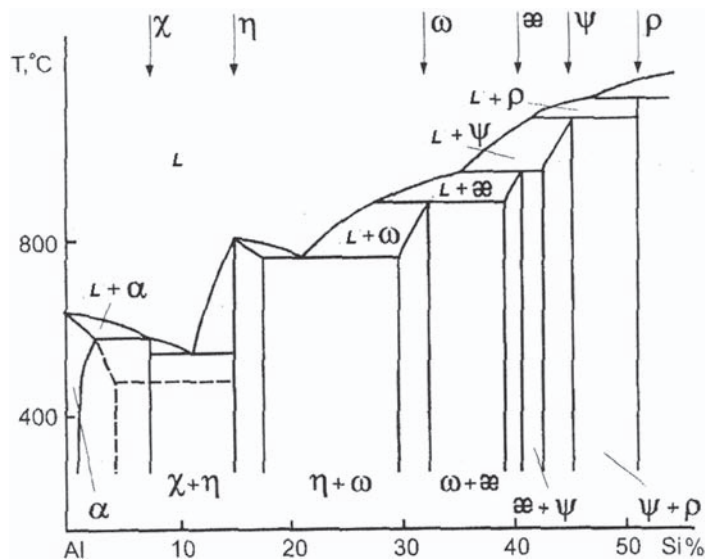
Figure 7 Experimentally found metastable phases in the Al–Si system and their estimated phase diagram.

temperature and soaking time at this temperature. Besides this, it was also found that the phase appearance sequence depends on the time for isothermal solidification [19,28]. Isothermal solidification of the Al–21.5% Si alloy was studied by overheating synthetic melt to different temperatures for 30–40 min, transferring it to the thermostat made on the basis of Al–Cu–Si alloy (for 1–100 sec), and quenching [19,25]. The resulting specimens were studied by different methods and phase stability areas, and their transformation kinetics were determined. As an example, Fig. 8 shows the phase formation sequence in the Al–21.5% Si alloy, overheated to 900°C, soaked for various times at different temperatures, and rapidly quenched. At ~ 700°C and ~ 750°C, the liquid remains stable for quite a long time, but the æ-phase appears in the liquid after incubation for ~ 20 sec above 700°C and the ω-phase, below 700°C. Unlike for the melt overheated to 830°C, the æ-phase does not turn into the ω-phase but disappears, forming liquid. At lower isothermal soakings, the ω-phase seems to be the first one, followed by the η-phase below 650°C. Approaching stabile eutectic temperature (577°C), the phase change sequence gets more complicated, but after ~ 100 sec of soaking, only α- and β-phases remain in the structure. It is important that after decomposition of the metastable phases, only silicon crystals and the eutectic will appear in the microstructure, but they will have different morphology and properties due to the various ways of metastable phase transformation [19,28]. Eutectic $\alpha + (Al, Si, Fe)$ is known to make alloys brittle, and thus, ways of stabilization of the ω-phase should be sought (here (Al, Si, Fe) means an iron-containing intermediate phase, the composition of which is not yet exactly
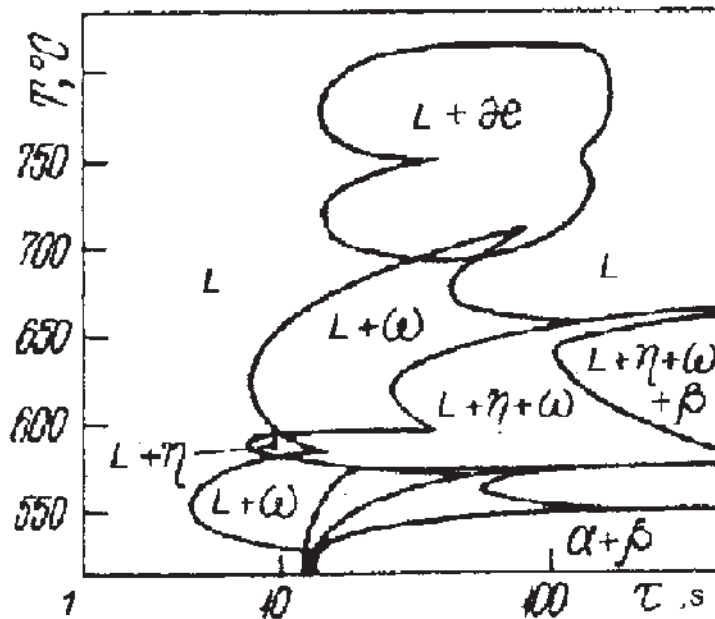
Figure 8 Formation of the metastable and stable phases in the Al–21.5% Si alloy overheated to 900°C and isothermally cooled at different temperatures.

known). This is an important measure for recycled Al–Si alloys with increased iron content.

The influence of iron additions on changes of the isothermal crystallization was also studied [28] (Fig. 9). In the Al–21.5% Si–2% Fe alloy, preheated at 830°C, during such crystallization, the starting liquid between 3 and 40 sec transforms to:

700–750°C   the stable liquid
680–700°C   liquid $+\omega$
590–680°C   liquid $+\omega\rightarrow$ liquid $+\omega+\eta$
570–580°C   liquid of increased stability$\rightarrow$ liquid $+\omega\rightarrow$ liquid $+\omega+\eta$
530–560°C   liquid $+\omega+\eta\rightarrow$ liquid $+\eta+(\omega+\eta)_{\text{eutectic}}$

In comparison with non-alloyed Al–Si melt, iron additions decrease the incubation time for ω-phase formation and increase its stability area. In the case of oxygen additions (Al–21.5% Si–2% O) (Fig. 10), the starting liquid has the following transformations [28]:

730–750°C   the stable liquid
590–720°C   liquid $+\eta$
570–580°C   liquid of increased stability $\rightarrow$ liquid $+\eta$
530–560°C   liquid $+\eta\rightarrow$ liquid $+\eta+\alpha+(\alpha+\eta)_{\text{eutectic}}\rightarrow$
            liquid $+\eta+\alpha+(\alpha+\beta)_{\text{eutectic}}\rightarrow\alpha+\beta+(\alpha+\beta)_{\text{eutectic}}+(\alpha+(\text{Al},$
            Si, Fe))$_{\text{eutectic}}$

Here, it is seen that oxygen suppresses formation of the ω-phase formation and promotes formation of the more brittle $(\alpha+(\text{Al, Si, Fe}))$eutectic.
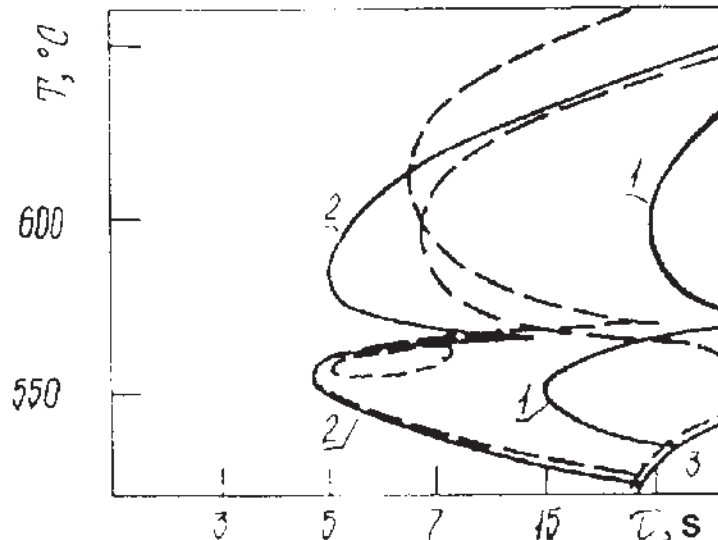
Figure 9   Isothermal crystallization of the Al–21.5% Si–2% Fe alloy, preheated at 830°C: 1 − liquid $+\omega+\eta$, 2 − liquid $+\omega$, 3 − $\alpha+\beta+(\alpha+\beta)_{\text{eutectic}}+(\alpha+(\text{Al, Si, Fe}))_{\text{eutectic}}$. Dashed lines are phase areas of unalloyed Al–21.5% Si alloy, as shown in Fig. 8.
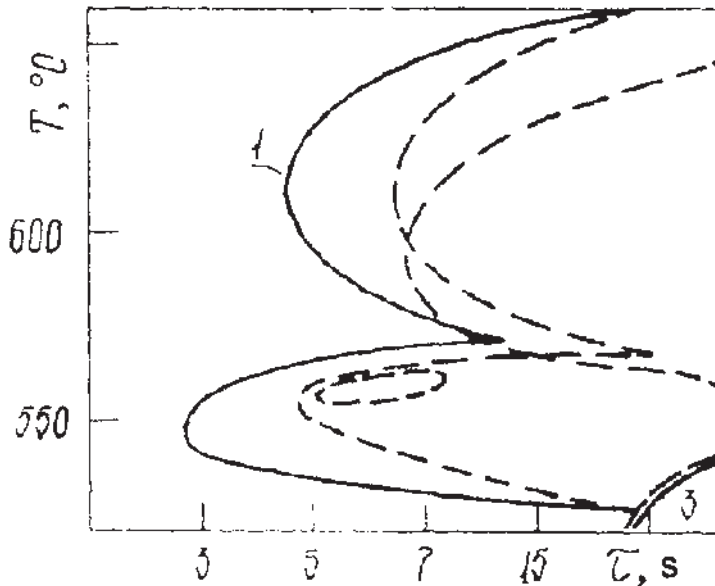
Figure 10 Isothermal crystallization of the Al–21.5% Si–2% O alloy, preheated at 830°C: 1 − liquid + η, 3 − α + β + (α + β)$_{eutectic}$ + (α + (Al, Si, Fe))$_{eutectic}$ (note that the stability area of "liquid + ω" is greatly suppressed). Dashed lines are phase areas of unalloyed Al–21.5% Si alloy, as shown in Fig. 8.

Comparison between microstructures of Al–Si alloys with and without oxygen and iron additions has shown significant differences in eutectic fractions and crystal morphology (cross-sectional area variation is 8−10 times). Metastable phases, which do form first during the crystallization, have always some amount of stabilizing additions (impurities) hindering their decomposition. For instance, the η-phase (Al$_4$Si) may contain substantial amounts of oxygen [19,25], and its lattice parameters are close to those of the sillimanite Al$_2$SiO$_5$. Also, the hexagonal lattice of the ω-phase is isomorphic to the intermetallide Al$_{12}$Fe$_3$Si$_2$ [9,25], and the ω-phase may thus be stabilized by iron additions. The lifetime of these phases under other equal conditions depends on the oxygen and iron concentrations in the alloy. That is why recognition of these metastable phases in pure Al–Si alloys is very difficult, due to their fast transformation into various forms of silicon and eutectics.

A. V. Mazur has studied the morphology of silicon crystals in the Al–Si alloys and revealed some peculiarities of their shape and surface correlation with practical properties of Al–Si castings [19]. It was found that special physical–chemical treatment (patent pending) of the Al–12% Si–2% Cu–1% Mg–1% Ni melt leads to certain transformations in silicon crystal growth and in its shape (Fig. 11). The silicon shape changes have led to simultaneous increase of both the strength and the ductility of the Al–Si alloy, which has led to a significant improvement in diesel engine piston quality.

Similar data for metastable phase formation and their influence on properties were obtained for hypoeutectic Al–Si alloys [35]. In that study, the influence of not only overheating temperature but also cooling rates (0.3−1000 K/s) on
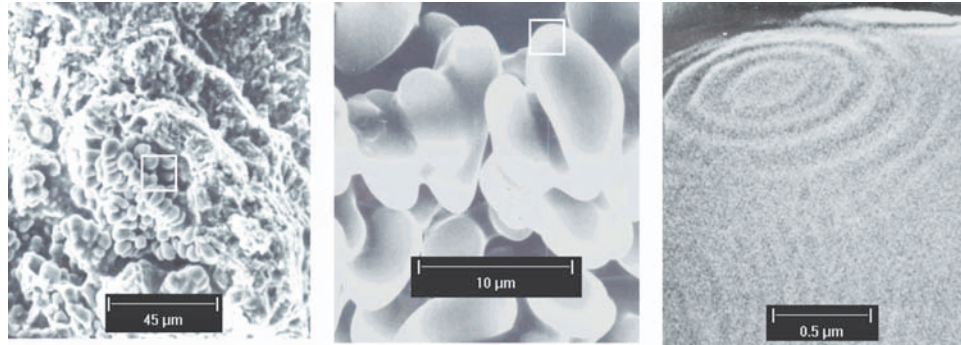
Figure 11 Shape and morphology of the silicon crystals in the Al–Si alloy. White squares indicate approximate places for the photos on the right.

microstructure and properties has been analyzed. Figure 12 shows a contour plot of the DAS parameter (μm) of the Al–7% Si alloy overheated to different temperatures. It is seen that higher overheating usually leads to smaller DAS, although there are clear anomalies in the range of 700–750°C and after 830°C (where double decrease in the DAS at low cooling rates $\sim 1\,$K/s may be obtained by increasing overheating). High cooling rates of 1000 K/s have smaller influence on DAS irrespective of overheating [35].

The hardness of the Al–7% Si alloy is shown in Fig. 13. The dependence of hardness on overheating is now similar for all cooling rates studied. There are peaks of hardness at 700°C and a "valley" between 740°C and 800°C, where the lowest hardness values have been found at any cooling rate. One may calculate that the target hardness level of 45–50 HRB of this alloy may be achieved at cooling rates <1K/s after overheating at either 700 ± 10°C or 860–880°C. If the overheating temperature is 740–760°C, then this hardness level may only be achieved at cooling rates >1000 K/s. Detailed X-ray diffraction analysis carried out in Ref. [35] has confirmed metastable X- and η-phase formation. Thus, the isothermal crystallization diagram for the Al–7% Si alloy may be drawn as shown in Fig. 14. This indicative diagram is valid for the alloy that was over-heated to 900°C. If the overheating temperature is different, a new crystallization diagram may be constructed, and the optimal temperature regimes could be determined.

Since the determination of crystallization fields and metastable phases is very time-consuming, it is important to develop measurement techniques for faster analysis of the solidification processes. A good choice for that are thermoanalytical techniques. Thermal analysis measurements by differential scanning calorimetry (DSC) were made by authors on the unmodified Al–7% Si, 21.5% Si, and 36% Si alloys [34] using a Netzsch STA449C "Jupiter" simultaneous thermal analyzer. It allows very precise measurement of the heat flow with a temperature resolution of 0.1°C and mass changes of $\pm 0.1\,$mg. Specimens of ~ 100–200 mg have been heated, held, and cooled in pure argon with different rates up to 830°C, 900°C (the abovementioned critical temperatures in the metastable Al–Si system), and 1480°C (to destroy possible associates, clusters, or
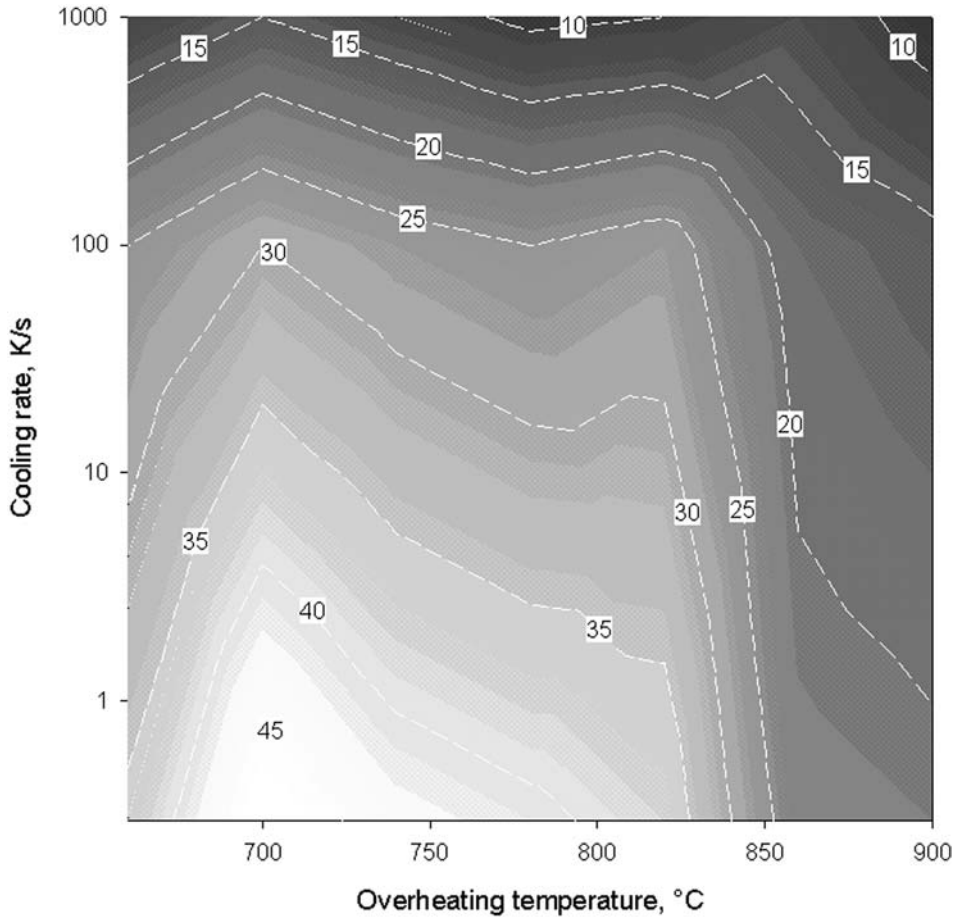
Figure 12 DAS (μm) in the Al–7% Si alloy overheated to different temperatures and crystallized from these temperatures at different rates.

other heritage effects). Thermal effects have been registered at heating and cooling that correspond to both stable and metastable equilibria (Figs 15 and 16). In Fig. 15, the effect of the cooling rate is seen to impact significantly on the metastable peritectic-type reaction, but causes only small oscillations in the onset and peak of the "liquidus" of this alloy. If these changes had been caused by faster cooling rates only, the extrapolation on the zero cooling rate would result in the "stable" liquidus of the Al–21.5% Si alloy ($\sim 723°$C), which is obviously not the case (the extrapolation of both peak and onset temperatures gives $\sim 680°$C). The complicated dependence of the metastable peritectic line (Fig. 15) suggests that such extrapolation is unlikely to be valid, because at lower cooling rates, the system would probably go via a different crystallization route according to another metastable phase diagram.

Figure 16 shows thermal effects in three alloys with different silicon concentrations on heating and cooling. Only hypereutectic alloys have metastable
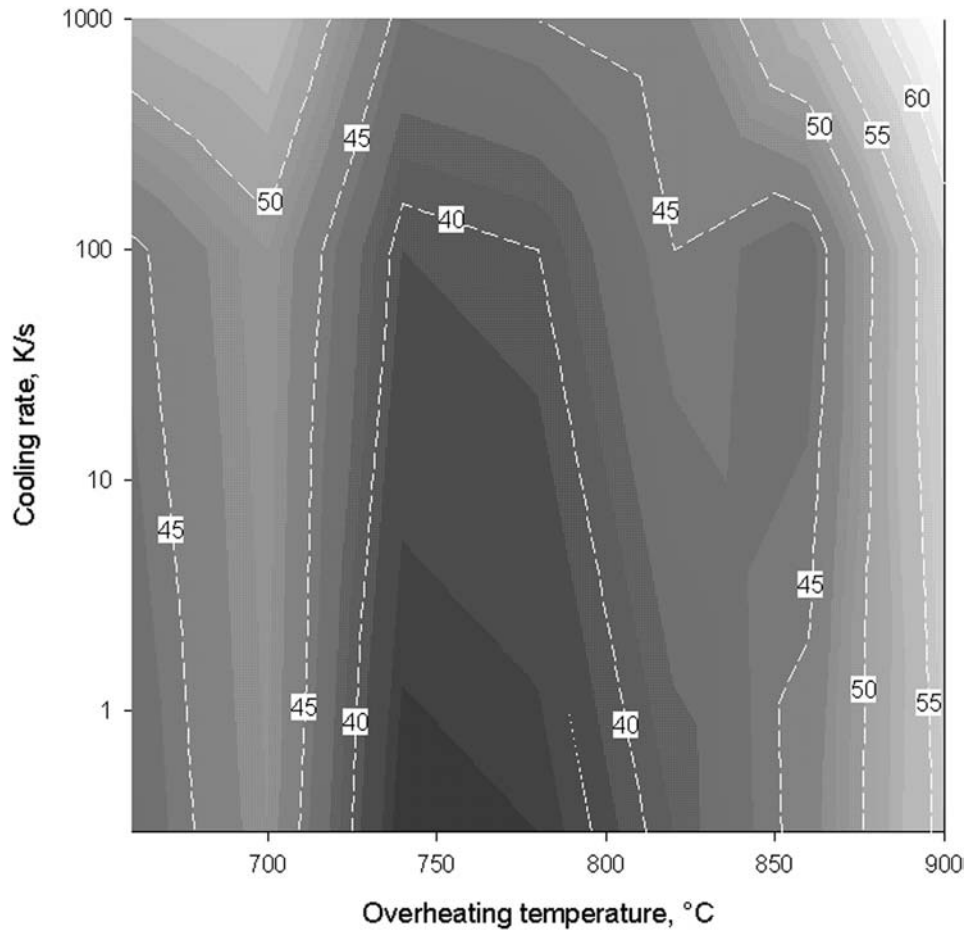
Figure 13 Hardness (HRB) in the Al–7% Si alloy overheated to different temperatures and crystallized from these temperatures at different rates.

peritectic peaks between "stable" liquidus and eutectic thermal effects [34]. What is more interesting is that there are reproducible thermal effects, associated with the "phase transformations" in the liquid state, far above the "stable" liquidus line for all the alloys. For hypereutectic alloys, the clear effect at 934–937°C exists at cooling, whereas, for a hypoeutectic alloy (7% Si), it is situated near 680°C. These findings demonstrate the usefulness of the thermal analysis application for studying crystallization phenomena (actually, these methods have been used for this purpose before). Here, the point is in very careful application of known techniques (e.g. activation energy, kinetics, or enthalpy determination with non-isothermal methods) and critical assessment of the results, because data received with different cooling rates may easily reflect different crystallization processes. In this case, simple comparison of the cooling curves makes very little sense for process optimization.

It has been experimentally confirmed [19,25] that pouring of the Al–Si melts from temperatures higher and lower than those mentioned above results in different
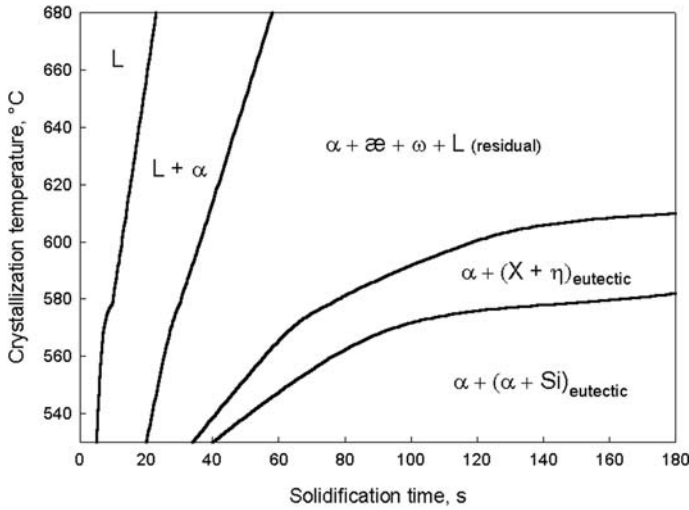
Figure 14 Schematic of the isothermal crystallization of the Al–7% Si alloy overheated to 900°C.

microstructural appearance of silicon and aluminum-rich crystals and, respectively, properties of castings. For example, the alloy Al–21.5% Si–2.5% Cu–2.5%Ni was heated up to 920–940°C (this corresponds to the phase transformation in the liquid phase; Figs. 15 and 16), held there about 15 min and then cooled down to the pouring temperature (810–940°C). As-cast specimens were analyzed for their microstructure and mechanical properties (ultimate tensile strength, fracture toughness, microhardness, hardness). On the basis of more than 1000 specimens,
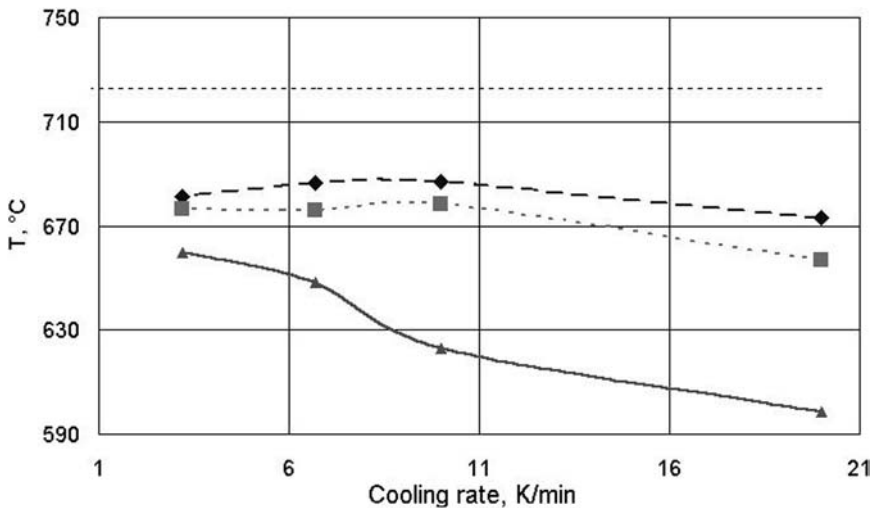
Figure 15 Thermal effects in the Al–21.5% Si alloy overheated to 1480°C and cooled at different rates – temperature onset point at liquidus (◆), peak temperature for liquidus (■), metastable peritectic (▲), stable liquidus border (- - - -).
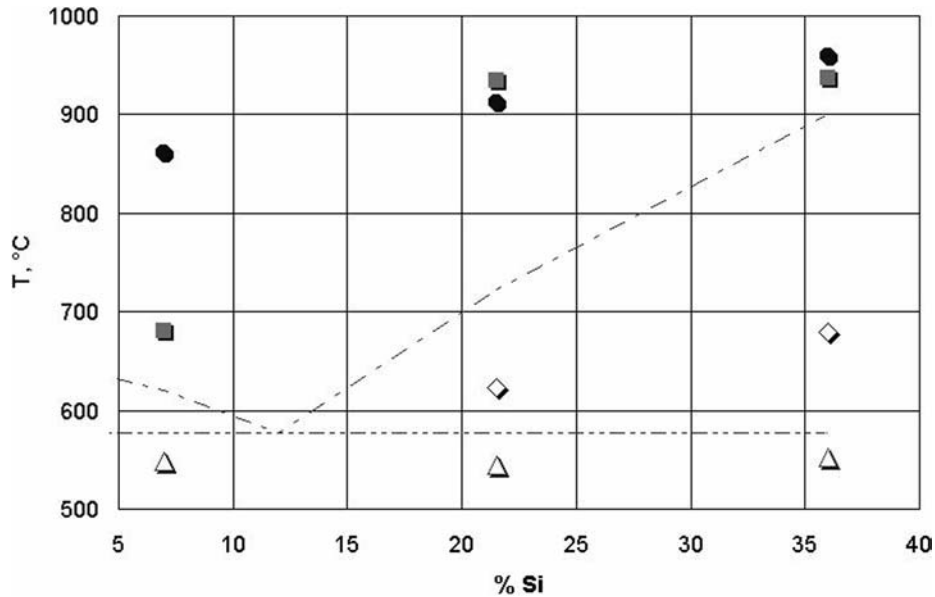
Figure 16 Thermal effects in three Al–Si alloys heated at 20 K/min to 1480°C and cooled at 10 K/min: transformations in the liquid phase (● — heating 20 K/min, ■ — cooling 10 K/min), eutectic temperature peaks (△), metastable peritectic (◇). Lines of the stable Al–Si diagram are also shown for comparison.

a correlation between basic mechanical properties (UTS and impact energy) and pouring temperature was revealed (Fig. 17). This dependence is not linear and cannot be described by a smooth function either — instead, it has clear peaks and valleys of properties in narrow temperature intervals (10–30°C). The level of properties (maximal UTS 205 MPa at 940°C, impact energy 3.9 J at 890°C) and their variations leave no doubts about a tight connection between microstructural transformations in the liquid state, metastable phase appearance, and decomposition and the final properties. Thus, "optimal" pouring temperature (taking into account industrial limitations and costs for very high overheating of the melt) of the Al–21.5% Si alloy was in the range of 845–860°C, which corresponds to the second viscosity anomaly in the Al–Si system discussed above.

It is worth nothing that these properties have been achieved without any grain refiners, master alloys, or modifiers — only by a simple optimization of the heating schedule. There is also a proof of finer microstructure of the optimally cast alloy in comparison with the non-optimized one (Fig. 18). These two photos present the microstructure of the alloy, which was heated and soaked according to the same temperature program, but was poured from two different temperatures, with the difference being only 12°C. The specimen with finer microstructure had a UTS of 150 MPa, impact energy of 3.3 J, hardness of 100 HB, and excess hexagonal crystal size of 15–30 μm, with the microhardness being 10–12 GPa. Another specimen had lower UTS (<140 MPa), impact fracture energy (2.65 J), and hardness (90 HB).

With this technique, the new technology of optimization of microstructure and phase composition as well as resulting properties has been developed [19,33] for
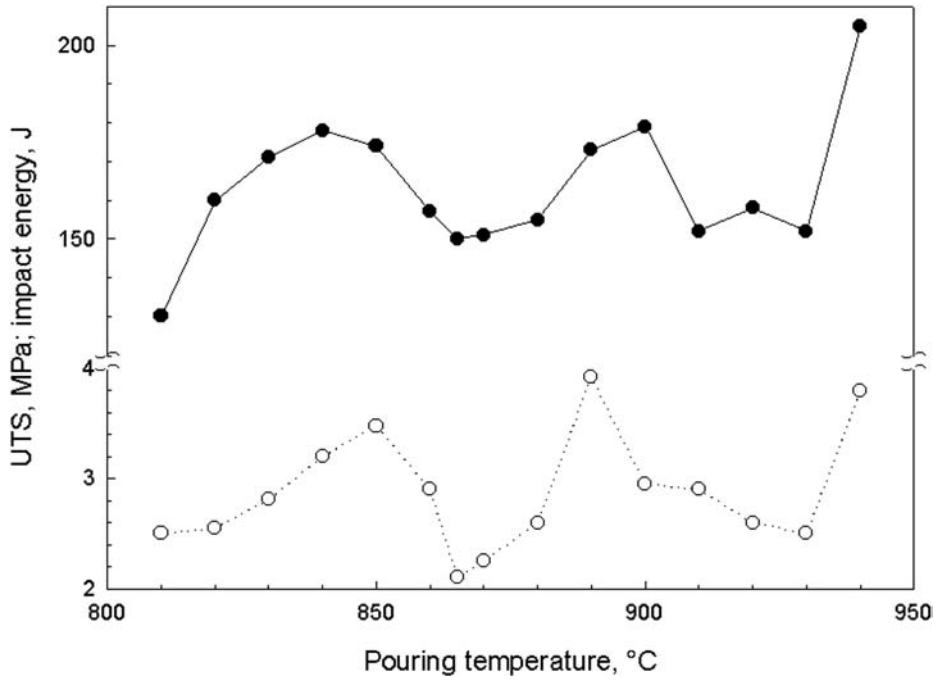
Figure 17 Mechanical properties of the Al–21.5% Si alloy, overheated to 920°C and poured from different temperatures: ● — ultimate tensile strength (MPa); ○ — impact energy (J).
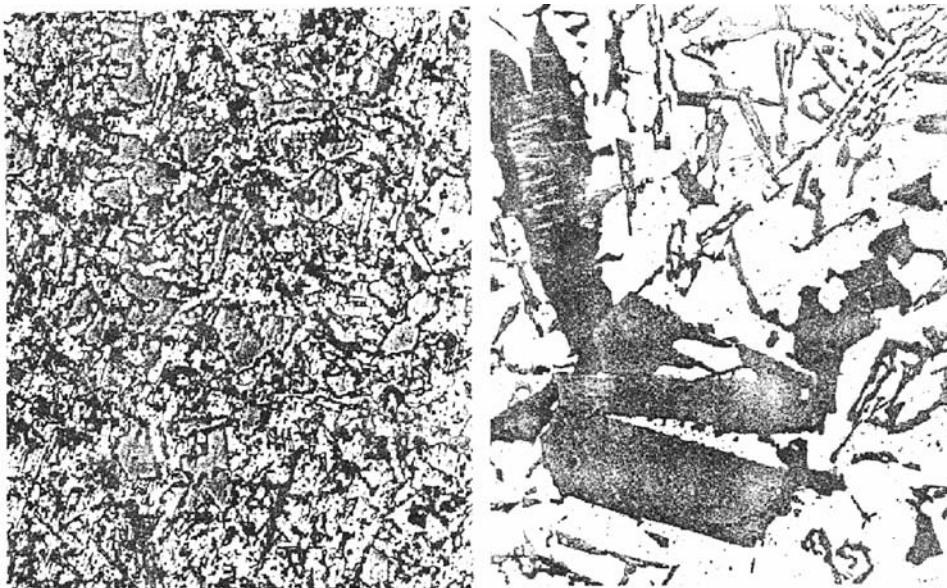
Figure 18 Microstructure of the Al–21.5% Si alloy, poured from two different temperatures, with the difference being only 12°C. Note large differences in microstructure and silicon crystals.

alloys of the Al–Si system. Therefore, use of master alloys or modifiers is not always necessary if the temperature and time parameters of the casting process can give the same or even better results with respect to grain sizes and mechanical properties.

6   CONCLUSIONS

The contents of this chapter may be summarized in the following statements or recommendations, which should be kept in mind for selection of the master alloys and performing the aluminum alloy modification and grain refinement. A good guideline rule tells us that there are no absolute guidelines, but all possible factors should be taken into account:

- Aluminum alloy history (processing, impurities, possible metallurgical heritage, charge composition, i.e. scrap content, etc.).
- Casting technology parameters (heating rate, overheating temperature, soaking time, pouring temperature, and soaking time at this point, crystallization rate at different parts of castings, etc.).
- The results of the microstructural, macrostructural, and physical investigation: are the alloy processing parameters optimal from the point of view of grain size, mechanical properties, foundry defects, etc.? If not, can temperature and time programs be adjusted, and what will the result be in that case?
- When the optimal processing parameters are fixed, the selection of modifying master alloy(s) may be done on the basis of the most desired effect (grain size refinement, DAS decrease, strength and elongation increase, hardness adjustment, etc.)
- The optimal amount of master alloy(s), introduction temperature and time, soaking, mixing, and pouring parameters have to be determined on the basis of target properties of castings. The formation of possible metastable phases and their influence on the morphology and microstructure of the final alloys should be taken into account. This requires more scientific effort to be applied to the problem before the operational window is set up for the casting process.

REFERENCES

1. L. Arnburg, L. Backerud, and H. Klang, "Production and Properties of Master Alloys of Al–Ti–B and Their Ability to Grain Refine Aluminum," Metals Technol. 1982, 1, pp. 1–17.

2. V. I. Napalkov and B. I. Bondarev, *Master Alloys for Aluminum And Magnesium Alloys*, Metallurgy Publ., Moscow, Russia, 1983, 160 pp. (in Russian).

3. U. Gruber, "Kornfeinnung von Aluminium und Aluminium-Gußlegierungen mit AlTi5B1 in Drahtform," Giesserei 1983, 7, pp. 209–211.

4. KB Alloys, Inc., Technical Bulletin, April 2000, Vol. 1, issue 4.

5. V. I. Nikitin, Heritage in Cast Alloys, Samara State Technical University, Samara, Russia, 1995, 248 pp. (in Russian).

6. T. B. Massalski, *Binary Alloy Phase Diagrams*, ASM, Metals Park, OH, 1986, 2224 pp.

7. N. P. Lyakishev, head ed., *Binary Phase Diagrams of Metal Systems*, Vol. 1, Mashinebuilding Publ., Moscow, 1996, pp. 112–244 (in Russian).

8. Y. N. Taran and V. I. Mazur, *The Structure of Eutectic Alloys*, Moscow, 1978, 312 pp. (in Russian).

9. L. F. Mondolfo, *Aluminium Alloys: Structure and Properties*, London, 1976, 971 pp.

10. H. Dieckmann and H. Arbenz, "Titanhaltige Aluminium–Kupfer–Gußlegierungen," Giesserei 1963, 50(12), pp. 372–378.

11. KB Alloys, Inc., Technical Bulletin, January 1999, Vol. 1, issue 1.

12. D. X. Tang, L. X. Wang, M. S. Zhao, H. Y. Lu, M. L. Zhang, and M. J. Cui, "The Rare Earth Aluminium Alloys in China," J. Rare Earth 1992, 10(1), pp. 66–71.

13. MetallurgAluminium Product Data Sheets, 2001 (available electronically via Internet at the address http://www.metallurgaluminium.com/).

14. M. I. Gasik, B. I. Emlin, N. S. Klimkovich, and S. I. Khitrik, *Electrical Smelting of Alumina-Silicates*, Metallurgy Publ., Moscow, 1971, 304 pp. (in Russian).

15. M. I. Gasik, N. P. Lyakishev, and B. I. Emlin, *Theory and Technology of Ferroalloys Production*, Metallurgy Publ., Moscow, 1988, 784 pp. (in Russian).

16. V. Z. Kutsova, "The Regularities of the Structure Formation and Phase Transformations in Al–Si Alloys," in: *Advanced Light Alloys and Composites*, Proc. NATO ASI Series, Vol. 59, 1998, pp. 29–34.

17. V. Z. Kutsova, O. V. Shvets, K. I. Uzlov, and A. G. Krimmel, "Effect of Sr on Phase Composition of Wrought Al–Cu Alloys," Light Alloys Compd. 1999, pp. 165–170.

18. V. P. Skripov, *Metastable Liquid*, Science Publ., Moscow, Russia, 1972, 312 pp. (in Russian).

19. A. V. Mazur, Principles of Hyper-Eutectic Al–Si Alloys Crystallization and Their Application in the Diesel Engine Piston Alloy Al–21% Si–2.5% Cu–2.5% Ni Technology, Ph.D. Thesis, National Metallurgical Academy of Ukraine, 1988, 131 pp. (in Russian).

20. B. H. Hu and H. Li, "Comparison of Effects of Master Alloys Containing Titanium and/or Boron on the Grain Size and Dendrite Arm Spacing of DIN226S Aluminium Alloy," J. Mater. Sci. 1997, 16, pp. 1750–1752.

21. V. Lavaert, M. Moors, E. Wettinck, R. Castañeda, and L. Goyos, "Evaluation of the Effect of Modifiers on the Mechanical Behavior and Shrinkage of Al7SiXMg Alloys," in: Proceed. Intern. Conf. ICCE/8 (D. Hui, ed.), Tenerife, Spain, 2001, Univ. New Orleans, pp. 523–525.

22. X. F. Liu, X. F. Bian, and J. J. Ma, "A New Grain Refining Technique for Pure Al by Addition of Molten AlTiB Master Alloys," Mater. Sci. Forum 2000, 331–337, pp. 385–390.

23. B. A. Balashov and G. G. Krushenko, "The Synthesis of the Composite Materials by Means of Interaction of Titanium Nitride Nanopowder with Aluminum Melts," in: Proc. Int. Conf. "Structure and Properties of Brittle and Quasiplastic Materials" , Riga, Latvia, 1994, p. 129.

24. G. G. Krushenko and S. P. Torshilova, "About Structural and Properties Anomalies in Aluminum," Dokl. Akad. Nauk SSSR, Rep. USSR Acad. Sci. 1984, 278(1), pp. 97–100.

25. A. V. Mazur, "Influence of Alloying by Iron and Oxygen on Kinetics of Isothermal Crystallization of the Al–22% Si Alloy," in: Proc. Intern. Conf. "Eutectica-V," Dnipropetrovsk, Ukraine, 2000, pp. 142–150.

26. K. Körber and K. Löhberg, "Einfluß von Phosphor und Natrium auf die Diffusion in flußigen Al–Si–Legierungen," Giesserei-Forsch. 1971, 23, p. 169.

27. M. Petrescu, "Liquid State Atomic Mobility of Silicon in the Unmodified Eutectic Silumin," Z. Metallkde. 1970, 61, pp. 14–18.

28. A. V. Mazur and V. I. Mazur, "Isothermal Solidification and Metastable Heterogeneous Equilibrium in the Al–21.5% Si Alloy," in: Proc. 2nd Intern. Conf. "Aluminum Alloys-90," Pergamon Press/CNPIEC, Beijing, P.R. China, 1990.

29. A. A. Popov, M. M. Gasik, and V. I. Freedman, "Nickel Powder Metallurgy (P/M) Superalloys with Isotropic and Gradient Carbide Reinforcement," J. Mater. Synth. Process. 1994, 2(3), pp. 143–150.

30. A. A. Popov and M. M. Gasik, "Nickel P/M Superalloys with Isotropic Carbide Reinforcement In Situ," Scr. Metall. Mater. 1995, 32(1), pp. 49–55.

31. A. A. Popov and M. M. Gasik, "Rapidly Solidified Fe–TiC Composites: Thermodynamics and the Peculiarities of Microstructure Formation In Situ," Scr. Mater. 1996, 35(5), pp. 629–634.

32. P. S. Popel and V. I. Nikitin, "Influence of Melt Structure on Silumines Crystallization," Melts (Rasplavy) 1987, 1(3), pp. 31–35 (in Russian).

33. V. I. Mazur, A. V. Mazur, and A. Y. Shportko, "About Metastable Phases in the Al–Si System," in: Proc. Intern. Conf. "Eutectica-V," Dnipropetrovsk, Ukraine, 2000, pp. 50–55.

34. A. V. Mazur and M. M. Gasik, Thermoanalytical Study of the Metastable Transformations in the Al–Si System, Report of the Academy of Finland No. 75199, Espoo, Finland, 2002, 6 pp.

35. N. Ryabova, Peculiarities of Phase Equilibria and Phase Transformations in Eutectic Al–Si Alloys, Ph.D. Thesis, National Metallurgical Academy of Ukraine, 1997, 113 pp. (in Russian).

36. E. V. Prikhodko, *Metallochemistry of the Complex Alloying*, Metallurgy Publ., Moscow, 1982, 184 pp. (in Russian).

4

Recycling of Aluminum

JORGE ALBERTO SOARES TENÓRIO and DENISE CROCCE ROMANO ESPINOSA

University of São Paulo, São Paulo, Brazil

1 SECONDARY ALUMINUM PRODUCTION

Aluminum is the most abundant metal in the earth crust [1]. Bauxite, which is the ore from which metallic aluminum is produced, is composed basically of aluminum hydroxide, iron oxide, hydroxides and silicates, carbonates of magnesium, calcium and iron, quartz, kaolin, titanium compounds and a small fraction of other metals; among the other metals are calcium, sodium, potassium, chromium, zirconium, phosphorus, gallium, and vanadium [2]. The aluminum hydroxide is the bauxite's main component, and also is the compound that makes the difference between bauxite and clay. Clays are composed essentially of hydrated aluminum silicates.

The main physical properties of aluminum appear in Table 1 [3]. Aluminum crystallizes in the face-centered cubic system; this metal is characterized by its high stretch, as indicated in Table 2 [3]. One can notice that the tensile strength and the yield point of pure aluminum are fairly low, which reduces the applicability of this metal in structural uses. However, aluminum alloys can achieve tensile strength over 600 MPa after heat treatment [3].

One can verify that the specific weight of the aluminum is about one-third of the specific weight of iron or copper. So, the resistance limit becomes a better value when mechanical components are compared.

Aluminum also has other characteristics that make it an important metal for various applications. Aluminum is the second most used metal in the world; some of the characteristics that make this metal so important are good corrosion resistance, low density, high electric conductivity, and good thermal conductivity.

Table 1 Main Physical Characteristics of Aluminum (From Ref. 3.)

Properties	
Crystalline structure	fcc
Atomic weight	26.98154
Specific mass at 20°C (g/cm^3)	2.6989
Solidification shrinkage (%)	6.5
Fusion temperature (°C)	660.4
Ebullition temperature (°C)	2494
Linear thermal dilation coefficient, from 20°C up to 400°C (°C × 10^6)	26.4
Specific heat at 25°C (J/kg °C)	900
Latent heat of fusion (kJ/kg)	397
Combustion heat (MJ/kg)	31.07
Volumetric electric conductivity (% IACS)	64.94
Thermal conductivity at 25°C (W/m °C)	247

Table 2 Main Mechanical Properties of Pure Aluminum (From Ref. 3.)

Property	Annealed	Hard-drawn (90%)
Tensile strength (MPa)	40–50	120–140
Yield point (MPa)	15–20	100–120
Brinell hardness (kgf/mm^2)	12–16	27
Stretch	50–70	8–12

Figure 1 [4] shows the main aluminum applications. The packing sector can be highlighted because this material is the one which present the shorter life cycle and, hence, it can be more rapidly recycled.

2 ALUMINUM RECYCLING: ENVIRONMENTAL AND ENERGETIC IMPACT

The secondary metallurgy of aluminum, as many other recycling processes, allows an economy of raw materials and energy. The production of one metric ton of aluminum from bauxite requires about 17,000 kWh of electricity, while the same amount of recycled aluminum consumes approximately 750 kWh. Therefore, the recycled aluminum substitutes primary aluminum with a gain of 95%.

Another positive aspect of aluminum recycling is the environmental impact. The production of one metric ton of primary aluminum requires about four metric tons of bauxite and it produces around two metric tons of red mud. Beside the generation of red mud, the production of primary aluminum also releases fluorides.

The mining of bauxite is also an activity that causes problems to the environment, because this activity devastates forests and needs space to dispose of the wastes.

Accordingly, the favorable energy balance makes the aluminum an excellent product to recycle, since the main cost of the production of primary aluminum is
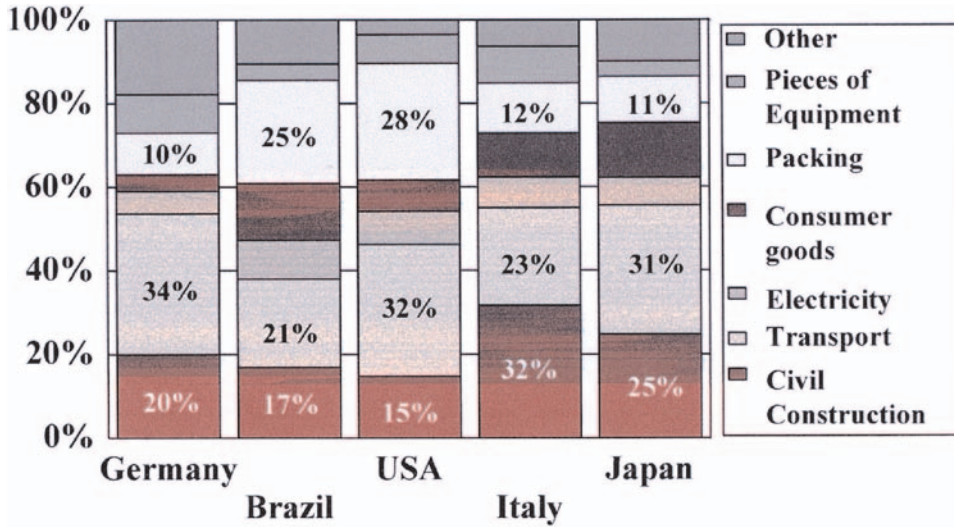
Figure 1 Aluminum consumption per sector. (From Ref. 4.)

electricity. Aluminum also presents another advantage, its scrap has a higher market aggregate value than many other materials used in packing.

For all these advantages, the recycling rate of aluminum cans has been increasing. Figure 2 [4] shows the evolution of the recycling rate of aluminum cans from 1991 to 1999. One can notice from Fig. 2 that the three countries that have the biggest recycling rate of aluminum cans are Japan, Brazil and the USA.
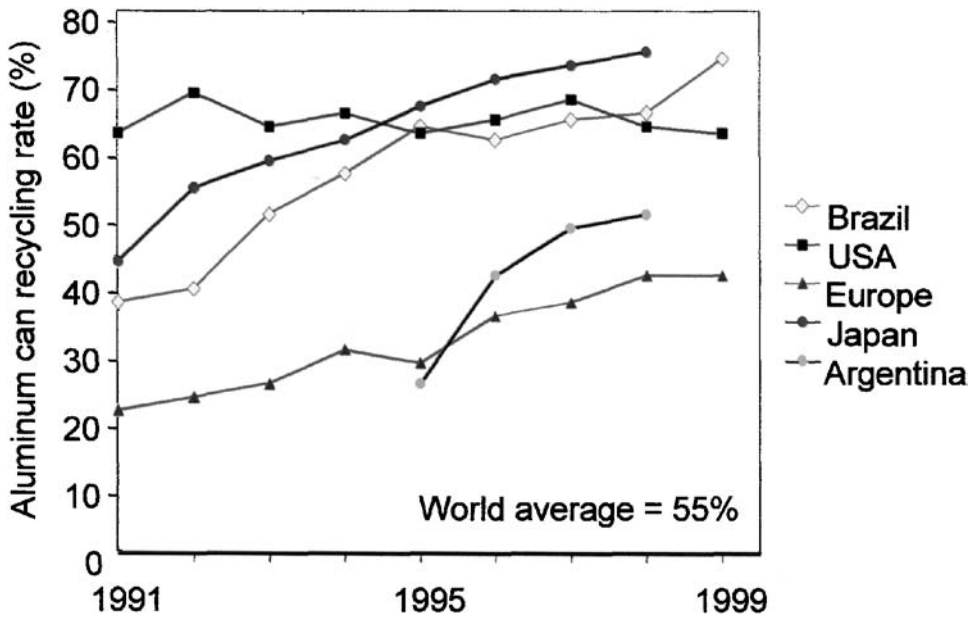
Figure 2 Evolution of the aluminum-can recycling rate in some countries. (From Ref. 4.)

Table 3 World Production of Secondary Aluminum (10^3 tons) (From Ref. 4.)

Continent	1996
Africa	37
North America	3,442
South America	183
Asia	1,285
Europe	1,822
Oceania	65
Total	6,834

Table 3 presents the data for the world production of secondary aluminum. North America is the leader of the world, and is responsible for around 50% of the world's secondary aluminum production. In fact, almost all the aluminum recycled in the North America is recycled in the USA (3,310,000 tons in 1996 [5]). Nevertheless, the USA data contemplate the so-called old and new scrap. In the USA, figures around 50–60% are for the new scrap [6,7]. New scrap indicates preconsumer sources.

The supply of old scrap (a product becomes an old scrap when it completes its useful product life) is affected by the amount of short-life products; so, used beverage cans (UBCs) are, in most cases, the main source of old scrap [7]. In the United States, UBCs account for more than one-half of the old scrap [8]. The greatest part of the UBCs is collected to produce new beverage cans. Most of the remaining old scrap is used to produce cast alloys for automotive or aerospace industry. Furthermore, the majority of secondary smelters is located near the automotive manufacturing areas [7,9,10]. All over the world, the prices of aluminum products are based on the values of the London Metal Exchange (LME).

The overall aluminum-recycling rate of different countries appears in Fig. 3 [4]. The world average is about 35%. Some countries such as Holland, USA, and England are above this average. European countries, in general, follow the world average; while less developed countries have usually a low aluminum-recycling rate.

3   ALUMINUM SECONDARY METALLURGY

Aluminum secondary metallurgy can be divided into two major groups: the melting of scraps, and the processing of white drosses.

Drosses are produced during the aluminum melting processes. Due to the high aluminum reactivity with oxygen, the formation of a superficial layer of oxide occurs. This layer becomes a physical barrier between the melt aluminum and the oxidant atmosphere, hence protecting the bath against oxidation. Turbulence generated by the handling of the liquid metal makes a new exposure of aluminum, which causes an increase in the oxide layer [11]. At the end of the melting process, this oxide layer is removed. During the removal of the oxide layer, some amount of aluminum is dragged along with the oxide; the material generated in this operation is composed
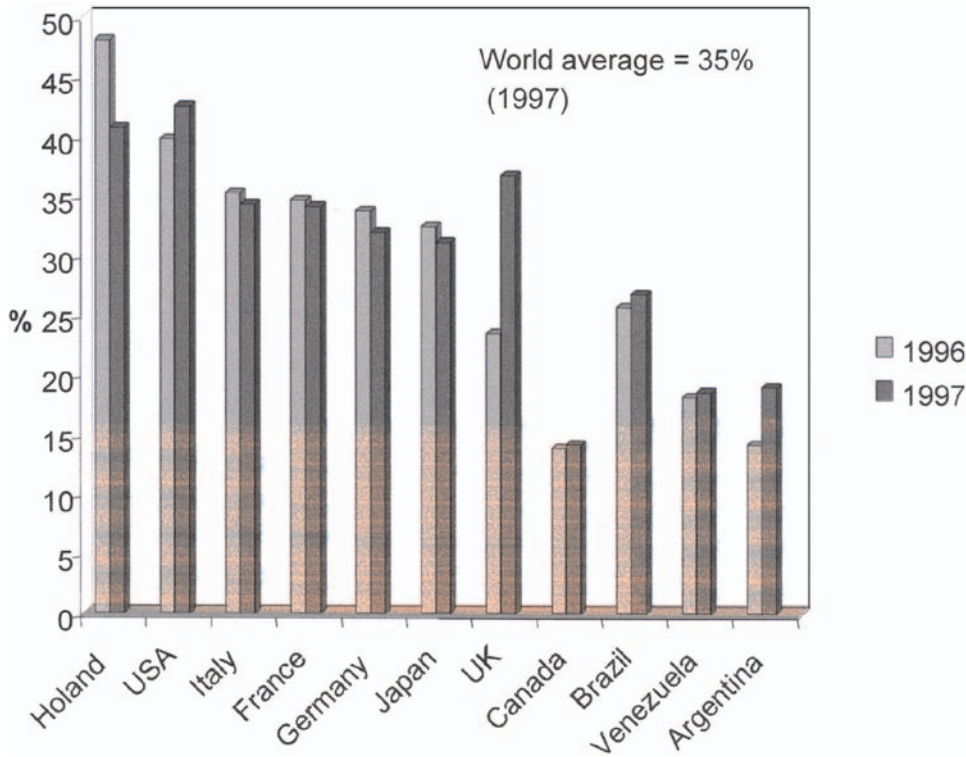
Figure 3 Overall aluminum-recycling rate per country. (From Ref. 4.)

basically of oxides, metallic aluminum and impurities that are found on the surface of the bath.

Oxidation is a superficial phenomenon; thus, the thickness of the raw materials to be melted influences the amount of oxide produced during the process, i.e., the thinner the raw material, the higher the oxide formation. This behavior occurs because the relation surface/volume becomes high when dealing with thin raw material.

Thus, the secondary metallurgy of aluminum aims to recover the aluminum of the scraps and drosses, as well as give an adequate destination to the sub-products of these operations.

3.1 High Temperature Oxidation of Aluminum

The oxidation process of aluminum alloys depends on the superficial condition of the sample, on the alloy composition and on the temperature. Some disagreement exists in the literature related with the oxides formed and also with the process kinetics [12–37]. During the initial stages of oxidation, aluminum and its alloys develop an amorphous, superficial, thin, and resistant layer (γ-Al_2O_3), which follows parabolic kinetics. This layer provides a barrier between the metal and the atmosphere, and the initial layer influences the oxide growth at high temperatures [12,13,28].

In the range of 425–450°C, the γ-Al$_2$O$_3$ amorphous coating suffers a discontinuous change of structure. A fast migration of oxygen through the oxide/metal interface happens at this range of temperature, providing the nucleation and growth of a γ-Al$_2$O$_3$ crystalline phase, under the initial amorphous layer [12].

However, the γ-Al$_2$O$_3$ amorphous layer continues to grow by cation diffusion to the oxide/gas interface, concurrent to the growth of the γ-Al$_2$O$_3$ crystals which are formed in the oxide/metal interface and grow to the interior of the metal after an incubation period [12].

The oxide films developed in aluminum alloys at high temperatures are characteristically coupled, composed of amorphous and crystalline structures [12,13,17].

Small Mg additions cause an increase of the γ-Al$_2$O$_3$ nucleation rate, which is responsible for the decrease to 400°C of the temperature formation of crystalline γ-Al$_2$O$_3$. Additions above 1 at.% Mg cause MgO crystal precipitation instead of crystalline γ-Al$_2$O$_3$ [17]. The MgO precipitation occurs initially at the oxide/metal interface. However, with the oxidation process advancement, the precipitation of MgAl$_2$O$_4$ may occur, which can be stable depending on the Mg concentration [17].

Smeltzer [16] performed a thermogravimetric study in an Al-3 wt% Mg alloy in the temperature range of 200–550°C. The surface condition effect on the oxidation rate was observed. Above 400°C, there was an increase in the oxidation rate. Between 450°C and 550°C, a parabolic behavior followed by a linear growth step was observed.

According to Hine and Guminski [37], the increase of Mg content in the alloy causes an increase in the oxidation rate.

Figure 4 refers to the oxidation kinetics of aluminum-can body alloy, Fig. 5 exhibits the aluminum-can lid alloy behavior and, in Fig. 6, a comparison between the results obtained for the two alloys in the temperatures of 500°C and 700°C is made.

The lid oxidation tests (Fig. 5) show different behaviors, according to the applied temperature. There is a linear behavior between 450°C and 500°C, suggesting a non-protecting characteristic for the oxides. Above 550°C, there is a parabolic behavior, evidencing the beginning of the protection by the superficial oxide [38].

The melting interval for the lid alloy ranges from 580°C to 637°C. Thus, in tests from 450°C to 550°C, the material was in the solid state, while in the 600°C test, the alloy was in the liquid + solid field, the other tests corresponding to the liquid condition. The test results for the bodies are shown in Fig. 4. The observed behavior is similar to the lid tests. The greatest difference is due to the deviation presented by the 800°C kinetic curve. The lid material oxidation kinetics are greater to the body. The 800°C curve exhibits a mixed behavior. The initial stage shows a parabolic progression, while the posterior shape (approximately after 10 min) indicates a change to a linear behavior. This change suggests loss in the initial protecting characteristic presented by the material (parabolic kinetics). This behavior was also found at 750°C, but for longer test periods [38].

Figure 6 shows greater kinetics for the lid alloy, in both temperatures. This result is related to the higher magnesium content [13,17,18,27,32,33]. This element has greater affinity for oxygen than for aluminum.

The largest practical interest result refers to the great deviation shown by the body alloy at 800°C oxidation. In fact, the body represents 75% of the
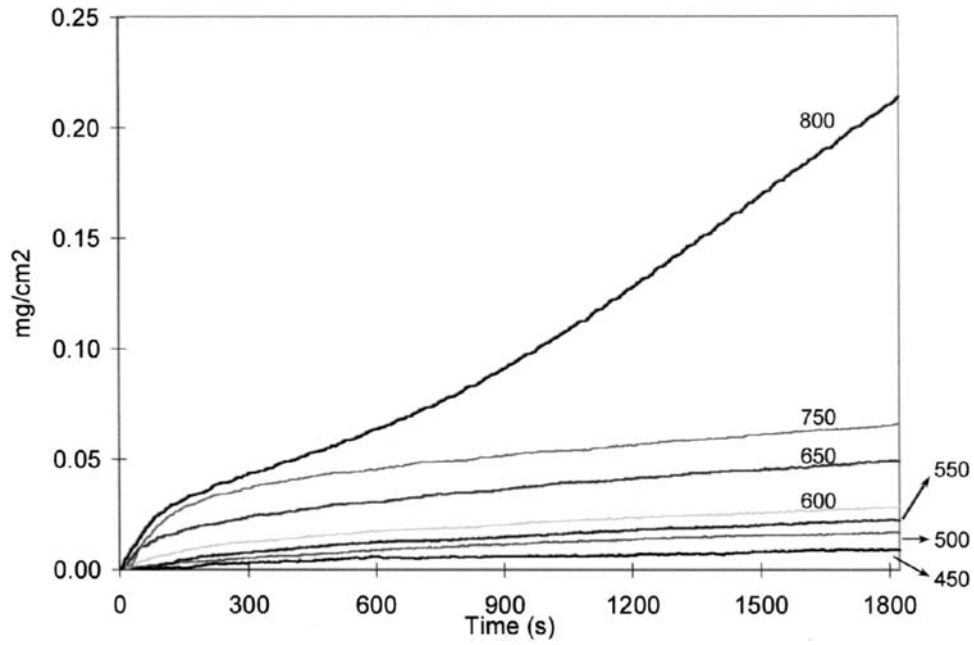
Figure 4 Alloy 3004 thermogravimetric tests. Atmosphere: air.
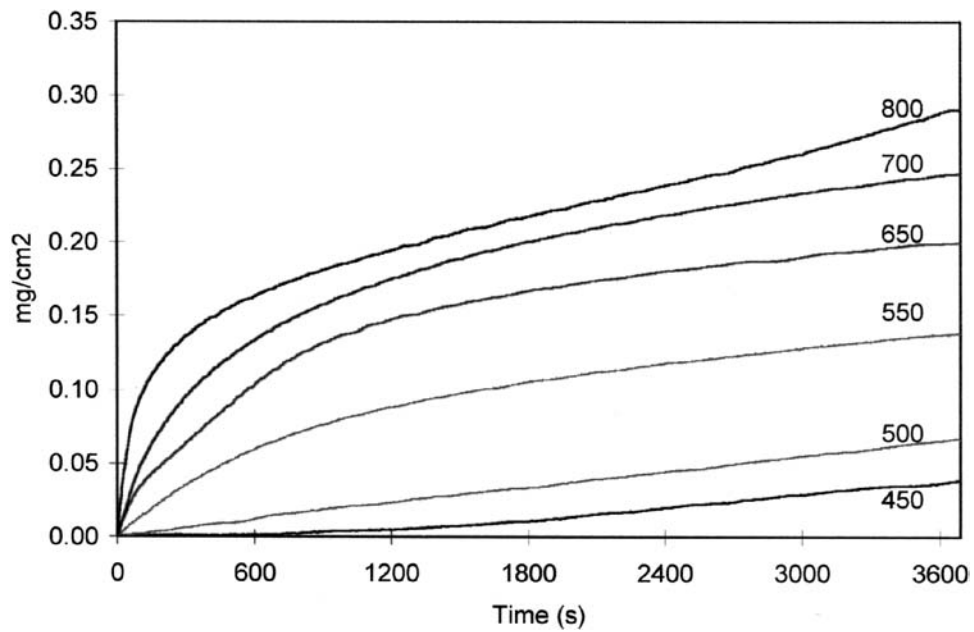
Figure 5 Alloy 5182 thermogravimetric tests. Atmosphere: air.
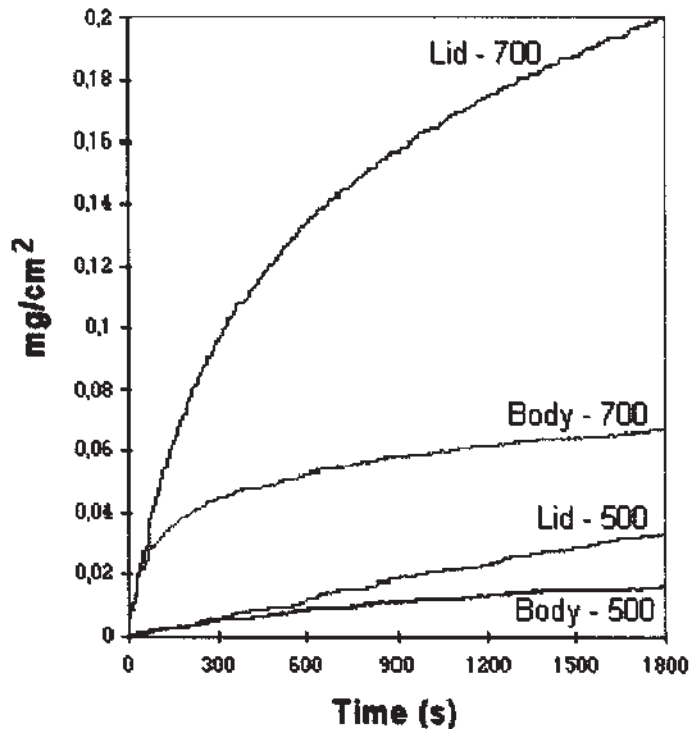
Figure 6 Comparison between the results obtained for the lid and body alloys in the temperatures of 500°C and 700°C.

aluminum-can mass, so it is considered that its behavior will dictate the general characteristics in the recycling process. The obtained data indicate that during the process of recycling of the considered alloys, the used temperatures should avoid permanence in the limits of 800°C.

According to Tenório and Espinosa [38], the oxidation products are constituted of MgO for the 5182 alloy and MgO· Al_2O_3 for the 3004 alloy.

The effect of the atmosphere composition is shown in Fig. 7 [38]. Ar + 5% O_2 simulate a weak rotary furnace oxidizing condition, while the atmospheres Ar + 1% O_2 and CO_2 simulate typical stoichiometric combustion conditions.

In order to analyze the atmosphere effect in the oxidation process, some values were extracted from the data which originated Fig. 7. These values are illustrated in Table 4. Comparison between the weight gain for different atmospheres is shown in Table 5.

The use of close stoichiometric atmospheres produces a 4–7 times reduction in the oxidation process. On the other hand, atmospheres with a small amount of excess oxygen produce quite the same oxidizing effect as air.

When stoichiometric mixtures are used, furnace atmospheres produced by burners, which use air, should not present high oxidation potential. Pieces of equipment, which use oxygen and operate with neutral atmospheres, can reach a significant aluminum loss reduction.
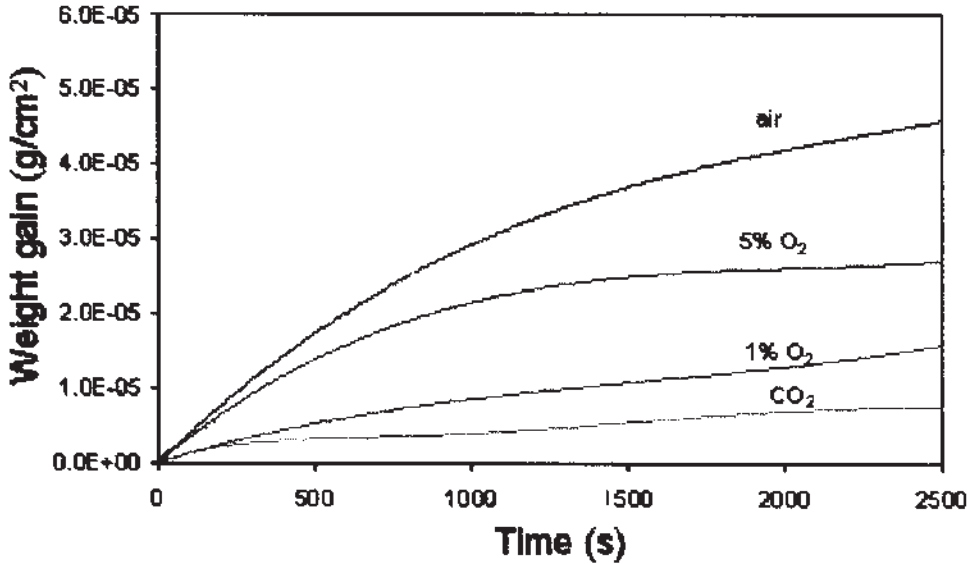
Figure 7 Atmosphere effect in the oxidation process for alloy 3004.

3.2 Scrap Melting

The melting of scraps is usually performed in rotary furnaces. Small foundries may also use crucible or reverberatory furnaces. Molten salt fluxes are often used in the melting of scraps. Typically, the volume of these fluxes is not higher than 2 wt% of the charge. The molten aluminum goes to the teeming or holding furnaces, and the white dross is stored in steel containers by the furnaces.

The volume of salt fluxes might be higher than 2 wt% of the charge when the raw material is composed mainly of light-gauge scraps or contains organic coatings such as aluminum chips, beverage cans or aluminum foils.

The range of temperature employed in the processes of aluminum melting is 700–800°C. The same range of temperature is used to recover aluminum from drosses.

Table 4 Weight Gains (mg/cm^2) from Different Oxidation Conditions

Time (s)	CO_2	1% O_2	5% O_2	Air
1000	4.07×10^{-6}	7.79×10^{-6}	2.03×10^{-5}	2.98×10^{-5}
1500	5.08×10^{-6}	9.50×10^{-6}	2.43×10^{-5}	3.56×10^{-5}

Table 5 Weight Gain Relation Among Different Oxidation Conditions

Time (s)	CO_2	1% O_2	5% O_2	Air
1000	7.32	3.83	1.47	1
1500	7.01	3.75	1.47	1

Table 6   Scrap Source and Its Main Impurities (From Ref. 11.)

Scrap	Physical impurities
Cables	Fe, Zn, Cu and organic material
Stamped material	Varnish, ink and organic material
Aluminum shapes	Fe, Cu, plastics
Aluminum sheets	Fe, oxides, earth
Aluminum foil	Oil, food remains, paper
Chips	Oil, Fe, Zn, Cu, Si
Blinds	Ink
Tubes	Fe, ink, plastic
Radiators	Fe, Sn, Pb, plastic
Engine block	Fe, Zn, Sb
Piston	Fe, Cr, oil, organic material
Beverage can	Fe, Cu, Mg, ink, varnish

Due to the higher formation of dross, the processes of recycling light-gauge aluminum scraps are analogous to the processes to recover aluminum from drosses.

Table 6 shows the main sources of aluminum scraps and the most common physical impurities associated with each kind of scrap.

3.3   Aluminum Recovery from Drosses

White drosses, as aforementioned, are a consequence of the process of aluminum melting. Typically, 15–25 kg of dross are produced per metric ton of molten aluminum [39]. The literature [11] separates drosses into three different classes:

- *White Dross*. It is from the primary metallurgy of aluminum, extrusion plants, lamination plants, and foundries. The main characteristics that differ this class of dross from the others are the absence of salt fluxes and the fact that this dross is usually whitish or grayish or presents a light color. The *white dross* may contain from 15% to 70% of metallic aluminum [39].
- *Black Dross*. It is produced in the secondary metallurgy of aluminum. This process uses salt fluxes and the produced dross presents dark colors.
- *Saltcake*. It is the result of the aluminum recovered from white and black drosses. It is characterized by the high concentration of salt and by its darkish color. This class of dross contains from 1% to 7% of metallic aluminum.

Actually, there are not much differences between the black dross and the saltcake.

3.3.1   Main Compounds of the Drosses

Drosses contain 5–60% of aluminum. Beside aluminum, drosses may contain other compounds. Stewart [40] suggested that the following phases might be present in drosses: Al_2O_3, AlN, AlC, MgF_2, $NaAlCl_4$, $KAlCl_4$, SiO_2, MgO, $MgO \cdot Al_2O_3$, and the salt constituents.

During the recovery of scraps that have Mg, Mg may react with fluorites forming the compound, MgF_2. It was suggested [39] that MgF_2 is a product of the reaction between the KF or NaF of the salt flux and the Mg of the alloy, because MgF_2 is thermodynamically more stable than KF and NaF. Thus, it is necessary to compensate this loss or use other types of salt flux.

Depending on the concentration of Mg in the alloy, the formation of $MgO \cdot Al_2O_3$ [41] in the dross is also possible, which turns the dross dark.

Unger and Beckmann [42] studied the treatment of drosses. They found the following compounds in the washed residue: Al_2O_3, $MgO \cdot Al_2O_3$, $(Ba, Sr, Ti)Al_2O_4$, $xAl_2O_3 \cdot ySiO_2$, AlN, $NaAl_{11}O_{17}$, Al_4C_3, SiO_2, $xNa_2O \cdot ySiO_2$, Na_4SiFe, CaF_2, $CaSO_4 \cdot 2H_2O$ and other components with concentrations lower than 1%.

Hryn et al. [43] analyzed the composition of the products from several plants that treat dross in the USA; the results are listed in Table 7 [43].

Aluminum hydroxide is not a natural product from the treatment of aluminum or drosses, but it is the result of the reaction between aluminum and alkaline water. The alkalinity of the water is due to the presence of AlN in the dross [43].

Aluminum recovery from drosses and light-gauge scraps with low thickness must be done under the condition of charge protection against oxygen. Thus, the processes can be divided into two classes:

1. Melting using salt fluxes.
2. Melting without salt fluxes.

The main by-product of these processes is a new type of dross. This dross has a low aluminum content, between 1 and 7 wt%, with or without salt depending on the melting process. The dross generated in the process that uses salt fluxes must be treated to remove the salt before the dross can be discharged. Next, the fundamental aspects of these two classes of melting processes are going to be described, as well as the main by-product destination.

3.4 Melting Using Salt Fluxes

The processes to recover aluminum from drosses or from light-gauge scraps are performed in rotary furnaces using salt fluxes. Typically, these furnaces have the capacity to process 1–10 metric tons of dross or scrap per cycle. It burns fuel as the

Table 7 Compounds Present in the Wastes from The Dross Treatment (From Ref. 43.)[a]

Compound	UBCs	2XXX e 7XXX	3XXX	5XXX	6XXX
Al_2O_3	+	+	+	+	+
Spinel	+	+	+	+	+
MgO	+	n.d.	n.d.	n.d.	n.d.
$Al(OH)_3$	+	+	+	+	+
Amorphous phase	+	+	+	+	+
AlN		−	−	−	−
Other phases			SiO_2		SiO_2
			CaF_2		CaF_2

[a] +, Large concentration; −, small concentration; and n.d., non-detected.

energy source; the selection of the fuel depends on the local availability of fuel. Rotary furnaces are used because they promote a constant agitation of the charge, enhancing the yield of the process [44].

The main function of the saline fluxes is to act as a barrier between the liquid aluminum and the oxygen of the atmosphere, diminishing the process of superficial oxidation of aluminum. The salt flux is added into the furnace at the beginning of the heating and the amount of salt flux varies between 20 and 65 wt% of the charge [42,45]. The fluxes should have the following characteristics:

- melting point below 720°C;
- low viscosity;
- easily detachable from the liquid bath;
- must not react with the metal;
- must not add impurities into the metal;
- must not be hygroscopic;
- low vapor pressure;
- low cost; and
- low treatment cost.

The main functions of the salt fluxes are the following:

- protect the molten metal from oxidation;
- help in the removal of the superficial oxide layer; this layer is formed previously or during the heating of the rotary furnace;
- promote the coalescence of aluminum drops; and
- maintain the oxides in suspension.

The salt fluxes are composed of NaCl and KCl mixtures; the most common composition is an equimolar mixture of these chlorides, because it constitutes the composition of the minimum point of the system. Besides these chlorides, fluorides are added to the fluxes. The main fluorides used are Na_3AlF_6, NaF, KF, CaF_2, and MgF_2 [41].

Figure 8 shows a sketch of the rotary furnace. The main advantages of the process using rotary furnaces and salt fluxes are the low-cost energy source (fossil fuel), the low initial investment, and the ease of operation.

The main oddity of this process is the production of a sub-product that contains salt and needs to be treated before the final disposal. Alternative processes for the use of the rotary furnace and salt flux aim to obtain comparable or even higher yields, but without the use of salt. In a process that does not use salt, the formed dross would not need treatment before the final disposal. These alternative processes also generate fewer fumes, as will be explained afterward.

3.5 Processing of the Saltcake

The dross from the processes that use salt flux needs to be treated before its final disposal. The soluble salt content is responsible for the necessity of treatment of this waste. The treatment aims primarily to diminish the content of soluble salt in order to meet the local environmental law. The saltcake presents the following composition:

- 60–63% salt;
- 7–8% Al; and
- insoluble material.
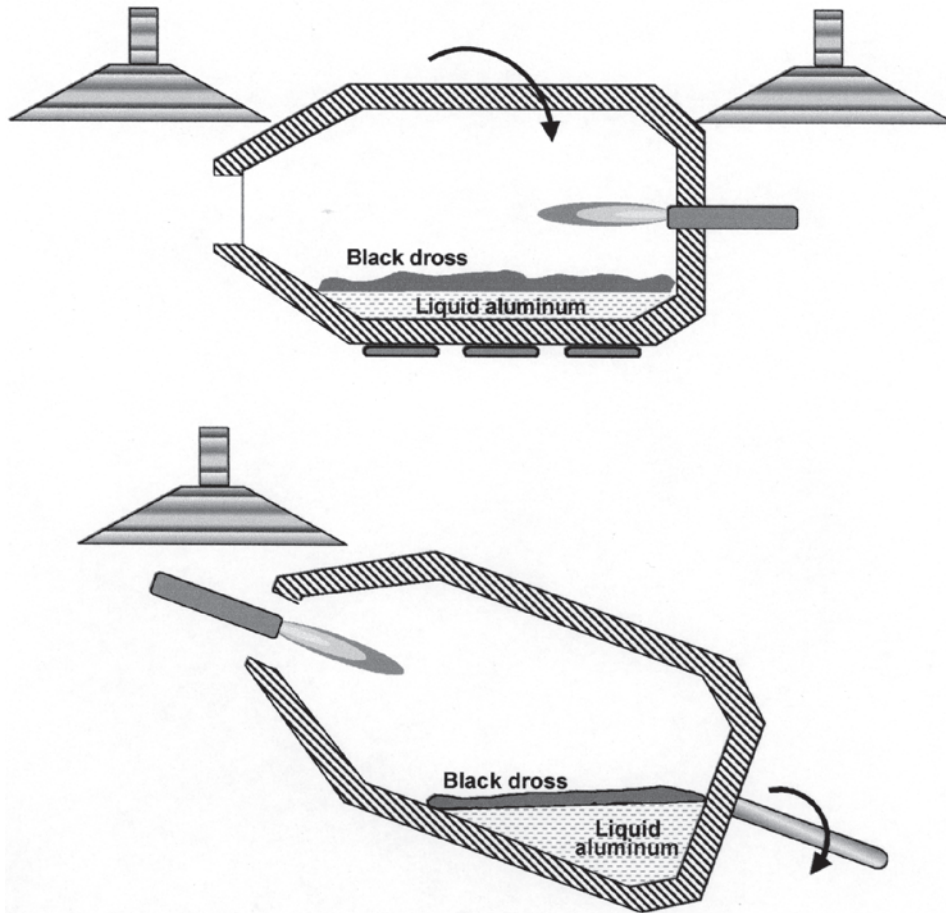
Figure 8 Sketch of the rotary furnace for aluminum recovery from drosses or light-gauge scraps.

Several steps compose the saltcake processing: milling, leaching with water, decantation, filtering, and drying of water for the salt recover [46–51]. During the leaching with water, a gas with bad smell is released. The gases that can be formed during this step are shown in the following [46,52]:

- Hydrogen: $Al + 3H_2O = Al(OH)_3 + \frac{3}{2}H_2$
- Ammonia: $AlN + 3H_2O = Al(OH)_3 + NH_3$
- Methane: $Al_4C_3 + 12H_2O = 4Al(OH)_3 + 3CH_4$
- Phosphine: $AlP + 3H_2O = Al(OH)_3 + PH_3$
- Sulfur hydroxide: $Al_2S_3 + 6H_2O = 2Al(OH)_3 + H_2S$
- Hydrochloric acid: $NaCl$ or $KCl + H_2O = NaOH$ or $KOH + HCl$
- Hydrofluoric acid: $NaF + H_2O = NaOH + HF$

Leaching is common in Europe and in other countries that have space restrictions. In the majority of the countries of the world, the saltcake is disposed of in landfills. The processing of the saltcake is predominantly performed by another

company, which treats the materials recovering the salt. The salt is sent back to the recycling aluminum plant.

As shown by the equations, the reactions produce some gases that need to be treated. Gases are cleaned through the use of activated carbon filters and also by the use of sulfuric acid that reacts with the NH_3 producing $NH_4(SO)_3$.

3.6 Fundamental Aspects of Aluminum Recovery Using Salt Flux

The main fundamental aspects highlighted in the literature related to the recovery of aluminum using salt fluxes are the following:

- phase diagrams of the chlorides and fluorides;
- metal/salt interfacial tension;
- coalescence power of the fluxes;
- interaction between aluminum and salt; and
- effect of composition on the yield.

3.6.1 Phase Diagrams Involving the Chlorides and Fluorides

NaCl–KCl System

The NaCl–KCl system is the base of the process that uses salt fluxes. The fusion temperature of the NaCl is 802°C and the KCl fusion temperature is 775°C. The system is isomorphic with a minimum point at the temperature of 667°C and the composition of 49.3 mol% of KCl, as shown in Fig. 9 [53]. The system also has an immiscibility line in the solid state.

The processes that melt drosses and light-gauge scraps generally use an equimolar mixture of NaCl and KCl, because this composition is near the minimum point of the NaCl–KCl system. This mixture melting point fulfills one of the requirements for a salt flux, which is, that the melting point must be under 700°C, the temperature compatible with the processes of aluminum melting.

NaCl–KCl–NaF System

Figure 10 [54] shows the ternary phase diagram, NaCl–KCl–NaF. NaF additions promote a lowering in the fusion temperatures of equimolar mixtures of NaCl and KCl until reaching a eutectic reaction.

NaCl–KCl–CaF₂ System

The ternary diagram, NaCl–KCl–CaF$_2$, shown in Fig. 11 [54], is fairly similar to the diagram shown in Fig. 10. The primary crystallization field is narrower.

Na₃AlF₆–CaF₂–Al₂O₃ System

The phase diagram, Na$_3$AlF$_6$–CaF$_2$–Al$_2$O$_3$ appears in Fig. 12 [55]. One can identify in this diagram a large solubility of alumina in calcium fluoride and cryolite. There is also an invariant eutectic reaction at 867°C.

3.6.2 Interfacial Tension

The effect of fluoride additions to equimolar NaCl–KCl salt fluxes on the interfacial tension between the salt flux and the liquid aluminum bath was first studied by Martin-Garin et al. [56]. Later on, other researchers [57–60] complemented these studies using a wider range of fluorides and also by verifying the effect of the Mg
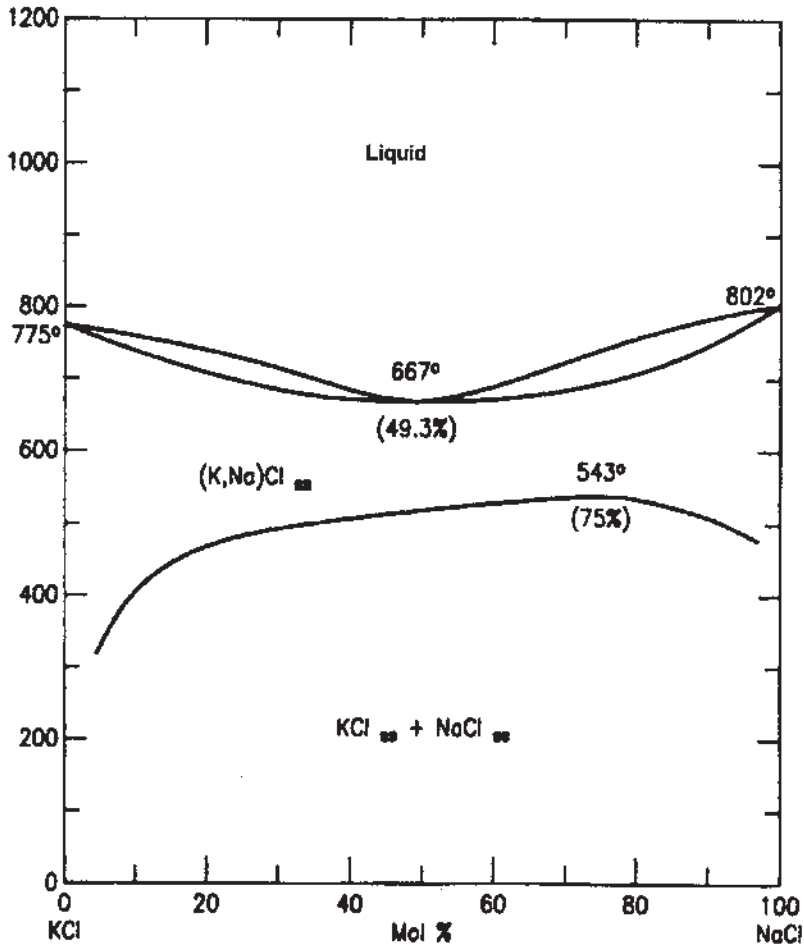
Figure 9 NaCl–KCl system. (From Ref. 53.)

content of the aluminum. Fluoride additions up to 10 mol% diminish the interfacial tension between the aluminum and the salt flux [56–59]. The fluoride efficiency was compared to each other and, were arranged in a crescent order of efficiency [58]:

$$Na_3AlF_6 < LiF < CaF_2 < MgF_2 < NaF < KF$$

Some of the results obtained by these researchers appear in Fig. 13.

According to Ho and Sahai [57,58], when adding fluorides to NaCl–KCl equimolar mixtures, NaF becomes more effective than KF for additions higher than 10 mol%.

The value of the interfacial tension between the pure liquid aluminum and the equimolar mixture NaCl–KCl at 740°C is 740 din/cm^2, while the value of the interfacial tension between the liquid alloy of the body of beverage cans and the equimolar mixture NaCl–KCl is 550 din/cm^2 [58]. The interfacial tension decrease is related to the increase in the activity of surface-active elements [61].
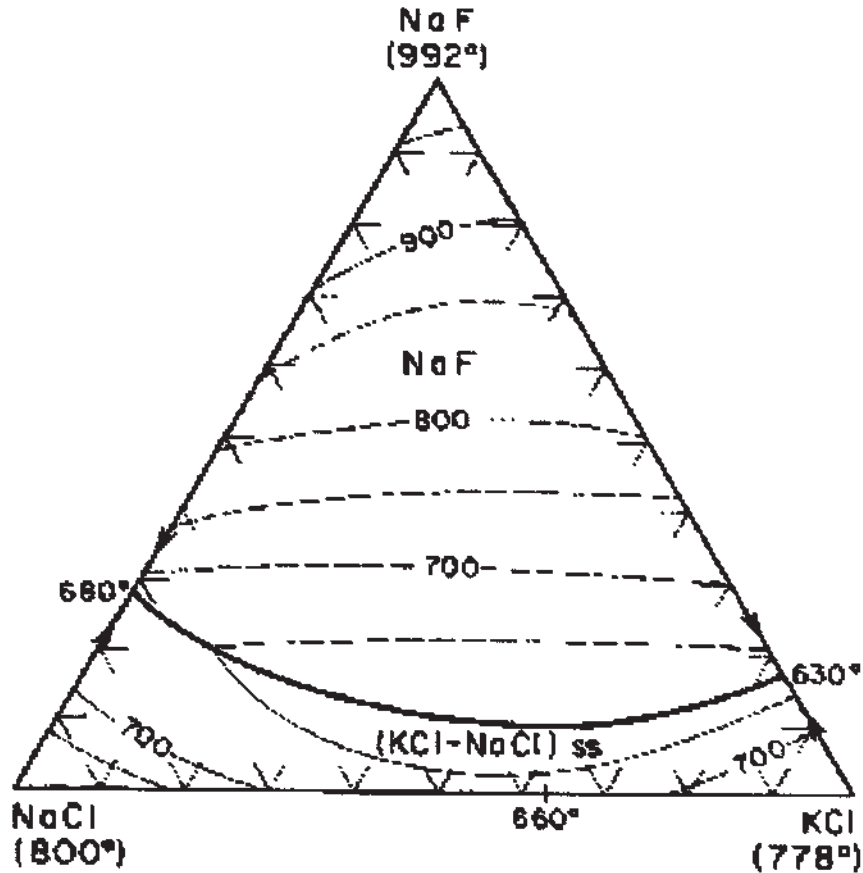
Figure 10   NaCl–KCl–NaF. (From Ref. 54.)

3.6.3   Aluminum Coalescence in the Salt Flux

When studying aluminum recycling, coalescence can be explained as the capacity of junction and growth of aluminum drops, firstly dispersing in the salt flux, forming bigger drops which have better conditions to be decanted into a pond under the flux. The literature [59,62] points to the fact that this coalescence is a consequence of the flux capability to remove the superficial layer of oxide.

Peterson [62] observed the behavior of aluminum fragments added to molten salt fluxes at 750°C. A transparent quartz crucible was used in these experiments and the aluminum fragments were of about 1–2 mm. Various kinds of fluorides and chlorides were added to an NaCl–KCl mixture. The NaCl–KCl mixture was done with the same weight of each salt. The results achieved are in Table 8 [62].

One can notice from Peterson results that the chlorides have little influence on the coalescence. On the other hand, in general, the fluorides effectively enhanced the coalescence. The fluorides, Na_3AlF_6, AlF_3, BaF_2, KF and SrF_2, stood out because they promoted the coalescence in less than 60 sec.

The impurities of the AlF_3 caused an increase in the time of the coalescence. The main impurity found in this compound was alumina. Thus, the presence of
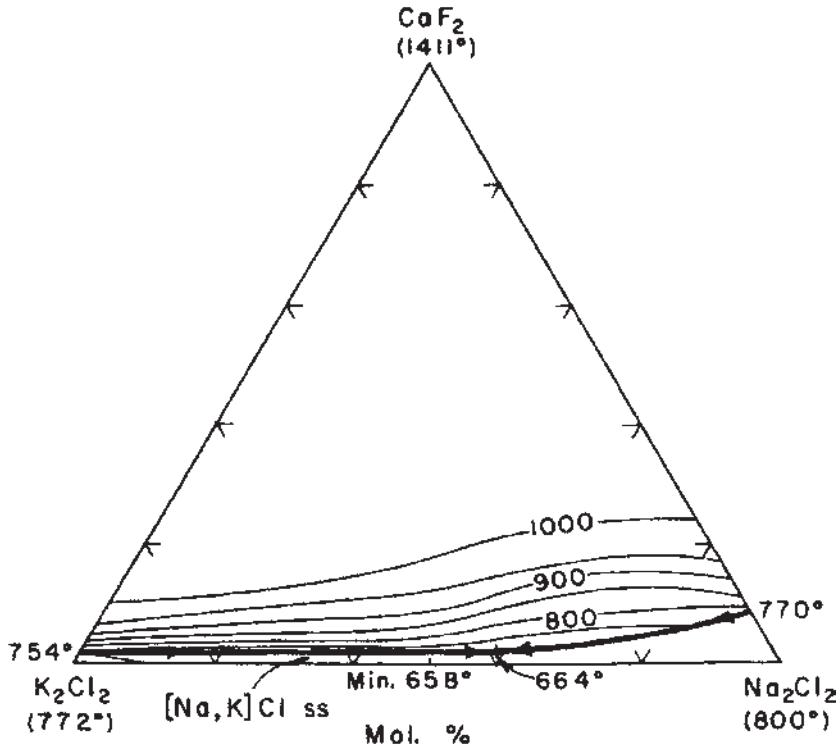
Figure 11 NaCl–KCl–CaF$_2$ ternary diagram. (From Ref. 54.)

alumina, even at low concentrations (0.65%), caused difficulties in the contact of aluminum drops, making the coalescence more difficult. Sully et al. [63] studied the effect of oxide concentration on the viscosity of molten salts. They observed that oxide additions higher than 10% decreased the aluminum coalescence.
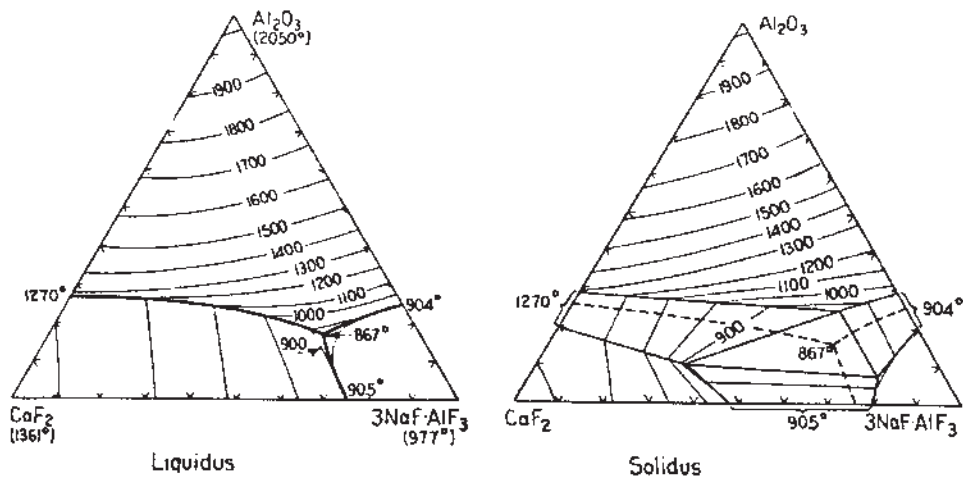
Figure 12 Na$_3$AlF$_6$–CaF$_2$–Al$_2$O$_3$ phase diagram. (From Ref. 55.)
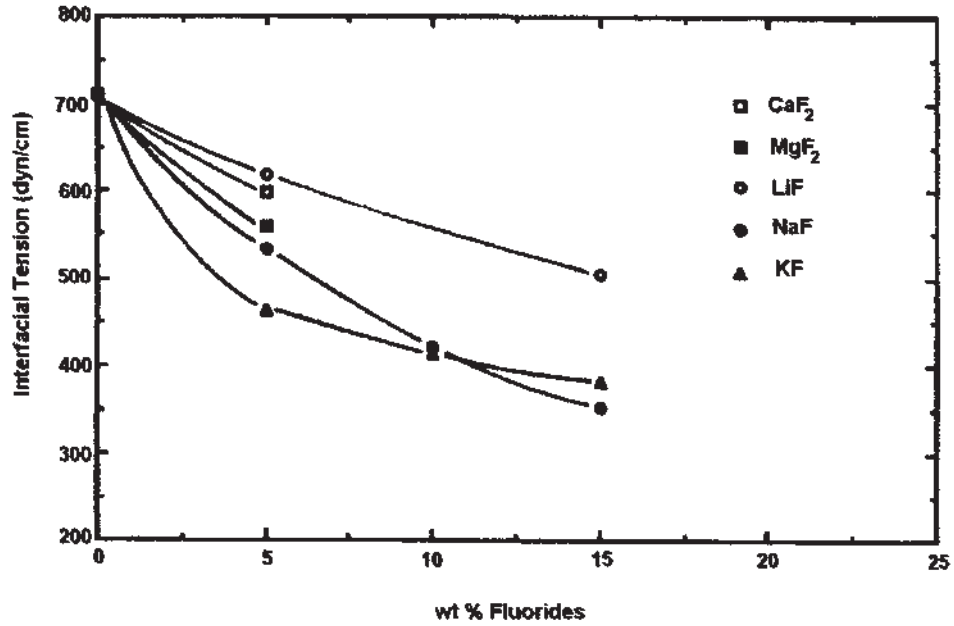
Figure 13 Effect of fluoride additions to equimolar NaCl–KCl mixtures on the interfacial tension salt/aluminum. (From Ref. 58.)

Table 8 Effect of Fluoride and Chloride Additions on the Coalescence of Aluminum in the Salt Flux (From Ref. 62.)

Additions (wt%)	Coalescence	Time for the coalescence (sec)
Cryolite	Complete	30
CaF_2	Moderate	222
NaF	Complete	> 270
AlF_3 (87% pure)	Moderate	> 60
$MgCl_2$	None	—
$CaCl_2$	None	—
MgF_2	Moderate	> 300
$NaAlF_4$	Almost complete	24
LiF	Complete	114
AlF_3	Almost complete	36
LiCl	Poor	> 600
$AlCl_3$	Poor	> 600
CaF_2	Moderate	216
KF	Complete	41
SrF_2	Almost complete	52
BaF_2	Almost complete	37

Figures 14 and 15 exhibit the curves for the variation of viscosity as a function of the fluoride addition to the equimolar mixture of NaCl and KCl at 760°C and 810°C, respectively [64].

Through the observation of Figs. 14 and 15, it is possible to conclude that the fluoride additions decrease the viscosity of the molten salt, and NaF presents approximately the same behavior as KF. The same behavior was observed by Roy et al. [65].

This decrease of the viscosity explains partially the results of Tenório and Delgado [66] who observed an increase in the aluminum recovery as the amount of fluoride increases in the molten salt. The decrease in the viscosity facilitates the movement of the small aluminum drops trapped within the salt layer, or in other words, promotes better conditions for aluminum coalescence.

The effect of interfacial tension seems to be less relevant in relation to aluminum recovery. Ho and Sahai [58] observed that Na_3AlF_6 exerts lesser influence on the interfacial tension than on other fluorides; on the other hand, Peterson [62] and also Tenório and Delgado [66] showed that Na_3AlF_6 causes an effective increase in the aluminum yield.

Although the literature presents little information about the solubility of alumina in molten chlorides [43,67], it is known that its solubility of the chlorides is much less than the solubility of molten fluorides; therefore, an increase in the solubility of alumina as the concentration of fluorides increases is expected.

Consequently, the result of fluoride additions can be considered as the sum of the capacity of attacking the oxide layer plus the decreasing of the viscosity.

The work of Ye and Sahai [68,69] describes tests performed in a way similar to the tests of Peterson [62]. The results for fluoride additions of 5 wt% to equimolar NaCl-KCl mixtures are shown in Fig. 16 [68]. One can notice that the additions of NaF, KF, LiF, and Na_3AlF_6 had excellent results [62].

These two studies presented conflicting results concerning NaF, LiF, and AlF_3. Additions of KF and Na_3AlF_6 presented excellent results in both the studies.

Ye and Sahai [68] also showed that the superficial condition of oxidation influences on the capacity of coalescence of the fluxes. Therefore, the equimolar NaCl-KCl flux could act as a real flux only for alloys with less than 1% of Mg. The flux made of only chlorides could not break the oxide layer of aluminum alloys containing more than 2% of Mg. Additions of only 1.5% of NaF were enough to promote the coalescence, breaking the oxide layer.

The salt flux, composed only of chlorides, was not capable of breaking the oxide layer of aluminum alloys that did not contain Mg when the oxide layer was grown thicker by heating the aluminum previously under air. Fluxes, with the addition of 1.5% of NaF, promoted the coalescence uniquely for alloys with less than 1% of Mg [68].

3.6.4 Salt/Oxide Interaction

The literature presents few data about the oxide solubility in molten salts [43,67]. Na_3AlF_6 can dissolve up to 12 wt% of Al_2O_3 at 955°C. NaF dissolves on the order of 5 wt% of Al_2O_3 at 985°C, and CaF_2 dissolves roughly 25 wt% of Al_2O_3 at 1275°C. NaCl additions to cryolite cause a strong decrease in the alumina solubility in cryolite. The solubility of alumina in chlorides is practically 0 and additions of NaF enhance the alumina solubility. Hence, not much can be affirmed concerning the
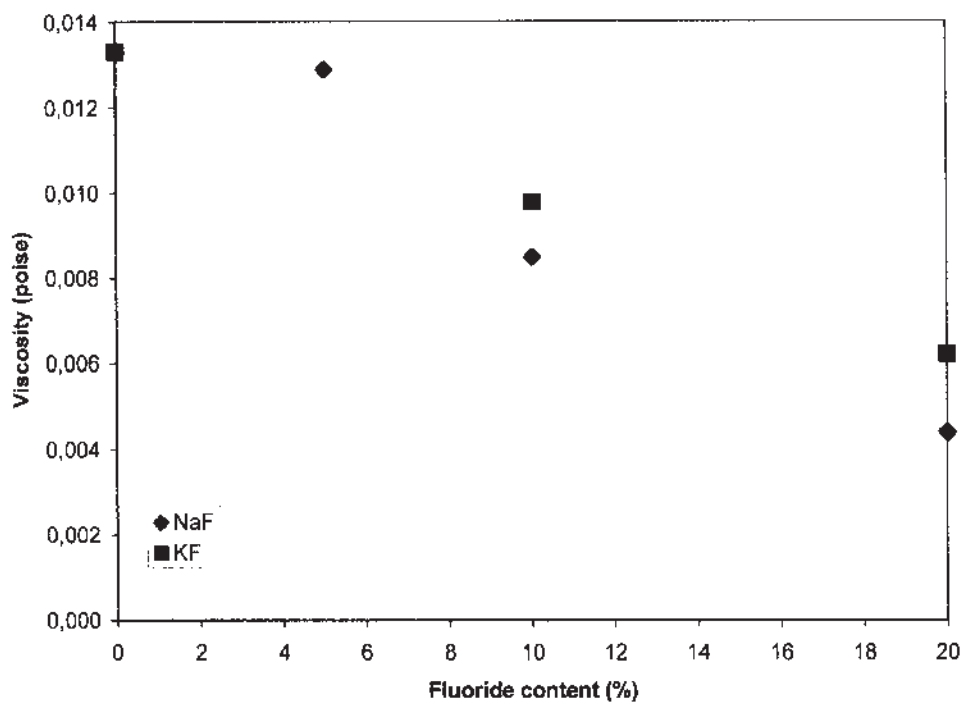
Figure 14 Variation of viscosity as a function of the fluoride additions to the equimolar mixture of NaCl and KCl at 760°C. (From Ref. 64.)
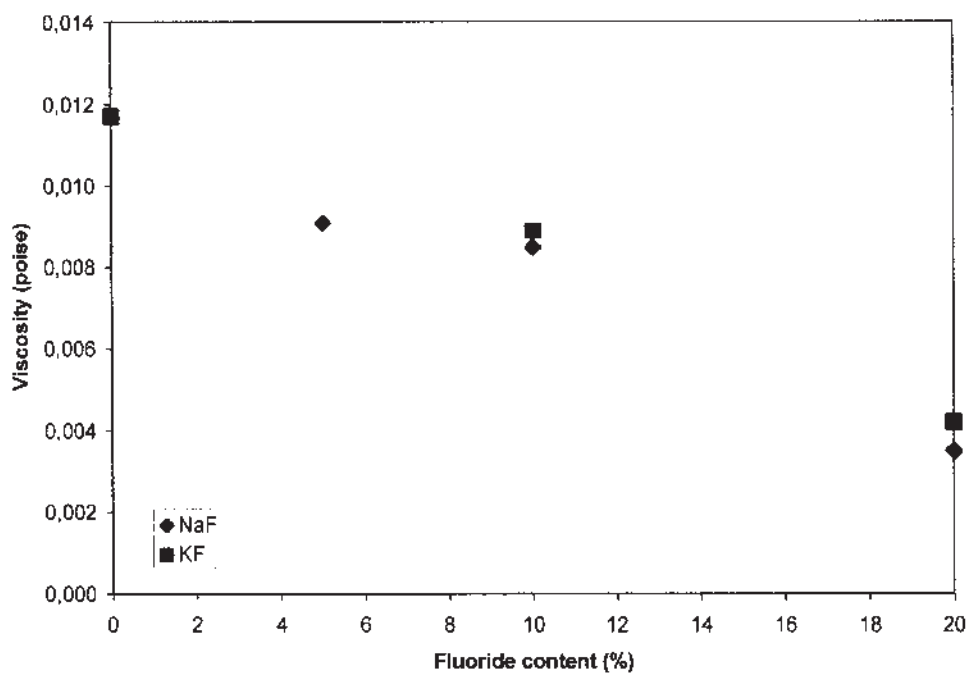
Figure 15 Variation of viscosity as a function of the fluoride additions to the equimolar mixture of NaCl and KCl at 810°C. (From Ref. 64.)
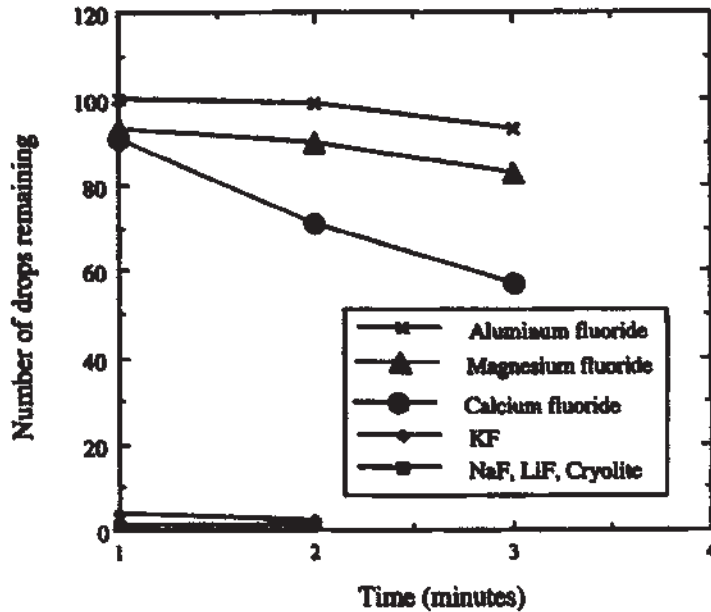
Figure 16 Effect of fluoride additions to salt fluxes on the aluminum coalescence. (From Ref. 68.)

effect of alumina solubility in the salt fluxes used in the aluminum-recycling processes.

Figures 17 and 18 [64] show the alumina dissolution in salt fluxes (P values) for the tests performed, respectively, using 2.5 wt% and 5 wt% of fluorides, and also present the behavior of the equimolar mixture of NaCl and KCl without any addition and also saturated in alumina. Figure 18 [64] shows the P values for the tests performed with different NaF concentrations.

According to Figs. 17–19, the addition of fluorides enhanced the alumina dissolution in the salt flux. The most effective was NaF, followed by CaF_2 and KF. As expected, the equimolar mixture of NaCl and KCl presented poor performance.

Even when the molten flux is saturated in alumina, the process of diameter reduction occurs, and this implies that dissolution is not the only effect observed. There is an effective diminution of diameter due to a corrosive attack process. This corrosive attack explains why the fluoride additions are so effective on the stripping of oxide layer, and stripping is exactly the step before coalescence.

The results of viscosity measurements and dissolution showed the complexity of the mechanisms involved in the recycling of aluminum in the rotary furnace using salt flux. Nevertheless, Ye and Sahai [68] showed only indirectly the action of fluorides. Thus, the increase in the recycling yield promoted by the additions of fluorides to the equimolar mixture of NaCl and KCl is attributed [64] to the simultaneous action of two factors: the decrease of flux viscosity that favors the aluminum coalescence and an increase in the corrosion attack of the oxide layer, which is related to the stripping of the oxide layer.
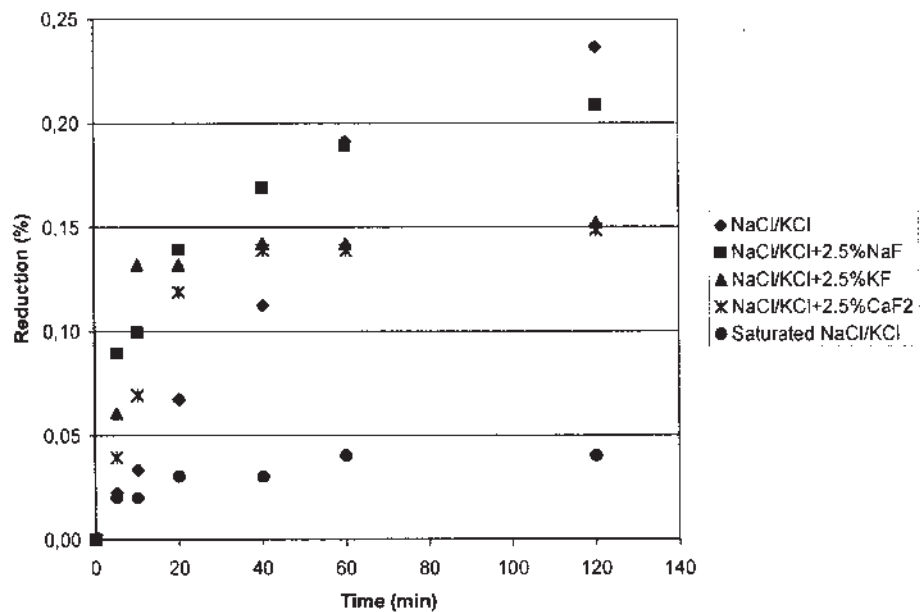
Figure 17 *P* values for the tests performed using 2.5 wt% of fluorides, and also present the behavior of the equimolar mixture of NaCl and KCl without any addition and also saturated in alumina. (From Ref. 64.)
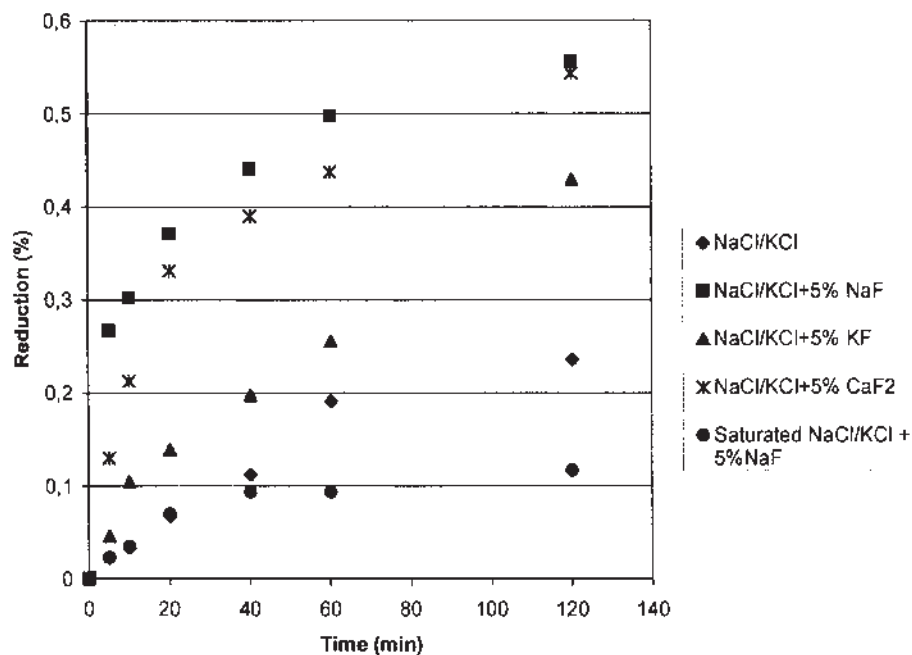
Figure 18 *P* values for the tests performed with 5 wt% of fluorides, and also present the behavior of the equimolar mixture of NaCl and KCl without any addition and also saturated in alumina. (From Ref. 64.)
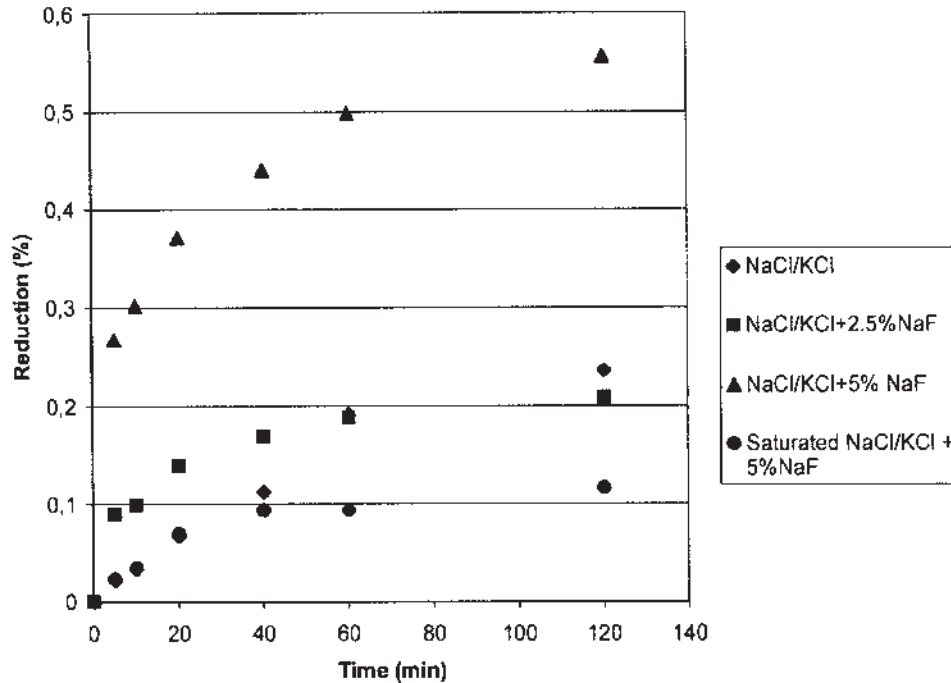
Figure 19 *P* values for different NaF concentrations in the equimolar mixture of NaCl and KCl. (From Ref. 64.)

Although it is believed that the salt fluxes promote the removal of the oxide layer, only few data are known about the mechanisms that rule this removal. West apud Ye and Sahai [68] suggested that the fluorides of the molten salts attack the external oxide layer, allowing the access of chloride ions to the interface metal/oxide. The removal of the oxide layer is attributed to the formation of aluminum chloride, which is volatile at the temperature of the recycling processes.

Sully et al. [63] suggested that the oxide layer removal is due to the interfacial tension strength between the liquid metal and the oxide layer. Later, this hypothesis was studied again by Ho and Sahai [58]. Ho and Sahai showed that with the available data, the relation $\gamma_{metal/salt} + \gamma_{oxide/salt} < \gamma_{metal/oxide}$ cannot be verified.

Van Linden and Stewart [45] suggested that the molten salt intensively wets the external oxide layer, which has its boundaries dissolved by the salt. This would lead to the disaggregation of the oxide layer.

Maason and Taghiei [70] proposed that the coalescence of metallic drops dispersed within the salt is influenced by the formation of a salt layer, which has a high melting temperature, on the interface between the molten alloy and the salt. The oxidation of the Mg present in the alloy is an important step for the formation of such salt layer. This layer was detected and identified as $KMgF_3$ or K_2NaAlF_6. Johnston and Peterson [41] showed that low yields could be caused by the effect of the Mg content of the alloy or by the high concentration of the spinel in the dross.
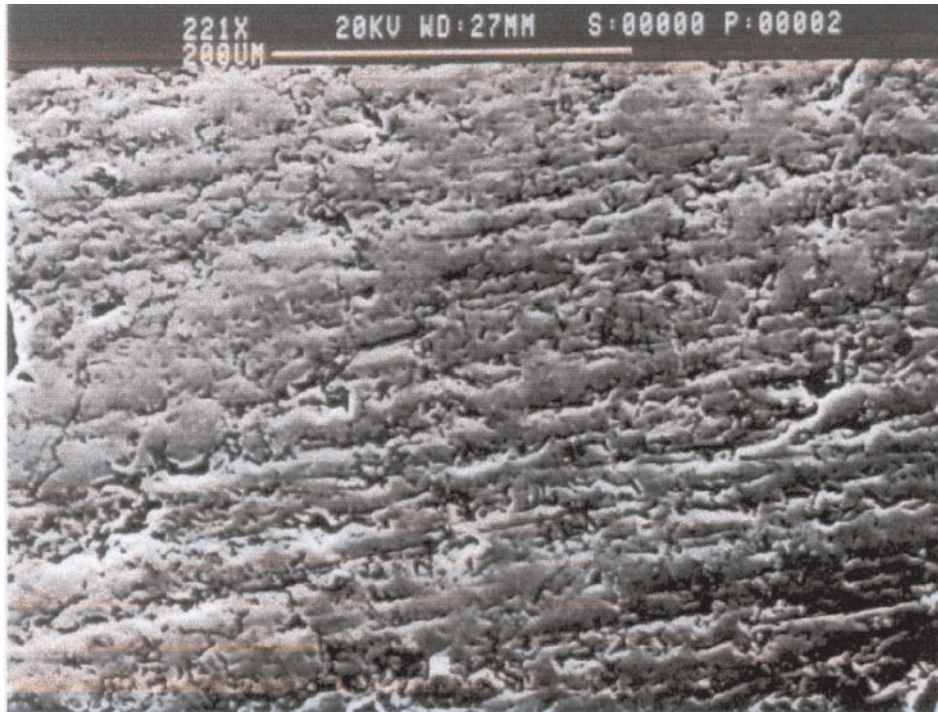
Figure 20 Back scattered image. Electron micrograph of the dross composed of oxide particles (dark) apparently in an aluminum matrix. (From Ref. 71.)

Based on the microscopic observation, Tenório and Espinosa [71] concluded that the drosses are essentially composed of oxides and trapped aluminum. Figure 20 [71] shows a typical microstructure of dross from a melting process where no salt was used. Clusters of grayish particles and an apparently continuous bright phase can be observed. The grayish phase corresponds to the oxide and the light phase to the aluminum. In this case, the oxide is mainly composed of $MgO \cdot Al_2O_3$ because the samples were obtained to form an aluminum-can recycling facility. XRD also confirmed the composition of the oxide.

The amount of aluminum in Fig. 20 could give the impression of being higher than the amount of oxides; however, this might be a false appearance. During the process of cutting and metallographic preparation, the aluminum may deform and spread. On the other hand, the oxides might break, because oxides are brittle phases.

After the selective etch to remove the aluminum, an apparent continuous net of oxides was observed, as shown in Fig. 21 [71]. The oxides exhibit the feature of slightly rough plates; consequently, this structure has a significant amount of specific surface area, which promotes the aluminum entrapment. The oxides in the dross remain solid at the standard temperatures of aluminum foundry, and the oxides have high strength at such temperatures. Hence, these two facts plus the high surface area explain the higher metal losses that occur in the aluminum recycling compared to other metals.

Studying the interaction between the molten salt and the dross at high temperature (700°C), Tenório et al. [64] observed that the liquid salt breaks the oxide
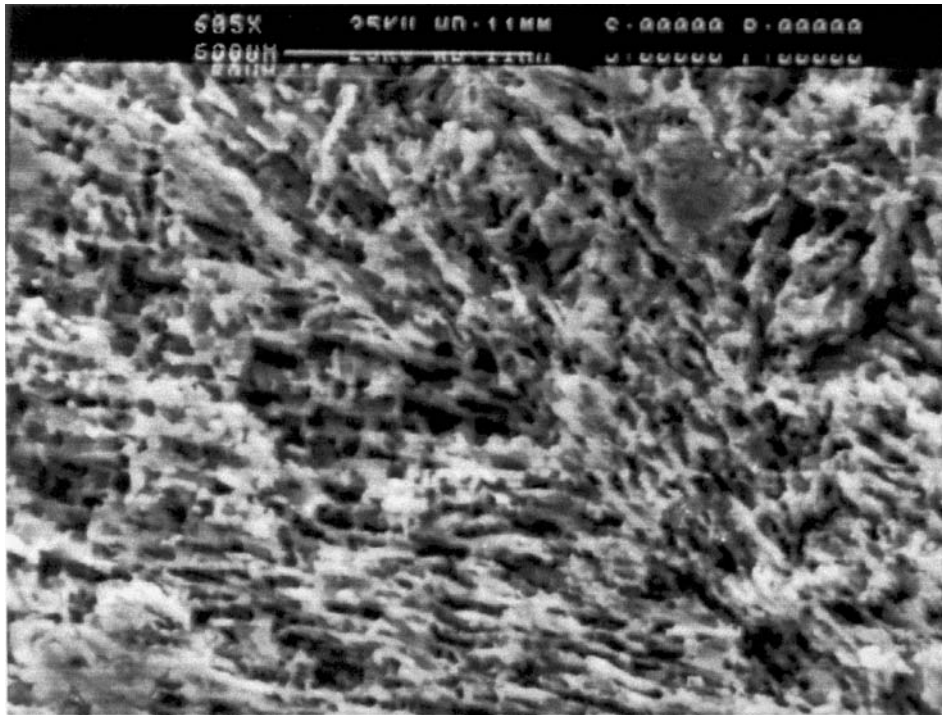
Figure 21 Back scattered image. Electron micrograph of the dross. After the selective dissolution of aluminum, it is possible to notice that the oxide particles are united in a net. (From Ref. 71.)

structure, and transforms this net into a great number of small fragments that turn into the shape of plates or small clusters of plates. Figure 22 [71] presents the microstructure of the oxides after the interaction with the liquid salt. In Fig. 22, all the aluminum was released from the dross due to the salt action. In this specific case, the salt was removed before the SEM observation by dissolution in water.

This mechanism of salt etch is favored by the low interfacial tension between the salt and oxide and also between the aluminum and salt. Fluoride additions decrease the interfacial tension between molten salt and aluminum [64], and consequently, this favors the overall process of the oxide net breaking. The viscosity of the molten flux can also influence this mechanism [64], once the salt has to percolate the dross and allow the separation of aluminum from the broken oxide particles.

The aluminum freed from the oxide entrapment started to coalesce, generating small drops of about 1 mm, which began to cross the viscous salt plus oxide layer. Figure 23 [71] shows the formation of such small drops in the inferior part of the dross layer.

Therefore, the molten flux breaks the long oxide nets and promotes the oxide fragmentation, as well as liberates the aluminum drops that have a propensity to coalesce.

Figure 24 [71] shows schematically the suggested process of oxide removal from the molten aluminum surface. Initially, the salt corrodes the oxide grain boundaries

Figure 22 Back scattered image. Electron micrograph showing the oxide net fragmented after the molten salt action. (From Ref. 71.)

(Fig. 24b). According to Tenório et al. [64], chloride salt flux has the capacity to corrode oxides even under saturated conditions, and this ability is increased by fluoride additions. The molten salt accumulates near or at the oxide aluminum interface (Fig. 24c), and causes the spalling of the scale (Fig. 24d). Finally, aluminum drops are generated, which keep distributed in the inner portion of the salt layer (Fig. 24e).

This behavior is similar to the hot corrosion process where a metal that develops a protective oxide layer is submitted to an atmosphere containing chlorides. Under these conditions, chlorine or chloride penetrates the scale and an increase in the chlorine concentration occurs near or at the interface between the oxide and the metal. This accumulation of chlorides causes a weakness in the mechanical stability of the oxide layer, which becomes susceptible to blistering, cracking or spalling [72]. Thus, a similar mechanism may be acting in the interaction between molten salt and oxides.

3.6.5 Effect of the Flux Composition on the Aluminum Recovery

As previously reported, salt additions make a remarkable effect on the yield of aluminum recovery especially in rotary furnaces. These fluxes typically contain 95–100% of an equimolar NaCl/KCl mixture with fluoride additions.

The effect of the equimolar mixture of NaCl/KCl on the aluminum recovery is shown in Fig. 25 [73]. Only a small amount of salt in the charge is enough and the recovery can reach an 85% yield level.
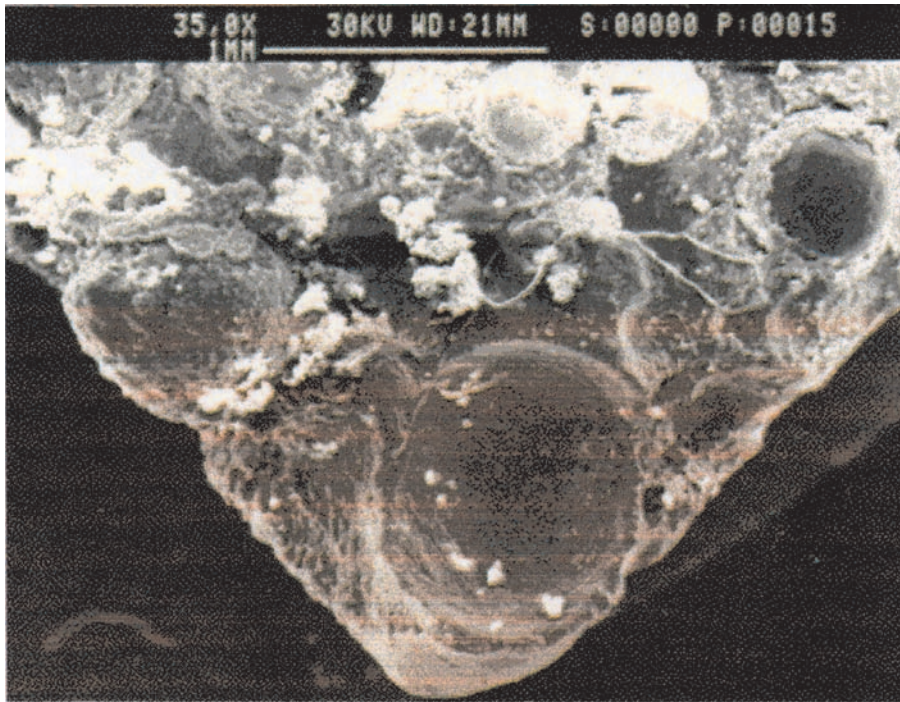
Figure 23 Back scattered image. Electron micrograph of aluminum drop generation subsequent to the introduction of the molten salt. (From Ref. 71.)

Van Linden and Stewart [45] showed that the metal recovery is more efficient with the increase of fluoride concentrations up to 10 wt%. Above this concentration, the yield decreases for CaF_2 or MgF_2 additions, and remains constant for NaF and KF.

Johnston and Peterson [41] studied the effect of magnesium in the fluxing process. It was found that magnesium promotes the formation of spinel $MgO \cdot Al_2O_3$ or $MgAlO_4$ in the dross, with a black color. It was also suggested that poor recoveries might be an effect of the magnesium concentration in the alloy or high concentration of spinel in the dross.

Tenorio and Delgado [66] studied the effect of fluoride additions to the equimolar NaCl/KCl mixture. Figures 26, 27, and 28 show, respectively, the influence of KF, CaF_2 and Na_3AlF_6 on the aluminum recovery.

An increase of KF in the flux increases the yield up to 95%. The yield observed for CaF_2 and Na_3AlF_6 additions are approximately 92% and 94%, respectively. Consequently, additions of 2.5 wt% of fluorides in the flux are sufficient to increase the yield by 5% at least, and also, the fluoride efficiency grows up in the following sequence: $CaF_2 < NaF < Na_3AlF_6 < KF$ [66].

3.7 Melting Without Salt Fluxes

Due to increasing environmental restrictions related to the disposal of the drosses that contain salt, since the end of the 1980s, alternative processes to recycle

a) Contact of molten salt and the oxide layer

b) Corrosion of oxide at the boundaries

c) Diffusion of Cl to the oxide/aluminum interface

d) Oxide detachment to the salt

e) Aluminum drops
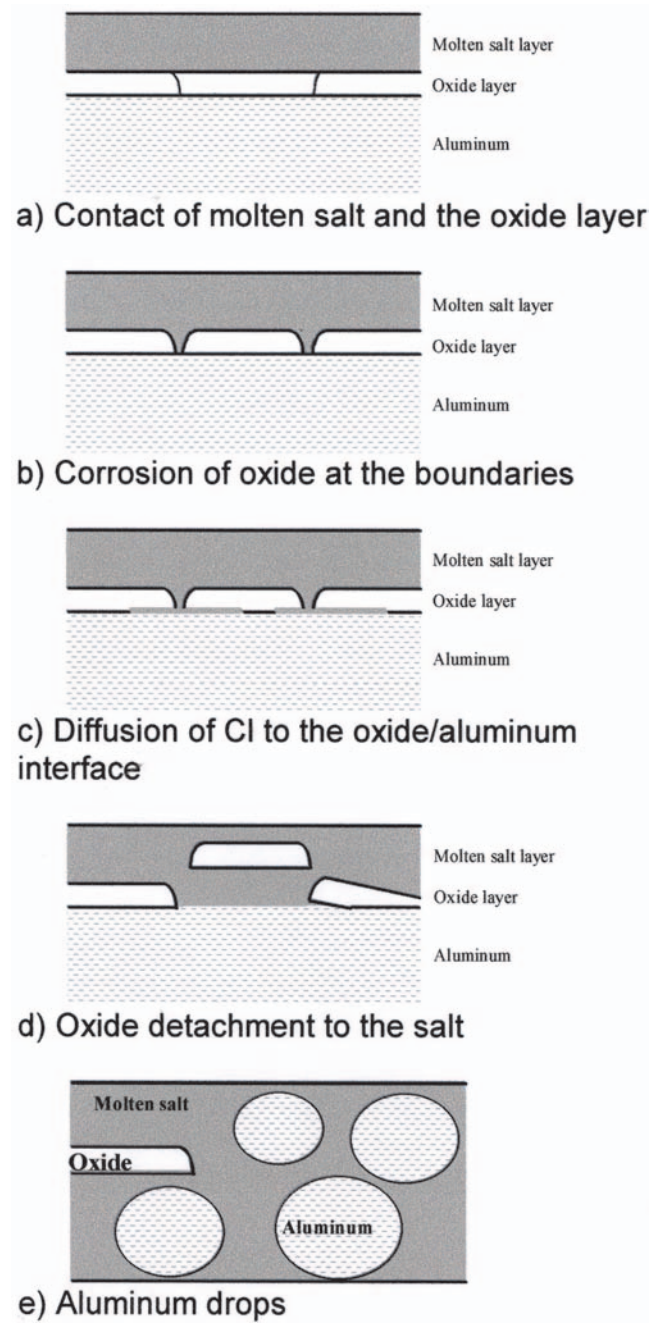
Figure 24 Sketch of the oxide detachment caused by the molten salt. (From Ref. 71.)
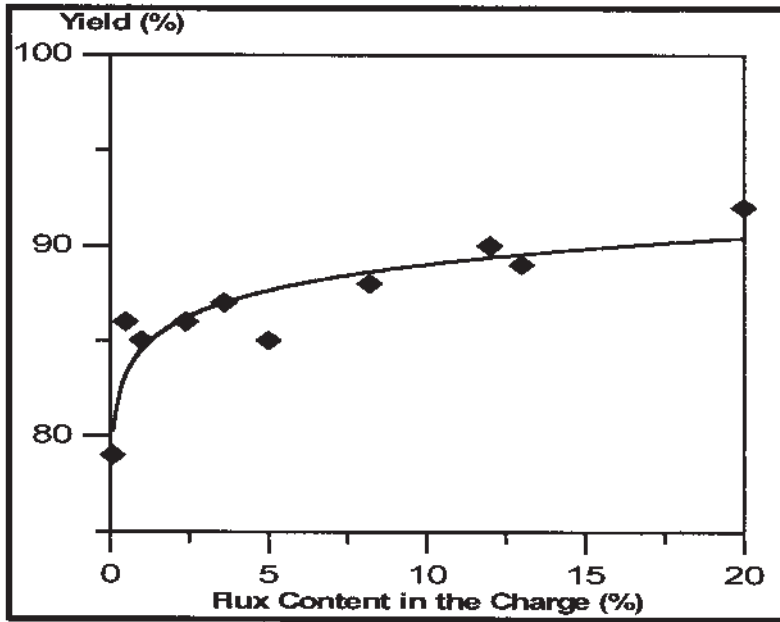
Figure 25 Effect of the NaCl/KCl equimolar flux in the aluminum recovery. (From Ref. 73.)
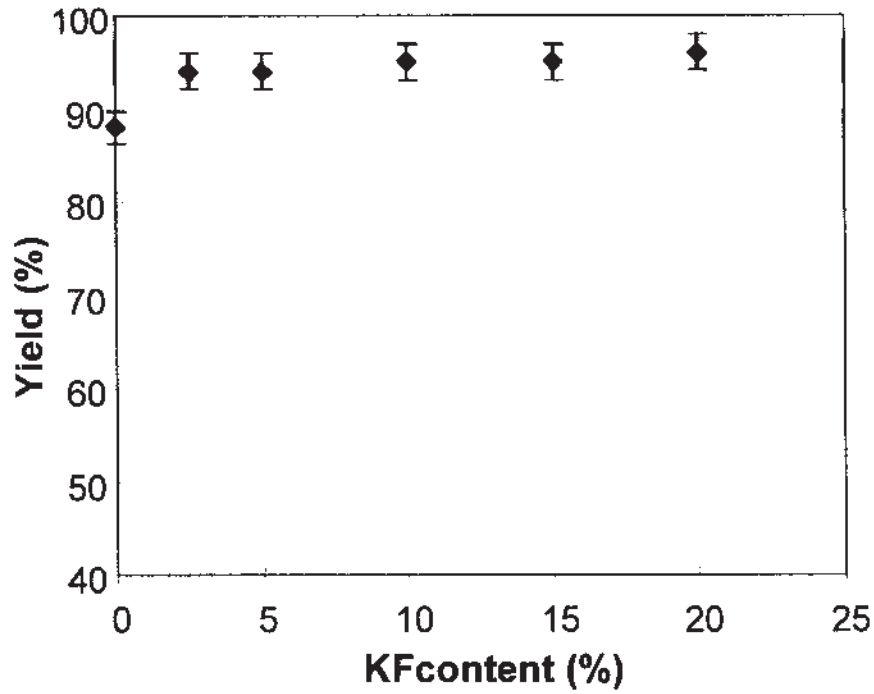
Figure 26 Effect of KF ratio in the equimolar flux on the yield. (From Ref. 66.)
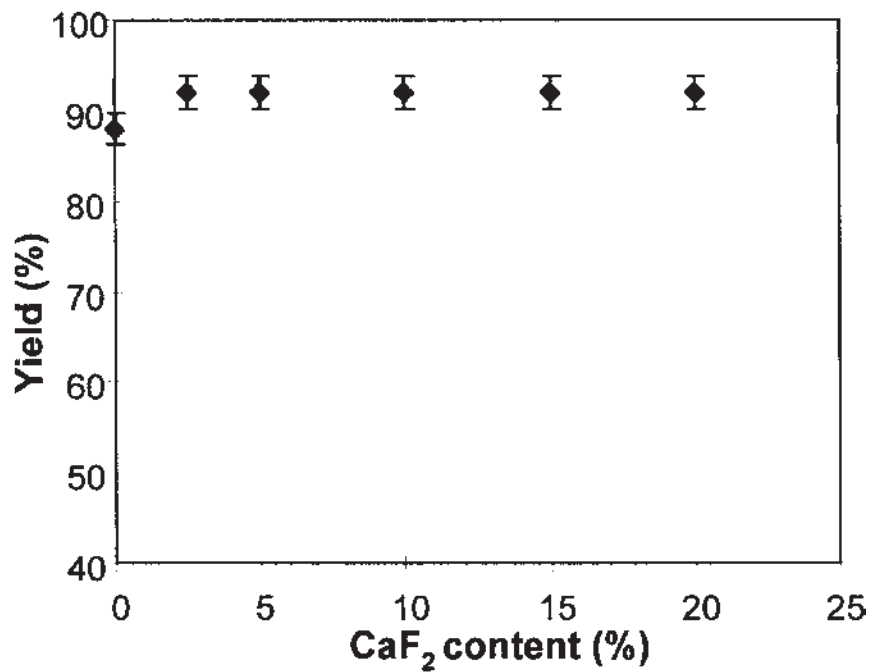
Figure 27 Effect of CaF_2 ratio in the equimolar flux on the yield. (From Ref. 66.)
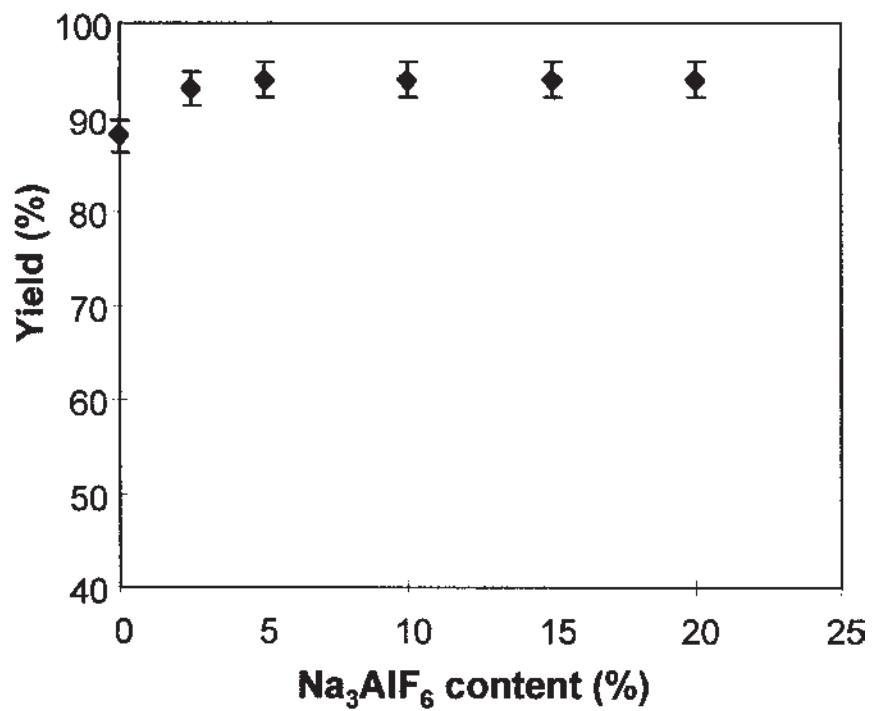
Figure 28 Effect of Na_3AlF_6 ratio in the equimolar flux on the yield. (From Ref. 66.)

aluminum scrap are being developed. This process does not use saline fluxes; the oxidation of the charge is avoided by means of lessening the oxygen potential inside the furnace. The main techniques that are being developed are the fusion in rotary furnace heated by a plasma torch [74–76], fusion in electric arc furnaces [77,78] and fusion burning fuel enhanced with oxygen [79,80].

3.7.1 Process Using the Plasma Torch

This process is being developed since 1991 by ALCAN. Nowadays, there are two plants that produce about 150,000 metric tons of aluminum per year. The process is performed in a rotary furnace heated by a plasma torch. The gas employed is oxygen, but nitrogen also can be used with an increase in the energy consumption. Figure 29 schematically depicts the process [77].

The main advantages of this process in comparison with the conventional one that uses salt fluxes are:

- the dross has no salt content;
- better heat transfer; and
- less amount of generated gases.

On the other hand, the main odds are:

- the use of electric energy; and
- the process needs trained people to operate and maintain the system.

3.7.2 Process Using Electric Arc Furnace

In 1992, Hydro-Quebec [77,78] began the development of a process to recover the aluminum content of drosses. This process uses an electric furnace, where the heating is done by an electric arc. The sketch of this furnace operation appears in Fig. 30 [77].

The DROSCAR process furnace consists essentially of a rotary furnace with graphite electrodes. Argon is purged for the stabilization of the arc and, to create an inert atmosphere.

According to the literature [77,81], the main advantage of this piece of equipment over the plasma furnace is that there is less generation of gases. Only three cubic meters of gas are generated by a metric ton of treated dross, while 30 and $300 \, m^3$ of gas are produced in the plasma and conventional processes, respectively.
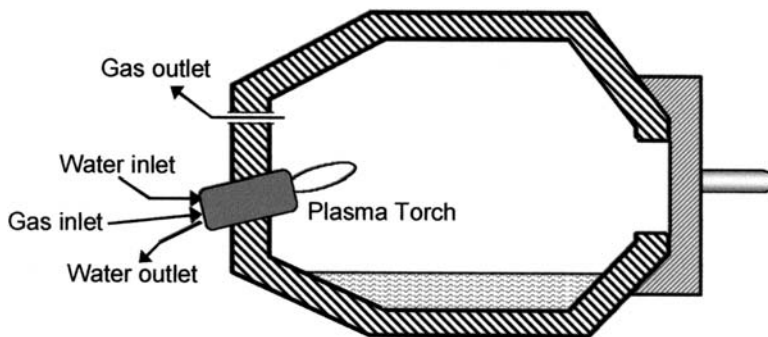
Figure 29 Sketch of the process that uses plasma torch. (Adapted from Ref. 77.)
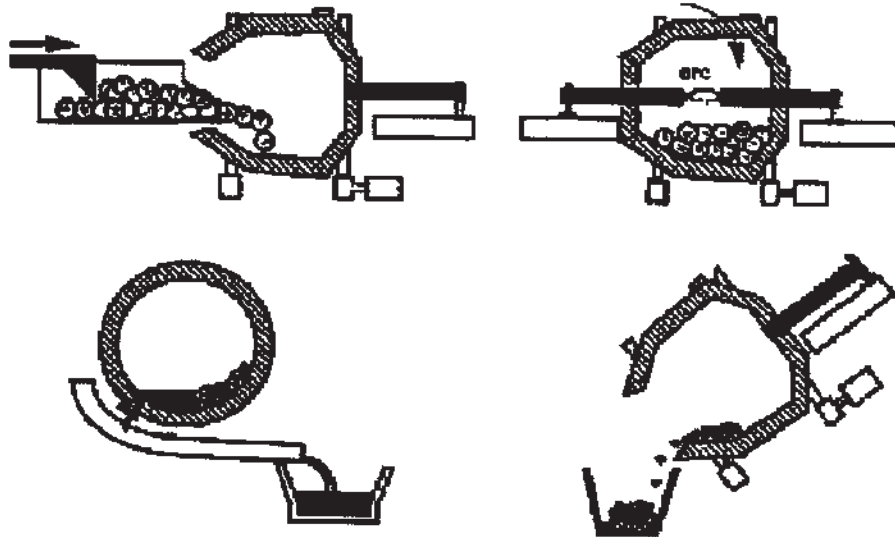
Figure 30   Sketch of the DROSCAR process. (From Ref. 77.)

3.7.3   Processes Based on the Control of the Burners

One of the most recent alternative processes developed is the ALUREC [79,80,82]. This process has been developed by AGA since 1994. It uses a rotary furnace similar to the furnace used in the conventional process, but no salt is used here. Instead, the heating is done by the burning of fossil fuel (oil or gas) with oxygen, not with air. Hence, with the optimization of the combustion, the generated gases are about 25% of the volume generated in the conventional process.

The literature points out a yield 5% superior than the yield reached in the conventional recycling process of dross. This increase in the yield is thought to be due to the non-utilization of air, so the formation of NO_x and aluminum nitrates is avoided. Figure 31 [79] shows a sketch of the ALUREC process.

3.7.4   Processes Based on Submerging the Scrap into the Bath

Another way to avoid or minimize the oxidation of aluminum is the fast immersion of the solid scrap into the molten aluminum pool. The vortex method [83] uses a
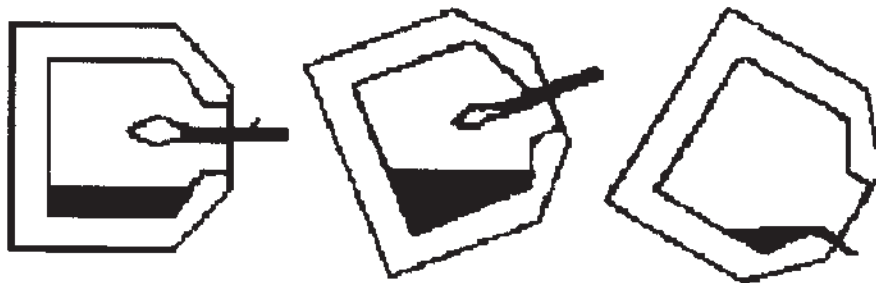
Figure 31   Sketch of the ALUREC process. (From Ref. 79.)

rotating disk to produce a vortex on the bath surface that sinks the scrap into the pool.

Another alternative to put the scrap quickly inside the bath is the metal pumping into a bowl, in which the pumped metal enters the upper portion of a cylindrical bowl tangentially causing a swirling action that quickly melts light-gauge scrap added into the bath. The charge is performed through the use of molten-metal pumps circulating the hot metal from the furnace hearth to an attached side-well melting area. Argon can be used to prevent oxidation of the scrap during the transport stage [84–86].

Both methods, vortex and side-well, are all employed on reverberatory-style melting furnaces.

3.8 Recovery of the Aluminum of the Drosses by Milling

Another alternative used in order to avoid the use of salt fluxes consists in the milling and classification of drosses that do not contain salt [87–89]. Firstly, the dross that does not contain salt is milled in a ball mill. Then, the milled material is classified into different size range (sieve classification).

The fraction of the material that has more than about 4 mm is rich in metallic aluminum; the metallic aluminum content is about 85%. This fraction goes to the melting. The fraction of material between 1 and 4 mm has approximately 70% of metallic aluminum; this material can be sold to plants that produce exothermic powders. The fraction of the classified material that has less than 1 mm contains roughly 25% of metallic aluminum. This fraction can be sold as an insulating cover for the teeming processes in steel making industries.

An analogous processing to the beforehand described is carried out on drosses that contain salts, before the leaching with water. In this case, the fraction composed of the biggest material is returned to the fusion and the fraction composed of the smallest material is forwarded to the leaching step [46,47].

3.9 Aluminum Scrap Preliminary Treatments

Frequently, aluminum scraps hold contaminants; in most cases, these contaminants are ferrous alloys or organic compounds. Consequently, it is necessary to clean the scrap before processing it.

Most of the time, the cleaning can be carried out as simply as separating the contaminants manually, which can be done by the furnace operators who know the scrap by looking at it and can identify the contaminants.

When the production is relatively high, manual separation becomes unfeasible; in this case, an investment in milling and physical separation techniques is necessary [87,90], as an example of these steps can be cited the magnetic separation followed by baling [91].

There are inks and varnishes on the beverage can scrap. These inks and varnishes correspond to 2–3% of the total mass of the can — they are applied on the interior and exterior walls for protection and product identification (brands). These coatings are composed essentially of organic resins (vinyl, epoxy, and acrylic) and pigments [92]. Besides these compounds, in the interior of the cans, it is possible to exist residues that are basically water and sugar. The sum of all these compounds represents in average 4–5% of the total mass of the scrap [90].

It was demonstrated that the presence of organic compounds on the scraps might diminish the yield of recovery by up to 10% [93].

Therefore, there exist some plants that perform a preliminary step of drying, and removal of ink and varnish [87,90,92,94–97]. This step must remove the contaminants without oxidizing the aluminum and releasing no pollutant. The cleaning process is carried out by heating the scrap up to 500–600°C [87,90,94–97]; during this treatment, there is a partial pyrolysis and oxidation of the coatings. The oxygen concentration in the emitted gas is within the range from 6% to 8% [94]. The gases from the furnace are further incinerated in order to diminish the emission of volatile organic compounds and CO_2 [90,94,96].

Drosses can also be submitted to a preliminary treatment to avoid oxidation during their slow cooling. The process must be performed where the dross is originally produced. It consists essentially of the quenching of the dross preferentially under an inert gas atmosphere [98,99]. Another usual treatment is to carry out a hot pressing of the dross. This technique promotes the flow of a part of the aluminum trapped in the dross [100,101].

REFERENCES

1. J. L. Bray, *Non-Ferrous Production Metallurgy*, John Wiley & Sons, New York, 1941, pp. 29–30.
2. N. Sevryukov, B. Kuzmin, and Y. Chelishchev, *General Metallurgy* (B. Kuznreev, Trans.), Peace Publishers, Moscow, s.d., pp. 290–291.
3. *ASM Metals Handbook*, Vol. 2, 10th Edn., 1990, pp. 62–122.
4. Source: ABAL- Associação Brasileira do Alumínio (Brazilian Aluminum Association).
5. http://minerals.usgs.gov/minerals/pubs/commodity/recycle/870497.pdf
6. R. P. Pawlek, "Secondary Aluminum Industry Annual Review," Light Metals Age, August 1998, pp. 6–13.
7. http://minerals.usgs.gov/minerals/pubs/commodity/recycle/870400.pdf
8. H. A. Oye, N. Mason, and R. D. Peterson, "Aluminum: Approaching the New Millennium," JOM, February 1999, pp. 29–42.
9. Laser Separates Aluminum by Alloy, Increases Scrap Value, JOM, January 2001, pp. 5–6.
10. D. V. Neff, "Recycling of Nonferrous Alloys," in: *ASM Metals Handbook*, Vol. 2, 10th Edn., 1990, pp. 1205–1231.
11. *Aluminum and Aluminum Alloys*, ASM International Handbook Committee, ASM International Metals Park, OH, 1993, pp. 47–55.
12. D. J. Field, G. M. Scamans, and E. P. Butler, "The High Temperature Oxidation of Al–4.2Wt Pct Mg Alloy," Metallurgical Transactions A1987, 18 A, pp. 463–472.
13. A. Bahadur, "Behaviour of Al–Mg Alloys at High Temperature," Journal of Materials Science 1987, 22, pp. 1941–1944.
14. M. Ahmad, "Thermal Oxidation Behavior of an Al–Li–Cu–Mg–Zr Alloy," Metallurgical Transactions A1987, 18A, pp. 681–689.
15. R. W. Bartlett, "Growth Kinetics of Discontinuous Thermal Oxide Films; Aluminum," Journal of The Electrochemical Society 1964, 111(8), pp. 903–908.
16. W. W. Smeltzer, "Oxidation of an Aluminum-3 Per Cent Magnesium Alloy in the Temperature Range 200–500°C," Journal of the Electrochemical Society 1958, 105(2), pp. 67–71.

17. G. M. Scamans and E. P. Butler, "In Situ Observations of Crystalline Oxide Formation During Aluminum and Aluminum Alloy Oxidation," Metallurgical Transactions A 1975, 6A, pp. 2055–2063.

18. M. H. Zayan, O. M. Jamjoom, and N. A. Razik, "High-Temperature Oxidation of Al–Mg Alloys," Oxidation of Metals 1990, 34(3/4), pp. 323–333.

19. K. Wefers, "Properties and Characterization of Surface Oxides on Aluminum Alloys," Aluminium 1981, 57, pp. 722–726.

20. O. Salas, et al., "Nucleation and Growth of Al_2O_3/metal Composites by Oxidation of Aluminium Alloys," Journal of Material Research 1991, 6(9), pp. 1964–1981.

21. K. C. Vlach, et al., "A Thermogravimetric Study of the Oxidative Growth of Al_2O_3/Al Alloy Composites," Journal of Material Research 1991, 6(9), pp. 1982–1995.

22. S. Fox, H. M. Flower, and D. S. McDarmaid, "Oxidation Induced Solute Depletion in Al–Cu–Mg–Li Alloys," Scripta Metallurgica 1986, 20(1), pp. 71–74.

23. S. Sako, K. Ohshima, and T. Fujita, "Surface Oxide Film of Metallic Small Particle," Journal of the Physical Society of Japan 1990, 59(2), pp. 662–666.

24. I. A. Shibli and D. E. Davies, "Effect of Oxidation on Sintering Characteristics of Al Powder and Effect of Some Minor Metallic Additions," Powder Metallurgy 1987, 30(2), pp. 97–101.

25. T. Rönnhult, U. Rilby, and I. Olefjord, "The Surface State and Weldability of Aluminium Alloys," Materials Science and Engineering 1980, 42, pp. 329–336.

26. C. Jardin and D. Robert, "AES and ELS Characterization of Surface Oxides on Al–Si Alloys," Applied Surface and Science 1989, 35, pp. 495–506.

27. W. Kahl and E. Fromm, "Examination of the Strength of Oxide Skins on Aluminum Alloys Melts," Metallurgical Transactions B 1985, 16B, pp. 47–51.

28. C. N. Cochran and W. C. Sleppy, "Oxidation of High-Purity Aluminum and 5052 Aluminum–Magnesium Alloy at Elevated Temperatures," Journal of The Electrochemical Society 1961, 108(4), pp. 322–327.

29. C. Lea and C. Molinari, "Magnesium Diffusion, Surface Segregation and Oxidation in Al–Mg Alloys," Journal of Materials Science 1984, 19, pp. 2336–2352.

30. C. N. Cochran, D. L. Belitskus, and D. L. Kinosz, "Oxidation of Aluminum–Magnesium Melts in Air, Oxygen, Flue Gas, and Carbon Dioxide," Metallurgical Transactions B 1977, 8B, pp. 323–332.

31. P. E. Blackburn and E. A. Gulbransen, "Aluminum Reactions with Water Vapor, Dry Oxygen, Moist Oxygen, and Moist Hydrogen between 500° and 625°C," Journal of The Electrochemical Society 1960, 107(12), pp. 944–950.

32. M. Drouzy and M. Richard, "Oxydation des alliages d'aluminium fondues," Fonderie 1974, 332, pp. 121–131.

33. M. Drouzy and M. D. Fontaine, "Oxidation de l'aluminium et des alliages aluminium–magnésium liquides," Revue de Métallurgie 1970, pp. 775–781.

34. M. J. Dignam, "Oxide Films on Aluminum," Journal of the Electrochemical Society 1962, 109(3), pp. 184–198.

35. M. J. Dignam, W. R. Fawcett, and H. Böhni, "The Kinetics and Mechanism of Oxidation of Superpurity Aluminum in Dry Oxygen," Journal of the Electrochemical Society 1966, 113(7), pp. 656–671.

36. I. M. Ritchie, J. V. Sanders, and P. L. Weickhardt, "Oxidation of a Dilute Aluminum Magnesium Alloy," Oxidation of Metals 1971, 3(1), pp. 91–101.

37. R. A. Hine and R. D. Guminski, "High-Temperature Oxidation of Aluminium–Magnesium Alloys in Various Gaseous Atmospheres," Journal of The Institute of Metals 1960–61, 89, pp. 417–422.

38. J. A. S. Tenório and D. C. R. Espinosa, "High Temperature Oxidation of Al–Mg Alloys," Oxidation of Metals April 2000, 53(3–4), pp. 361–373.

39. S. Freti, J. D. Bornand, and K. Buxmann, "Metallurgy of Dross Formation on Al Melts," Light Metals 1982, AIME, pp. 1003–16.

40. Aluminum Company of America, D. L. Stewart, et al., Salt-Based Melting Process, Int. Cl. C25C 3/06 US 5,057,194. Oct. 15, 1991.

41. T. J. Johnston and R. D. Peterson, "The Role of Magnesium in Fluxing UBC," in: International Symposium on Recycle and Secondary Recovery of Metals and Fall Extractive and Process Metallurgy Meeting, Fort Lauderdale, 1985, Proceedings, AIME, 1985, pp. 417–428.

42. T. W. Unger and M. Beckmann, "Salt Slag Processing for Recycling," in: *Light Metals* 1992, The Minerals, Metals & Materials Society, 1991, pp. 1159–1162.

43. J. N. Hryn, et al., "Products from Salt Cake Residue-Oxide," in: International Symposium Recycling of Metals and Engineered Materials, 3, Warrendale 1995, Proceedings, TMS, 1995, pp. 905–936.

44. P. N. Crepeau, M. L. Fenyes, and J. L. Jeanneret, "Solid Fluxing Practices for Aluminum Melting," Modern Casting 1992 July, pp. 28–30.

45. J. H. L. Van Linden and D. L. Stewart Jr., "Molten Salt Flux Composition Effects in Aluminum Scrap Remelting," Light Metals 1988, AIME, pp. 391–398.

46. R. M. Reynolds and M. Olper, "The ENGITEC System for Treatment of Salt Slag from Secondary Aluminum Smelters," in: International Symposium Recycling of Metals and Engineered Materials, 2, Warrendale, 1990, Proceedings TMS, 1990, pp. 439–450.

47. J. H. L. Van Linden, *Dross Treatment*, Alcoa Center, Alcoa Technical Center, Pensylvania 15069 (report MW/1551P).

48. Aluminum Company of America, Yerushalme et al., Process for the Recovery of Values from Secondary Aluminum Dross, Int. Cl. B01F 1/00 US 5, 198, 200. March 30, 1993.

49. D. Yerushalmi, "Production of Aluminum Oxide Products (NMP) from Secondary Dross and Saltcake Treatment," Light Metals 1993, The Minerals, Metals & Materials Society, 1992, pp. 1083–1084.

50. J. M. Cassells, et al., "Removal and Reuse of Aluminum Dross Solid Waste," Light Metals 1993, The Minerals, Metals & Materials Society, 1992, pp. 1075–1081.

51. Electrodialysis Process Trims Cost of Recycling Aluminum, Advanced Materials & Processes 1997, p. 19.

52. R. Narayanan and Y. Sahai, "Chemical Interactions of Dross with Water and Water Vapor in Aluminum Scrap Remelting," Materials Transactions, JIM 1997, 38(1), pp. 85–88.

53. The American Ceramic Society, Inc., *Phase Diagrams for Ceramists*, Vol. 7, Columbus, 1989, p. 206.

54. The American Ceramic Society, Inc., *Phase Diagrams for Ceramists*, Vol. 1, 5th printing, Columbus, 1985, p. 482.

55. The American Ceramic Society, Inc., *Phase Diagrams for Ceramists*, Vol. 2, 5th printing, Columbus, 1985, p. 409.

56. L. Martin-Garin, A. Dinet, and J. M. Hicter, "Liquid–Liquid Interfacial Tension Measurements Applied to Molten Al-halide Systems," Journal of Materials Science 1979, 14, pp. 2366–2372.

57. F. K. Ho and Y. Sahai, "Interfacial Tension in Molten Aluminum and Salt Systems," Light Metals 1990, The Minerals, Metals & Materials Society, 1990, pp. 717–720.

58. F. K. Ho and Y. Sahai, "Interfacial Phenomena in Molten Aluminum and Salt Systems," in: International Symposium Recycling of Metals and Engineered Materials, 2, Warrendale, 1990, Proceedings, TMS, 1990, pp. 85–102.

59. R. R. Roy and Y. Sahai, "Interfacial Tension in Molten Aluminum Alloys and Salt Systems," Light Metals 1993, The Minerals, Metals & Materials Society, 1992, pp. 1067–1072.

60. R. R. Roy and Y. Sahai, "Wetting Behavior in Aluminum–Alumina–Salt Systems," Materials Transactions JIM 1997, 38(6), pp. 571–574.

61. R. R. Roy and Y. Sahai, "Interfacial Tension Between Aluminum Alloy and Molten Salt Flux," Materials Transactions JIM 1997, 38(6), pp. 546–552.

62. R. D. Peterson, "Effect of Salt Flux Additives on Aluminum Droplet Coalescence," in: International Symposium Recycling of Metals and Engineered Materials, 2, Warrendale, 1990, Proceedings, TMS, 1990, pp. 69–84.

63. A. H. Sully, H. K. Hardy, and T. J. Heal, "An Investigation of Thickening and Metal Entrapment in a Light Alloy Melting Flux," Journal of the Institute of Metals 1953–1954, 82, pp. 49–58.

64. J. A. S. Tenório, M. C. Carboni, and D. C. R. Espinosa, "Recycling of Aluminum. Effect of Fluoride Additions on the Salt Viscosity and on the Alumina Dissolution," Journal of Light Metals December 2001, 1(3), pp. 195–198.

65. R. R. Roy, J. Ye, and Y. Sahai, "Viscosity and Density of Molten Salts Based on Equimolar NaCl–KCl," Materials Transactions, JIM 1997, 38(6), pp. 556–570.

66. J. A. S. Tenório and F. Delgado, "Optimization of Salt Composition in the Recycling of Aluminum Cans," in: Light Metals 97, 1997, Proceedings, Orlando, FL, TMS, pp. 505–509.

67. H. Rossel, "VAW Recycling Research Center − Recycling Techniques for Post-consumer Packaging," in: International Symposium Recycling of Metals and Engineered Materials, 3, 1995, Proceedings, Warrendale, TMS, 1995, pp. 663–675.

68. J. Ye and Y. Sahai, "Role of Molten Salt Flux in Melting of Used Beverage Container (UBC) Scrap," in: International Symposium Recycling of Metals and Engineered Materials, 3, Warrendale, 1995, Proceedings, TMS, 1995, pp. 639–650.

69. R. R. Roy and Y. Sahai, "Coalescence Behavior of Aluminum Alloy Drops in Molten Salts," Materials Transactions JIM 1997, 38(11), pp. 995–1003.

70. D. B. Masson and M. M. Taghiei, "Interfacial Reactions Between Aluminum Alloys and Salt Flux During Melting," Materials Transactions 1989 June, 6, pp. 411–422.

71. J. A. S. Tenorio and D. C. R. Espinosa, "Effect of Salt/Oxide Interaction on the Process of Aluminum Recycling," Journal of Light Metals December 2002, 2(1).

72. P. Kofstad, High Temperature Corrosion, Elsevier, London, 1988, pp. 502–506.

73. J. A. S. Tenorio, F. Marchini, and A. M. Barros, "Effect of the Saline Flux Composition on the Recycling of UBC and Aluminum Chias," in: Second International Conference on The Recycling of Metals, A.S.M. − American Society for Metals, Amsterdam, 1994, Proceedings, October 1994.

74. F. L. Kemeny, D. J. Sosinsky, and R. J. Schmitt, "Development of a dc Plasma-arc Furnace for Processing Aluminum Dross," in: Light Metals 1992, The Minerals, Metals & Materials Society, 1991, pp. 1147–1153.

75. S. Lavoie, C. Dubé, and G. Dubé, "The ALCAN Plasma Dross Treatment Process, a New Salt-free Dross Processing Technology," in: International Symposium Recycling of Metals and Engineered Materials, 2, Warrendale, 1990, Proceedings, TMS, 1990, pp. 451–462.

76. S. Lavoie and J. Lachance, "Five Years of Industrial Experience with the Plasma Dross Treatment Process," in: International Symposium Recycling of Metals and Engineered Materials, 3, Warrendale, 1995, Proceedings, TMS, 1995, pp. 791–801.

77. M. G. Drouet, et al., "A Rotary Arc Furnace for Aluminum Dross Processing," in: International Symposium Recycling of Metals and Engineered Materials, 3, Warrendale, 1995, Proceedings, TMS, 1995, pp. 803–812.

78. M. G. Drouet, et al., "Dross Treatment in Rotary Arc Furnace with Graphite Electrodes," JOM 1994, May, pp. 26–27.

79. H. Gripenberg, et al., "ALUREC – A New Salt-Free Process for Recovery of Aluminum from Dross and Aluminum Containing Materials," in: International Symposium Recycling of Metals and Engineered Materials, 3, Warrendale, 1995, Proceedings, TMS, 1995, pp. 819–827.

80. H. Gripenberg and P. Hagenfeldt, "Alurec – A New Technology for Salt Free Recovery Melting (alumelt) and Recycling of Aluminum and Dross," in: Seminário da Indústria e Tecnologia do Alumínio, 5, São Paulo, 1985, Anais, ABAL, 1985, pp. 42–52.

81. Electric Melter for Aluminum Recycling Undergoes Field Testing, JOM, November 1999, p. 8.

82. I. De Faro, "Alumex: Brazil's Newest Aluminum Recycling Facility," Light Metal Age 1999, August, pp. 112–114.

83. J. H. L. Van Linden, Aluminum Can Recycling in the USA – The Technical, Economical and Environmental Success Story, Alcoa center, Alcoa technical center, Pennsylvania 15069 (report Q-173).

84. A. I. Nussbaum, "Greenmelt Aluminum Scrap Recycling Plant at Hoogovens Aluminium in Duffel, Belgium," Light Metal Age 1998, February, pp. 32–41.

85. D. V. Neff, "Efficient Melting of Low Density Scrap," in: International Symposium on Recycle and Secondary Recovery of Metals and Fall Extractive and Process Metallurgy Meeting, Fort Lauderdale, 1985, Proceedings, AIME, pp. 51–64.

86. R. J. Reed and R. W. Marshall, "Upgrading Scrap Melter Operation," in: International Symposium on Recycle and Secondary Recovery of Metals and Fall Extractive and Process Metallurgy Meeting, Fort Lauderdale, 1985, Proceedings, AIME, pp. 397–406.

87. J. Wojciechowski and W. F. Fundine, "A State-of-the Art UBC Recycling Facility at IMSAMET's Idaho Plant," in: International Symposium Recycling of Metals and Engineered Materials, 2, Warrendale, 1990, Proceedings, TMS, 1990, pp. 215–220.

88. D. Grimm, "A Dross Processor for the Year 2000 – 15 Million Pounds Per Month," Light Metal Age August 1999, pp. 34–38.

89. G. J. Kulik and J. C. Daley, "Aluminum Dross Processing in the 90's," in: International Symposium Recycling of Metals and Engineered Materials, 2, Warrendale, 1990, Proceedings, TMS, 1990, pp. 427–437.

90. L. Knutsson and G. Sjöberg, "Aluminium can Recycling in Sweden," Light Metals 1992, The Minerals, Metals & Materials Society, 1991, pp. 1137–1141.

91. W. Pietsch, "Roller Presses for Secondary Metal Recycling," in: International Symposium Recycling of Metals and Engineered Materials, 2, Warrendale, 1990, Proceedings, TMS, 1990, pp. 233–241.

92. D. L. Stewart Jr. and J. H. L. Van Liden, "Measurement of Residual Carbon on Used Beverage Containers to Monitor Delacquering Effectiveness," Light Metals 1992, The Minerals, Metals & Materials Society, 1991, pp. 1143–1145.

93. A. L. Stewart, J. G. McGUBIN, and J. Sulzer, "Melting Aluminum and Aluminum Alloys," Light Metals Age December 1977, pp.13–15.

94. V. Newberry and R. F. Jenkins, "Advanced Technology Delacquering and Melting at Alcan Rolled Products, Oswego, New York," in: International Symposium Recycling of Metals and Engineered Materials, 3, Warrendale, 1995, Proceedings, TMS, 1995, pp. 685–701.

95. W. Trosch, "An Integrated Aluminum Scrap Delacquering and Melting System Meeting Strong Environmental Requirements," in: International Symposium Recycling of Metals and Engineered Materials, 2, Warrendale, 1990, Proceedings, TMS, 1990, pp. 237–267.

96. B. Mcavoy, J. Mcneish, and W. Stevens, "The ALCAN Decoater Process for UBC Decoating," in: International Symposium Recycling of Metals and Engineered Materials, 2, Warrendale, 1990, Proceedings, TMS, 1990, pp. 203–214.

97. K. A. Bowman, "Alcoa's Used Beverage can (UBC) Alloy Separation Process," in: International Symposium on Recycle and Secondary recovery of Metals and Fall Extractive and Process Metallurgy Meeting, Fort Lauderdale, 1985, Proceedings, AIME, pp. 429–443.

98. W. Fragner, "An Economical Aluminum Dross Processing System," Light Metals 1987, The Minerals, Metals & Materials Society, 1991, pp. 799–403.

99. R. P. Roberts, "A System For Processing Aluminum Dross in a Reduced Oxygen Environment," AFS Transactions 89(37), pp. 281–284.

100. New Dross Processing Technology at Hydro Aluminum's Holmestrand Rolling Facility, Light Metal Age 2001, February, 59, p. 136.

101. O. H. Perry, "The Development of the Modern Dross Press," in: Light Metals 2000, The Minerals, Metals & Materials Society, 2000, pp. 675–678.

5

Analytical Techniques for Aluminum

ALEXIS DESCHAMPS

Institut National Polytechnique de Grenoble, Domaine Universitaire
Saint Martin d'Hères, France

Many of the mechanical properties of aluminum alloys are obtained through the presence of precipitates. This can be in a favorable way, as precipitation hardening by a fine dispersion of particles, or control of the grain structure by the presence of dispersoids. But second phase particles can also control the properties in an unfavorable way, when iron or silicon-containing particles limit the fatigue life or the toughness of alloys.

In any case, the need for a detailed characterization of these precipitates is obvious if a good control and optimization of the process routes of aluminum alloys is to be obtained. This is of course a very complex task. First of all, precipitation characterization does not mean much by itself. The level of detail to be sought and the type of techniques to be used can be determined only if one decides the scale of the precipitates and the type of information that one needs.

- The scale of precipitates depends very much on the type of precipitate family. Typically, second-phase particles (non-soluble intermetallics) are larger than 1 μm, dispersoids are of the order of 100–500 nm, heterogeneous precipitates (e.g. on grain boundaries) are of the order of 10–50 nm, and homogeneous precipitates of the order of 2–10 nm. These vastly different scales demand different characterization techniques. Moreover, the different scale of these precipitates families is accompanied by vast differences in precipitate volume fraction: homogeneous precipitates can have volume fractions larger than 4%, whereas intermetallics usually show volume fractions of the order of 0.1%.
- A complete determination of the precipitation structure in a single sample would be an formidable task. Usually a few parameters have to be chosen,

155

which are relevant for the property to be understood, and will dictate the choice of techniques. These information include size (and possibly geometry), volume fraction, crystallographic structure, chemistry, distribution in space.

This chapter will try to constitute a guideline for choosing adapted experimental techniques for various problems. It will outline the possibilities and limits of numerous experimental techniques, from simple ones to implement (such as resistivity), to highly specialized experiments which can be performed only in a few laboratories in the world (such as the Tomographic Atom Probe).

This chapter will be divided in two main parts:

- characterization of hardening precipitates. The main characteristics of these precipitates are their small size ($D < 20$ nm), relatively simple geometry (sphere, rod, disk . . .) and high volume fraction (generally $>1\%$);
- characterization of second-phase particles. Their main characteristics are their heterogeneous dispersion, large size, low volume fraction and very complicated shape.

1 CHARACTERIZATION OF HOMOGENEOUS HARDENING PRECIPITATES

1.1 Crystallographic Structure

The characterization of the crystal structure of very small precipitates (nanometer size) is often not an easy task: it is difficult to obtain the crystallography of one single precipitate as the spacing between precipitates is very small. Actually, the procedure is quite different if the precipitates are presumed to be of one known structure and the aim is to check this, or if the aim is the complete determination of the structure of an unknown precipitate. In any case, by far the most widespread technique to achieve this is Transmission Electron Microscopy, since it enables to obtain simultaneously the image of precipitates and their diffraction patterns, ensuring in this way that the diffraction spots observed are actually generated by the hardening precipitates.

In the first case, the complete structure (including space group, lattice parameters and atom positions) of the presumed precipitate can be found from crystallography books [1]. The first task is to check if the diffraction patterns are compatible with this structure. Then it may be of interest to determine if there are any orientation relationships between the matrix and the precipitates.

The Al-Zn-Mg-Cu (7XXX) system can be taken as an example. Precipitation in this system starts by GP zones, which cannot be normally imaged in diffraction patterns since they have the same structure as the aluminum matrix. These GP zones are then replaced by $\eta\prime$ precipitates, which are platelets lying coherently on the (111) planes of the aluminum matrix. Until recently [2], the complete structure of η' had not been resolved, due to their very small thickness: it was presumed that they were hexagonal with their basal plane parallel to (111), but other structures were proposed as well. However, the determination of the type of precipitate was still possible, since their signature in the $[111]_{Al}$ diffraction pattern was well known [3], as shown in Fig. 1(a): a diffuse
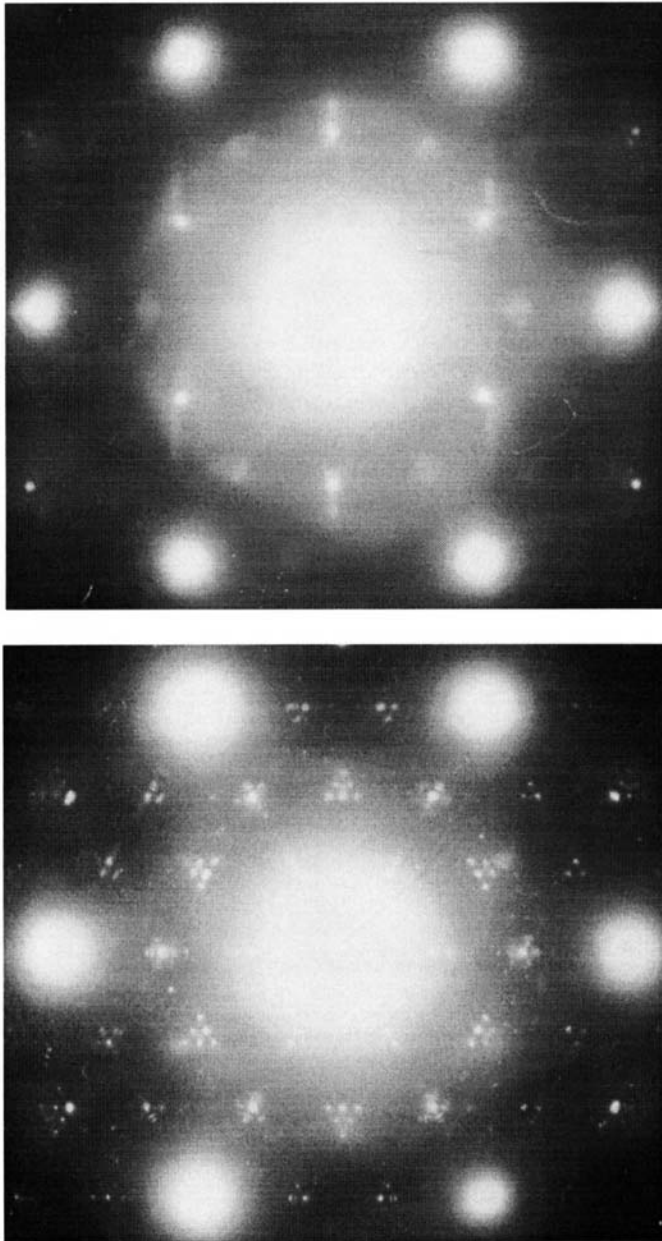
Figure 1 [111] Diffraction patterns of the aluminum matrix in an Al-Zn-Mg alloy showing diffraction spots for the precipitates. (a) Most precipitates are of the metastable η' phase; (b) most precipitates are of the stable η phase in the η_1, η_2 and η_4 variants.

spot in the center of the spots (000), (220) and (2–20) of aluminum, ensuring complete coherency with the matrix on the basal plane and resulting in a lattice parameter $a = 4.96$ Å.

The equilibrium phase of this system is η (which in the ternary system Al-Zn-Mg is $MgZn_2$). Its complete structure is known, the lattice parameters are $a = 5.14$ Å and $c = 8.6$ Å. However, a great number of orientation relationships between the precipitates and the matrix can be observed [4], depending notably on the type of nucleation site (in the bulk, on dislocations, grain boundaries, dispersoïds, etc.). These different types of orientation relationships (which are noted as η_1, η_2, etc.) can be distinguished from the location of their diffraction spots, for instance on the [111]Al diffraction pattern again (Fig. 1) [3]. In this case, one can clearly distinguish the presence of the η_2 and η_4 variants.

This type of diffraction pattern analysis applies to most of the aluminum hardening systems. Examples of analysis on various systems can be found in the literature for Al-Mg-Si-Cu [5,6], Al-Cu-Mg [7], Al-Li based alloys [8,9], etc.

High Resolution Electron Microscopy can also provide useful information, since this technique enables the spatial resolution of the atomic columns. From such images, one can obtain information such as lattice parameters or coherency strains between precipitate and matrix. Figure 2 shows a HREM image of a β'' precipitate in an Al-Mg-Si-Cu alloy.

The complete determination of the structure of unknown precipitates is a much more complicated task. Most precipitates have a complicated structure and large lattice parameters, thus the combination of several diffraction and imaging techniques is necessary, and some simplifications have to be made (e.g. the determination will be much more difficult if several phases are present in the same sample). The point group of the precipitates can be determined by symmetry analysis of dark field High Resolution TEM micrographs. A good example of structure determination is the work of Donnadieu et al. On alloy 6056 (Al-Mg-Si-Cu) in T6 condition
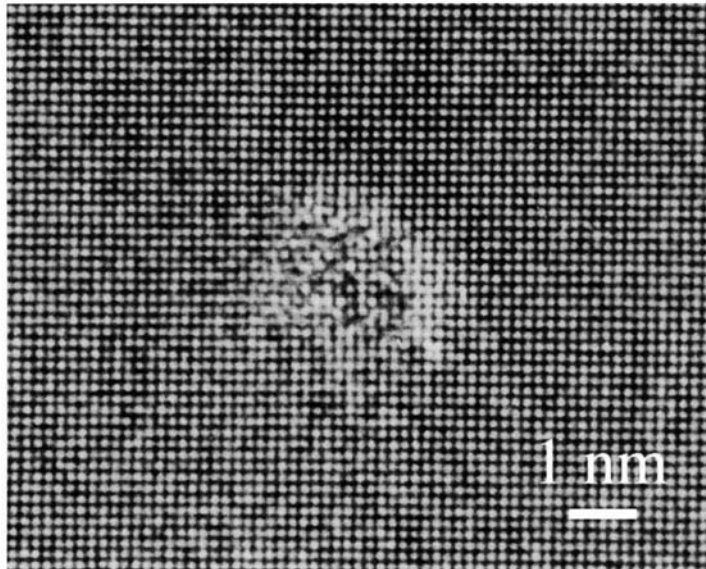
Figure 2   High resolution electron micrograph of a β'' precipitate in an Al-Mg-Si-Cu alloy in the T6 condition (Courtesy of P. Donnadieu, LTPCM, France).

[2,5]. Through a combined use of DSC, diffraction patterns and dark field HREM in the TEM, the authors were able to determine that the precipitates were of the 2/m 2/m 2/m point group (orthorombic structure), with lattice parameters $a = 0.8$ nm, $b = 1.1$ nm and $c \sim 0.6$ nm. An other good example of structure analysis using HREM is the work of Edwards and co-workers [10] on alloy 6061.

1.2 Precipitate Size

The determination of precipitate size and its evolution with process conditions is extremely important for the understanding of the final properties of the material. Most obviously, in most precipitation hardening materials, a critical size of precipitates provides the highest level of yield strength. Two main methods enable the determination of precipitate size: small-angle x-ray or neutron scattering provides an average indirect measurement on a very large population of precipitates, and Transmission Electron Microscopy provides a local direct measurement, but with poorer statistics.

1.2.1 X-ray and Neutron Scattering

The principle of Small-Angle x-ray or Neutron Scattering (SAXS and SANS) is simple. We consider a material containing two phases which have a different electronic density (for SAXS measurements), a matrix and precipitates. X-rays traversing the sample are scattered by these precipitates, and the intensity as well as the angle at which this scattering occurs depends on the size of the precipitate, the electron density contrast between the precipitates and the matrix, and the volume fraction of the precipitates. Thus, a detailed study of the scattered spectrum can provide useful information on the size of the precipitates, and a relationship between their composition and volume fraction [11].

SAXS can be performed with a variety of x-ray sources. They range from cheap and easy to handle x-ray tubes (used as well in diffractometers), more powerful rotating anodes, up to synchrotron radiation sources which provide the highest brilliance and density of x-rays. The range of precipitate sizes which can be investigated by this technique is of the order of $R = [0.5$ nm, 10 nm] in usual experiments, and newly developed USAXS (Ultra Small Angles) enable to measure precipitate radii up to 100 nm [12]. In most cases aluminum samples are studied close to the wavelength of CuK_α (1.54 Å, ~ 7.8 keV), with an associated sample thickness of about 70–80 μm.

A condition for using the SAXS method is of course that a high contrast in electronic density exists between the precipitates and the matrix, since the scattered intensity is proportional to the square of this $\Delta \rho$. This contrast is proportional to the difference in average atomic number, except if it is enhanced by taking an x-ray wavelength close to an absorption edge of one of the elements (this case is referred to as *anomalous scattering*). Thus it is hopeless, for instance, to study by SAXS precipitation processes in an Al-Mg-Si alloy where the contrast between precipitates and matrix is almost zero. In this case, one can use Neutrons Scattering (SANS), since the scattering cross-section to neutrons is not proportional to the atomic number [13]. SANS measurements can only be performed on scientific neutron nuclear reactors, which offer limited access, and are thus less

straightforward than x-ray experiments. However, SANS measurements average on a much larger volume of the material, since neutrons can traverse more than 1 cm of aluminum.

SAXS and SANS share also one other advantage: they can be performed in-situ during complex thermal histories, if the sample is heated under the beam. Thus it is possible to obtain direct access to precipitation kinetics of isothermal or non-isothermal phase transformations.

The measurement of scattering provides a number of x-rays (or neutrons) per sec and per solid angle, as a function of the scattering vector $q = 4\pi\sin\theta/\lambda$, where θ is the scattering angle and λ the wavelength. This spectrum can be transformed easily in normalized intensity I (or scattering cross section $d\Sigma/d\Omega$), which will be used for subsequent analysis [11]. One way to determine an average precipitate radius is then to consider Guinier's approximation, which states that in an intermediate range of scattering vectors the intensity is as follows:

$$I \propto \exp\left[-\frac{q^2 R_g^2}{3}\right]$$

R_g is the Guinier or gyration radius of the precipitates, and represents the mean square distance in the precipitate from the center of gravity, where the role of mass is played by the electrons. Under the initial calculation, this approximation was supposed to be true only for $q.R_g \ll 1$. However, in practice it is best followed in the range $0.8 < q.R_g < 2$.

One other useful result from the scattering spectrum is the large angle asymptotic behavior (so-called Porod behavior): at large angles the intensity becomes proportional to q^{-4}. If this is adjusted to the experimental behavior, the complete scattering behavior can be known until $+\infty$. Then the so-called invariant of the spectrum Q_0 (also called integrated intensity):

$$Q_0 = \int\limits_0^\infty I(q)q^2\,dq$$

can be calculated with precision. The integrated intensity yields several applications. The first is the calculation of an other average radius of the microstructure, the Porod radius, which is defined by:

$$R_p = \frac{3}{\pi(1-f_v)}\frac{Q_0}{\lim\limits_{q\to\infty} I.q^4} = 3\frac{V}{S}$$

The Porod radius is actually proportional to the ratio between the volume of the precipitates and the surface of their interface with the matrix.

The integrated intensity can also be used for the determination of precipitate composition and volume fraction. This will be presented in Secs. 1.3 and 1.4.

An ideal system for using Small-Angle X-ray Scattering is the 7000 series system, since the contrast between (Cu,Zn) atoms and (Al,Mg) atoms is very high. An example of a Guinier plot (ln(I) vs. q^2) is given in Fig. 3(a) In situ measurements at different temperatures can provide a detailed description of the precipitation kinetics, including nucleation radii, growth and coarsening rates [14].
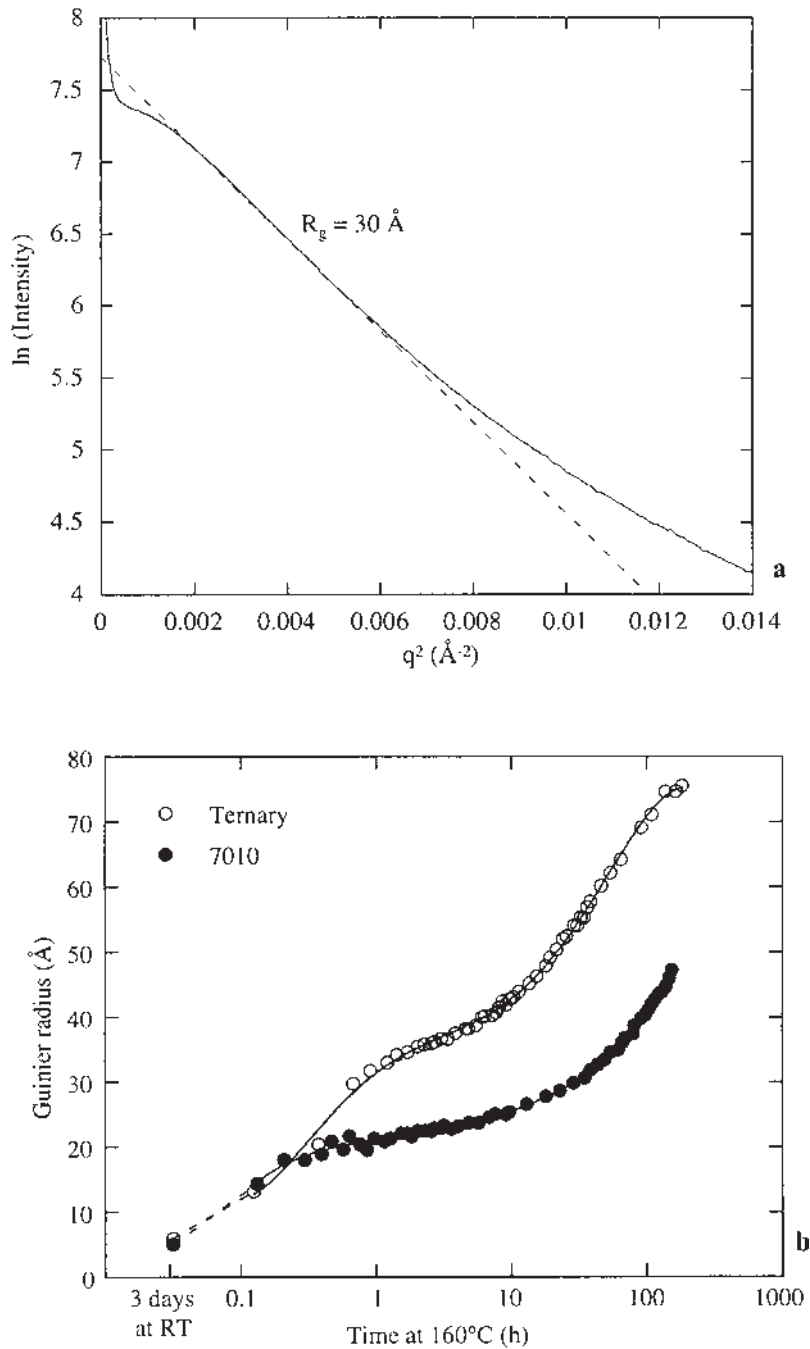
Figure 3 (a) Guinier plot representing the logarithm of scattered intensity vs the square of the scattering vector. The linear part of the curve enables to calculate the Guinier radius of the microstructure (peak-aged 7050 alloy), which is 30 Å; (b) Evolution of Guinier radius vs aging time at 160°C measured by in-situ SAXS in a ternary Al-Zn-Mg alloy and in the 7010 alloy (Al-Zn-Mg-Cu) (from Ref. 14).
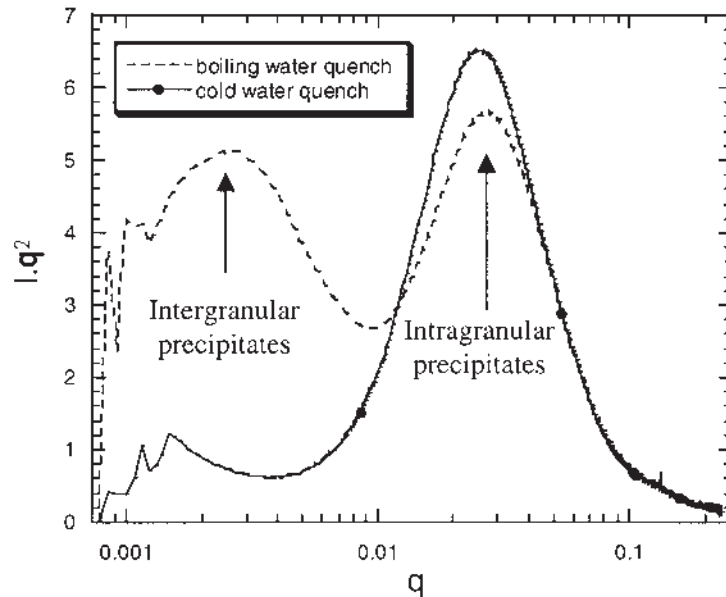
Figure 4   Ultra-small Angle X-ray Scattering measurements on 7040 alloy enable to separate in this I.q2 vs q plot two families of precipitates: intergranular coarse precipitates of radius approximately 700 Å, and intragranular precipitates of radius 60 Å. Note that in the case of a rapid quench intergranular precipitates are almost absent (Courtesy of D. Dumont, LTPCM, France).

Such results are shown in Fig. 3(b). In some more complicated cases, where several precipitate families are present, the Guinier approximation is much less straightforward, because of the overlap of the scattering behaviors of the different families. In this case, the two distributions of precipitates can be visualized in a plot of the "integrated intensity", i.e. $I.q^2$ vs. q. In the example shown in Fig. 4, it was possible using Ultra-Small Angle X-ray Scattering to separate homogeneous precipitates and heterogeneous precipitates on grain and sub-grain boundaries [12]. The respective precipitate sizes could be determined and the volume fraction of heterogeneous precipitates was found to represent approximately 10% of the total volume fraction.

Finally, the precipitate shape can *in principle* be determined by SAXS and SANS experiments [11]. In practice this is not an easy task due to experimental limitations. However, if a geometry is known from TEM experiments for instance, its parameters (such as aspect ratio, parameters of a precipitate size distribution) can be determined in favorable conditions (high contrast, low background noise, good counting statistics, etc.); For instance, Donnadieu and co-workers were able to fit the parameters of a log-normal size distribution to a SANS spectrum of β'' precipitates found in the 6056 Al-Mg-Si-Cu alloy in the peak aging condition [5], as shown in Fig. 5. In this analysis, both the average diameter and the average height of the precipitates were determined, namely 2.5 nm and 8 nm.
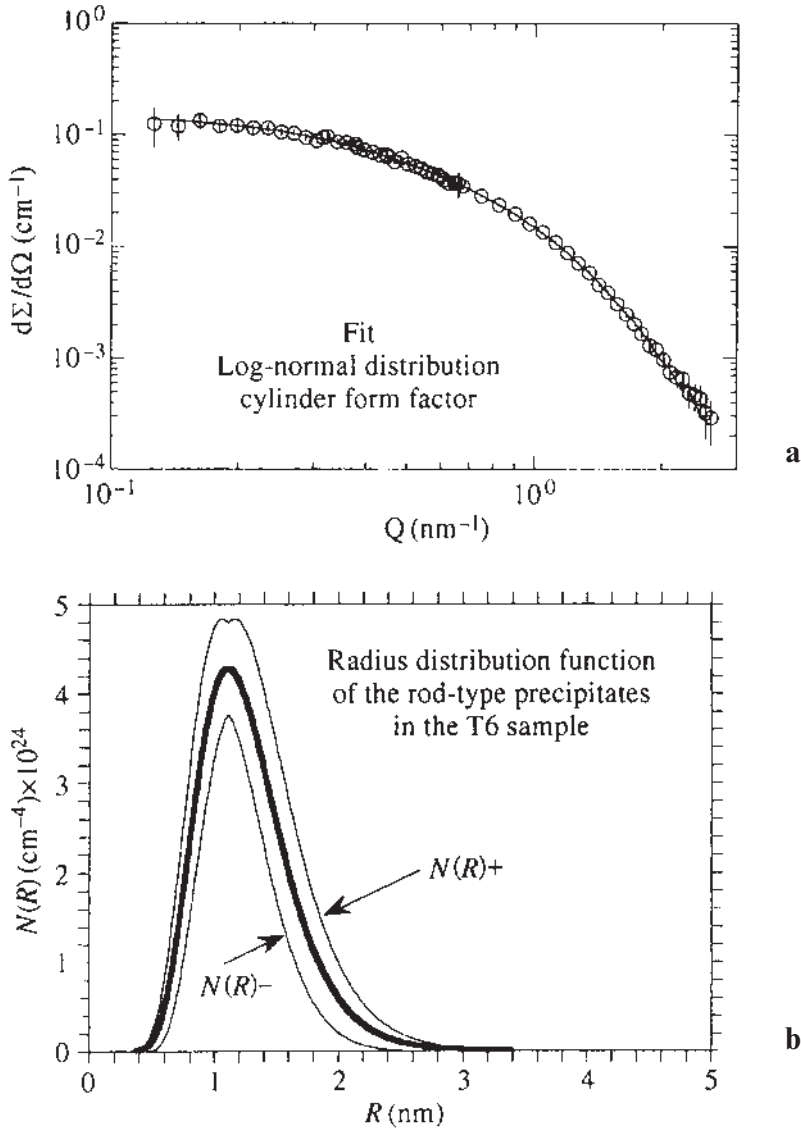
Figure 5 Small-Angle Neutron Scattering analysis on an Al-Mg-Si-Cu alloy in the T6 temper. (a) Fit of the experimental scattering spectrum supposing a log-normal distribution function of the precipitate sizes; (b) corresponding precipitate size distribution (from Ref. 5).

1.2.2 Transmission Electron Microscopy

The alternative technique to study precipitate sizes is TEM. This technique is the exact complementary of SAXS and SANS: it provides direct information on size, morphology and spatial distribution of precipitates, but is limited by poor statistics and the problem of a 2D projection of a 3D structure.

TEM is probably the only technique which can really provide information on the spatial distribution of precipitates. One can for instance characterize the precipitates-free zones which in some systems can exist close to grain boundaries (Fig. 6(a)) or around coarse quench-induced precipitates (Fig. 6(b)), or the nucleation mechanism of a coarse quench-induced precipitate on a previously existing dispersoïd (Fig. 6(c)) [15]. Such microstructural features are extremely important to characterize as they control many mechanical properties such as the fracture behavior [16]. The morphology can also be determined by TEM, such as the aspect ratio of plate-like η' precipitates in a 7108 alloy (Fig. 7). An average value of the aspect ratio can even be determined in some specific cases by the study of the diffraction patterns if the precipitates are very small, since the size of the precipitates in one given direction is inversely proportional to the size of the diffraction spot in the orthogonal direction. This was used for instance by Donnadieu to determine the aspect ratio of precipitates in the 6056 alloy [5].

TEM micrographs can also be used to obtain directly the precipitate size distribution, using image analysis. This can be done on bright field images, however the non-uniform contrast of the precipitates and matrix make automatic detection of the precipitates impossible, and underlining the precipitate border by hand is necessary. This type of analysis can thus best be performed on dark field images, which provide high contrast between white precipitates and black matrix. However, the selection of one single diffraction spot for constructing the image means that many precipitates are not counted. Therefore it is necessary to know the different variants of the orientation between the precipitates and the matrix in order to be sure that the measured distribution is representative of the complete microstructure. The other issue concerns the shape of the precipitates. If the precipitates are spherical, the diameter measured on the pictures is the actual diameter of the precipitates, except if the sample thickness is of the order of the precipitate size: in this case, many precipitates are cut by the thinning process during sample preparation and thus only partly imaged. Some mathematical corrections exist to calculate the actual distribution from the measured one, provided of course that the sample thickness is known [17] (see next section for the sample thickness measurements). Other simple cases may be easier to characterize: for instance, if the precipitates are rods oriented parallel to the [001] directions of the matrix, as is the case of β' in 6000 series, for instance, the analysis of a thin foil oriented as such will provide an exact measurement of the cross-section of these rods.

1.3   Volume Fraction

The evaluation of the volume fraction of precipitates is very important for the understanding of the mechanical properties of age-hardenable aluminum alloys. However, it is very seldom carried out because it is extremely delicate to perform. An independent quantitative determination of the volume fraction of small precipitates in one sample is a long and difficult work by itself. In this section we will describe which techniques may lead to a determination of volume fraction, and the degree of self-consistency of these methods.
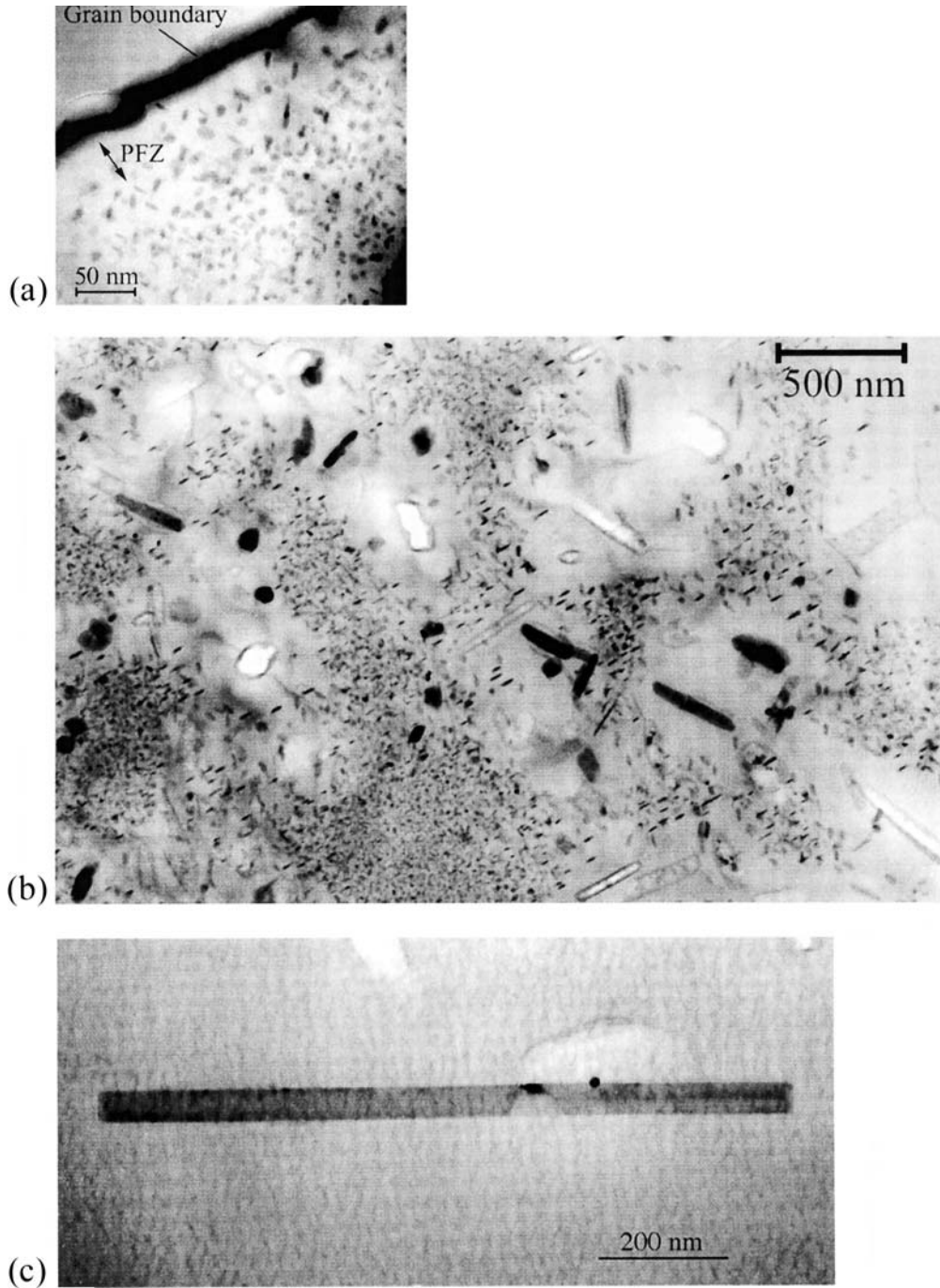
Figure 6 Microstructures in the 7010 alloy (Al-Zn-Mg-Cu) as characterized by TEM; (a) Precipitate-Free Zone around a grain boundary; (b) Precipitate depleted zones around coarse quench induced precipitates after a slow quench from the solution treatment; (c) Quenched-induced precipitate nucleated on an Al₃Zr dispersoïd (b and c from Ref. 15).
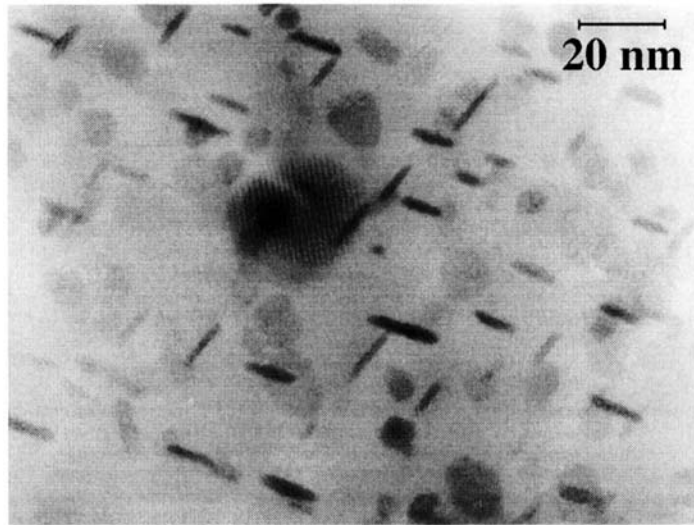
Figure 7 η' precipitates in a Al-Zn-Mg alloy in the T6 condition. One can clearly see in this near-$\langle 111\rangle$ zone axis the four crystallographic variants of the precipitates. The precipitates are platelets lying on the $\{111\}$ planes of the matrix.

1.3.1   Resistivity

The easiest method to determine precipitate volume fraction is also the most difficult to interpret. The principle of resistivity measurements during precipitation processes is to determine the loss of solute concentration in the matrix and convert this loss in volume fraction. This is achieved by the fact that precipitates do not scatter significantly electrons (except if they are too small, see below), contrarily to solute atoms. Therefore, as far as resistivity is concerned, a precipitation process is equivalent to a loss of solute from the matrix. In order to interpret such measurements, a number of conditions have to be fulfilled:

- Mathiessen's law has to be verified, i.e. there should be a linear addition law between the solute contribution and the other contributions to resistivity (such as phonon scattering);
- the relationship between solute concentration and resistivity has to be known or calibrated [18];
- the composition of precipitates should be known, in order to relate the volume fraction to the amount of solute lost;
- One should be aware of the resistivity anomaly [19]: when precipitates are small, they scatter more efficiently than solute atoms, and therefore the precipitation process induces an increase of resistivity instead of a decrease. This resistivity anomaly is strong up to sizes of about 2 nm in diameter. Therefore, in order to be completely aware, one should not attempt to measure precipitated volume fractions from resistivity below an average precipitate diameter of 5–10 nm.

This method has been successfully applied for instance by Godard in order to determine the evolution of precipitate volume fraction during continuous cooling from the solutionizing temperature in the 7010 alloy [20]. Figure 8 shows the evolution of volume fraction for different cooling rates. These measurements can then be modeled in a Johnson–Mehl–Avrami framework. In this case, precipitates were presumed to be equilibrium Mg-$(Zn,Cu)_2$ precipitates, and their size was relatively coarse because of their high nucleation temperature.
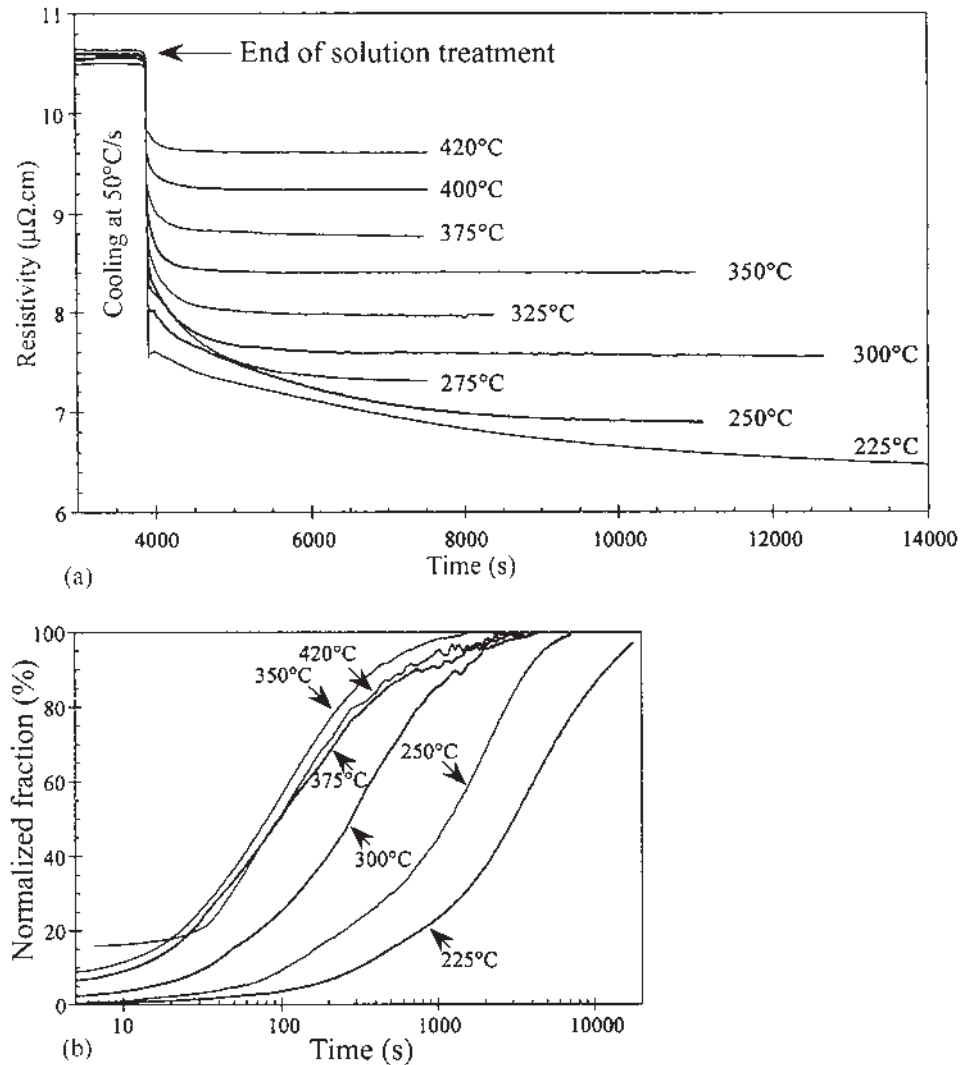
Figure 8 Study of precipitation processes during high temperature isothermal treatments using in-situ resistivity measurements. (a) evolution of resistivity during the heat treatments; (b) analysis in terms of normalized fraction of precipitates (Courtesy of D. Godard, LSG2M, France).

1.3.2 Small-Angle Scattering

The other relatively easy method for volume fraction measurements is Small-Angle Scattering (SAXS or SANS) [21]. For the principle of these experiments, see Sec. 1.2.1. In a two-phase model (matrix of uniform concentration and all precipitates of the same composition), the integrated intensity Q_0 depends on the volume fraction in a simple manner:

$$Q_0 = 2\pi^2 f_v (1 - f_v)(\Delta\rho)^2$$

$\Delta\rho$ is the difference in electronic density between the precipitates and the matrix, and is expressed as:

$$\Delta\rho = \frac{Z_p}{\Omega_p} - \frac{Z_m}{\Omega_m}$$

where Z and Ω are the average atomic number and average atomic volume in the precipitate and the matrix, respectively.

The integrated intensity is first calculated in the measurement range $[q_0, q_1]$. If the precipitates are not too large (and therefore do not scatter to too small scattering angles), the integrated intensity from the $[0, q_0]$ part of the reciprocal space can be neglected. However, the large angle contribution $[q_1, \infty]$ is usually not negligible and has to be calculated. This can be simply achieved by fitting the asymptotic Porod behavior in the large angle section of the measured spectrum, and thus adjusting the constant A:

$$I(q) \underset{q \to \infty}{\longrightarrow} \frac{A}{q^4}$$

Then the large-angle contribution to the integrated intensity is simply:

$$\Delta Q = \int\limits_1^\infty I(q).q^2 dq = \int\limits_{q1}^\infty \frac{A}{q^2} dq = \frac{A}{q_1}$$

1.3.3 Transmission Electron Microscopy

The volume fraction of precipitates can be estimated directly by means of Transmission Electron Microscopy [22]. This technique has the advantage of being direct (no assumption is needed regarding the precipitates composition), but is very cumbersome and time consuming. The principle is as follows: precipitates need to have a simple geometry (rods, spheres ...) and to be aligned with respect to the electron beam. The contrast between the precipitates and matrix should be high, and therefore the dark field imaging mode is privileged. The number density and size of the 2-dimensions projection of the precipitates is measured on the photographs by image analysis. The total number density is then calculated by simple geometrical rules. For instance, in case of rods aligned along the <100> directions of the matrix, three variants are equivalent, and the total number density of precipitates is three times the one measured on a dark field micrograph taken with a single precipitate

reflection. Finally, the sample thickness has to be measured. Several techniques are possible, but the two most suitable are the measurements by Convergent Beam Diffraction (CBED) and Electron Energy Loss Spectroscopy (EELS).

The EELS method is based on the measurement of the zero-loss peak intensity I_0 [23]. It is related to the total intensity I_t by the relation $\log(I_t/I_0) = \exp(t/\lambda)$, where t is the sample thickness and λ is the mean free path of the electron in the material. Thus if the mean free path of the electron is known, the sample thickness can be known with good accuracy (about 10%) since the method relies on the ratio between two intensities, and therefore does not depend on an absolute intensity measurement. λ can be computed from the experimental conditions provided that the material is appropriately described by the free-electron approximation.

However, the CBD method appears to be both more precise and more easy to use than the EELS method [22]. It is applicable to most TEM samples, and fails only with extremely thin samples (< 10 nm), heavily distorted structures (large strain deformations) or very absorbing elements (e.g. gold). The method and range of applications can be found in [24]. It consists of making a CBD pattern under two-beam conditions and in measuring the positions of the fringes inside the diffraction disc, which leads for each fringe to the related deviation s to Bragg angle (Fig. 9). According to the two-beam dynamical theory [25], the distribution of diffracted intensity depends on the deviation s, the extinction distance $\xi\mathbf{g}$ and the local thickness t. The white and black fringes correspond to the intensity extrema. Therefore, it is possible to derive the local sample thickness from the fringe positions. The range of measurable thickness depends on the extinction distance and on the absorption coefficients. According to Allen [24], measurements are possible in aluminum between 22 and 790 nm under two-beam conditions with a diffraction vector $\mathbf{g}(220)$.

Finally, corrections have to be made in order to take into account the fact that some precipitates are cut by the sample surface. Again, the correction depends on the geometry of the precipitates. In case of spheres, the apparent diameter of the precipitates will be changed by this artifact. Mathematical corrections such as the Schwartz–Saltikov method [17] are available to get the actual precipitate size distribution and number density. In the case of rods perpendicular to the sample surface and parallel to the electron beam, the correction will depend on their height as compared to the thickness of the sample. If the experimental height distribution is known (for instance by the analysis of the diffuse scattering in the diffraction patterns), as well as the sample thickness, the actual distribution and number density of precipitates can also be calculated.

1.3.4 Tomographic Atom Probe

Perhaps the most direct method of all, the newly developed Tomographic Atom Probe (TAP) enables the direct reconstruction in 3 dimensions of the sample volume, atom by atom [26]. Thus it enables the full determination of the microstructure, including the precipitate size, composition and volume fraction, as well as the solute content of the matrix, for instance (see Fig. 10). Of course the disadvantage is a very poor statistics as compared to other experimental techniques, since only a few 100,000 atoms can be analyzed per needle.
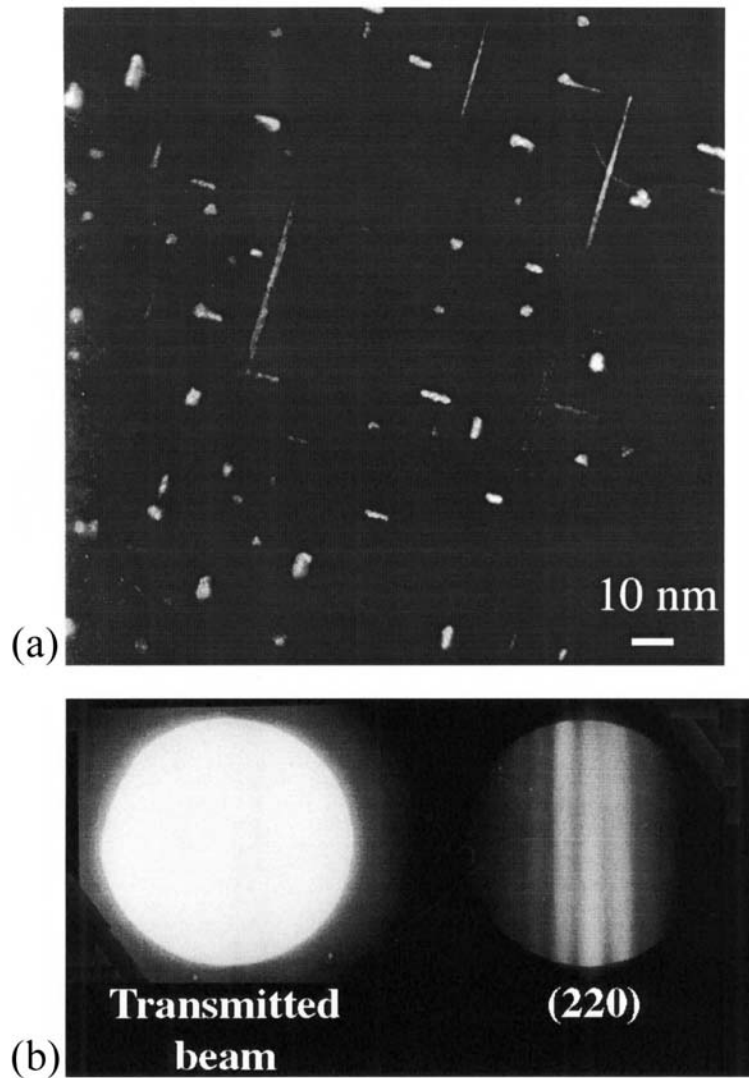
Figure 9   (a) Dark Field Electron Micrograph of 6056 aluminum alloy ([100] zone awis), showing the presence of β'' precipitates in the shape of rods. The volume fraction of these precipitates could be determined in this sample by analyzing the sample thickness using a convergent beam diffraction image of a (220) spot (b). The thickness of the sample can be determined from the spacing between the fringes in the diffracted spot (Courtesy of P. Donnadieu, LTPCM, France).

This technique can also lead to some artifacts due to local magnification effects, and therefore the measurements obtained by this technique should be checked with other experimental techniques such as the ones exposed above.
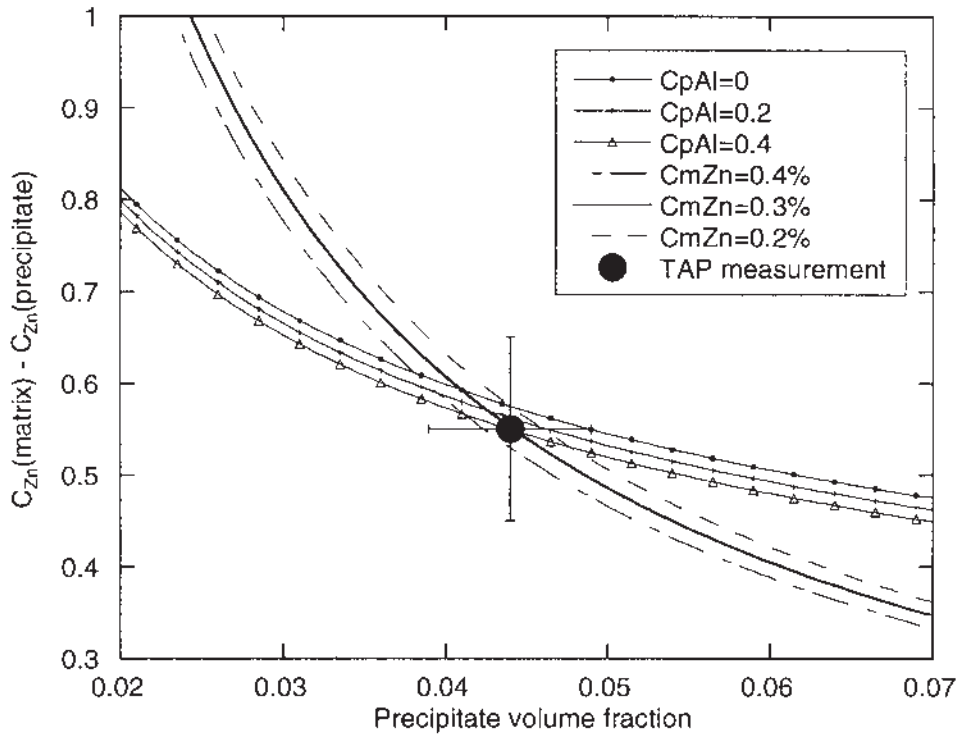
Figure 10   Determination of the Zn composition of η precipitates in an Al-Zn-Mg alloy by SAXS measurements: Given some hypothesis concerning the Zn concentration in solid solution (measured by Tomographic Atom Probe) and the composition in aluminum of the precipitates, the intersection of the two curves gives the Zn concentration of the precipitates and the precipitated volume fraction. TAP measurements showed very good agreement with the result.

1.4   Chemistry

Determining the chemistry of fine scale precipitates is a rather complicated task. First, fine scale precipitates in aluminum alloys are usually not of the same crystallographic nature as the stable phase predicted by the phase diagrams. There is therefore no reason for which their composition should be similar to the bulk composition of the intermetallic phases. Moreover, fine precipitates are usually associated with a strain field, which can substantially change their equilibrium concentration in the various solutes.

Knowing the precipitate composition is however very important. The balance between the various solutes (in the frequent case of a multi-constituent alloy) can determine the evolution of the precipitation driving force during the aging process. The total content of solute in the precipitate (or in other words the aluminum content in the precipitates) determines in turn the volume fraction attainable for a given amount of solute and can have substantial consequences on the resulting mechanical properties of the material.

Several methods are available for measuring the composition of nanoscale precipitates. However, even though these methods are probably relatively reliable for the determination of the relative balance between the different elements other than aluminum (for instance, the ratio between Si, Mg and Cu in a 6000 series aluminum alloy), the aluminum content of the precipitates cannot be measured with certainty in the present state of the art.

1.4.1   Transmission Electron Microscopy

A first method for analysis of precipitates composition is the use of TEM. However, dealing with nanometer sized precipitates means that the beam has to be focused on a similar scale. This is best achieved in STEMs (Scanning Transmission Electron Microscopes), which can provide when equipped with a Field Emission Gun probe sizes down to 1 nm. Two methods can be then used to measure the precipitate composition: Energy Dispersive X-Ray Spectroscopy (EDXS) [27,28], or Electron Energy Loss Spectroscopy (EELS) [29,30]. The main problem for both methods is to isolate the precipitate from the matrix. Ideally, the sample should be prepared in form of carbon replica so that there is no interaction with the aluminum of the matrix in the measurement. In practice, this is often impossible for very small precipitates and the main result of the measurement will be the ratio between the different solutes inside the precipitates.

EDX analysis can be achieved for all elements except the lightest elements (when the detector is equipped with a beryllium window lighter than Na). With EDX the spatial resolution of the analysis is about 2 nm in diameter, due to the beam broadening across the sample. With an acceleration voltage of 100 kV, the sample should be thinner than 100 nm in order to get a reasonable signal-to-noise ratio. Quantitative analysis is possible provided that more than 1 at% of the element to measure is present. Of course quantitative analysis are best performed if some standards (e.g. pure materials, phases of known composition) are measured in parallel. If not, the composition can be calculated from EDX spectra using appropriate softwares, but only if the K or L edges of the element can be recorded. An example of the type of study which can be carried out with STEM is shown in Fig. 11. The composition across a Precipitate-Free zone is measured by EDX in an Al-Zn-Mg-Cu alloy. The solute depletion of the PFZ is clearly identified on the calculated Zn profile, and also an increase in Zn concentration is observed on the grain boundary, due to the presence of a grain boundary precipitate.

EELS analysis enables the detection of all elements, but is particularly sensitive to light elements. In this mode spatial resolution of the analysis is better than in EDXS mode and can be as low as 1 nm. When the sample is free of contamination and is thinner than 50 nm, quantification can be performed with a good accuracy for all light elements as well as for elements up to the transition series. For heavier elements quantification becomes impossible without standards. Thus EELS and EDXS are complementary techniques that can be used to solve micro and nanoanalytical problems. Figure 12 shows an example of the use of the combination of these techniques to determine the composition of very small oxide particles $MgAl_2O_4$ at the interface of SIC particles in a SIC-reinforced Al-Mg alloy. Moreover, in this case the composition profiles of the various elements could be measured across the interface from EDX measurements.
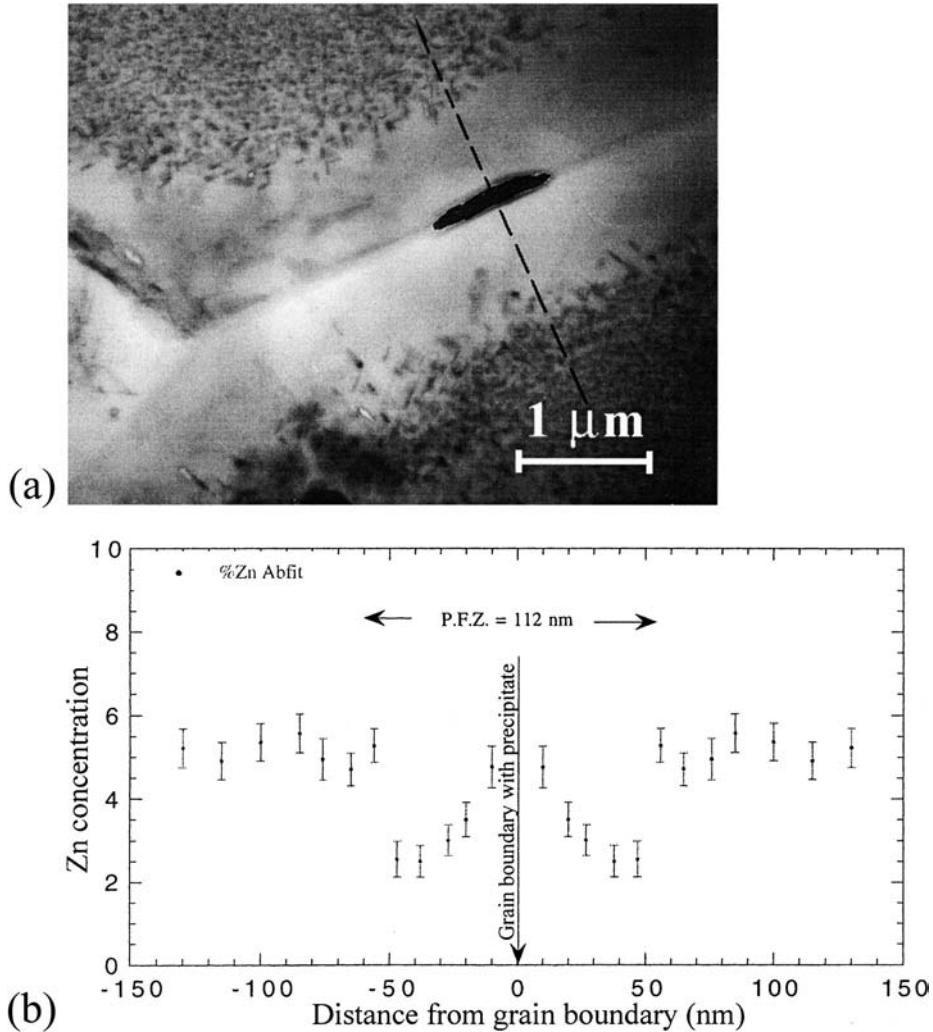
Figure 11 (a) TEM micrograph showing a Precipitate-Free Zone around a grain boundary in an Al-Zn-Mg-Cu alloy; (b) Zn profile of across this Precipitate Free Zone showing the solute depletion in the PFZ and the sole enrichment close to the grain boundary precipitate (Courtesy of D. Solas and M.C. Cheynet, LTPCM, France).

1.4.2 Small Angle X-ray Scattering

Small-Angle X-ray Scattering can be used for determining some characteristics of the composition of precipitates. However, a complete determination is impossible, since a high density of dilute precipitates will give the same SAXS signal as a low density of concentrated precipitates, as discussed in Sec. 1.3.2. Actually, if one is able to determine one parameter of the microstructure, the others can be fully determined by SAXS measurements. For instance, the determination of the solute
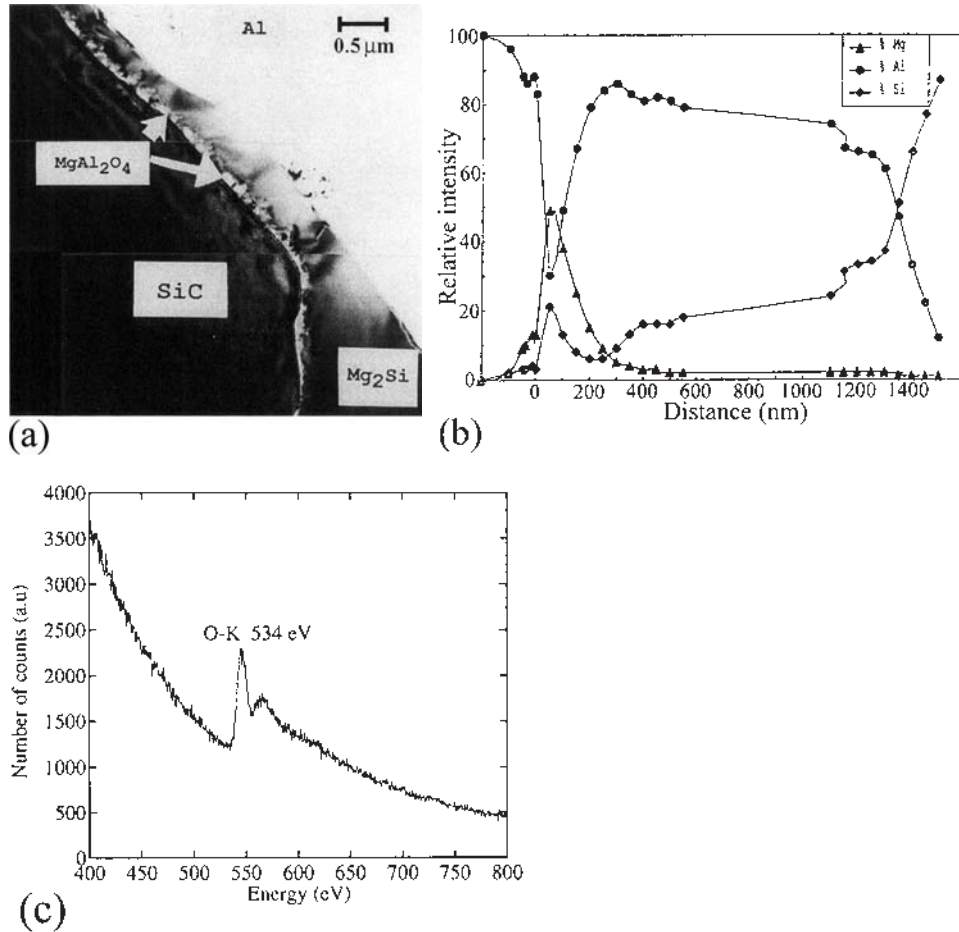
Figure 12 Analysis of a sample of Al-Mg alloy reinforced with SiC particles by STEM using EDX and EELS detection. (a) TEM micrograph showing the interface between the Al matrix and the SiC particle; (b) Calculated concentration profiles from EDX measurements in the different elements across the interface: (c) EELS spectrum on the very small particles at the SiC/Mg_2Si interface showing the presence of the $MgAl_2O_4$ phase (Courtesy of M.C. Cheynet and J.J. Blandin, LTPCM and GPM2, France).

concentration left in the matrix enables the calculation of the solute concentration in the precipitates and of the volume fraction, at least in a binary alloy or a pseudo-binary alloy, as shown in Fig. 10.

In complicated systems including several elements, information about the relative fraction of elements can be obtained from anomalous SAXS (so-called ASAXS). This technique consists in measuring the SAXS spectra for various wavelengths of the x-rays, approaching the absorption edge for one of the elements in the alloy.

We have seen above that intensity of the scattered signal is proportional $|\Delta\rho|^2$, where $\Delta\rho$ is the difference in electronic density between the matrix and precipitates. The electronic density writes $\rho = f/\Omega$, where f is the scattering factor to x-rays

at the wavelength used in the experiment and Ω is the atomic volume. Far away from an absorption edge, f is almost equal to the atomic number Z. However, when approaching an edge, the scattering factor writes:

$$f = f_0 + f' + if''$$

where f' accounts for the absorption and f'' accounts for the fluorescence of x-rays (i is the imaginary number). The values of f' and f'' are tabulated in the Cromer–Liberman formalism [31], which can be used if the energy shift to the edge $\Delta E / E_e$ is larger than the energy dispersion of the beam used.

Thus one can monitor precisely the contrast between one atomic specie and the others in a given microstructure. This enables to obtain very useful information on the presence of this atom in a given family of precipitates. This was achieved for instance on a Al-Li-Cu-Mg alloy, where the authors were able to separate the contributions of the Al_3Li (δ') precipitation and the GPB zones containing (Al,Cu,Mg) by studying the copper edge (Fig. 13) [32]. In an other study, the same authors were able to show that the composition in copper of η' precipitates in an Al-Zn-Mg-Cu alloy was very small, ranging from 1 to 7 at% depending on the state of aging [33].

Of course, using this technique is quite delicate, since one needs to be able to control and change precisely the wavelength of x-rays near a given value which is the edge of interest. Thus the x-ray experiment should be equipped with a good
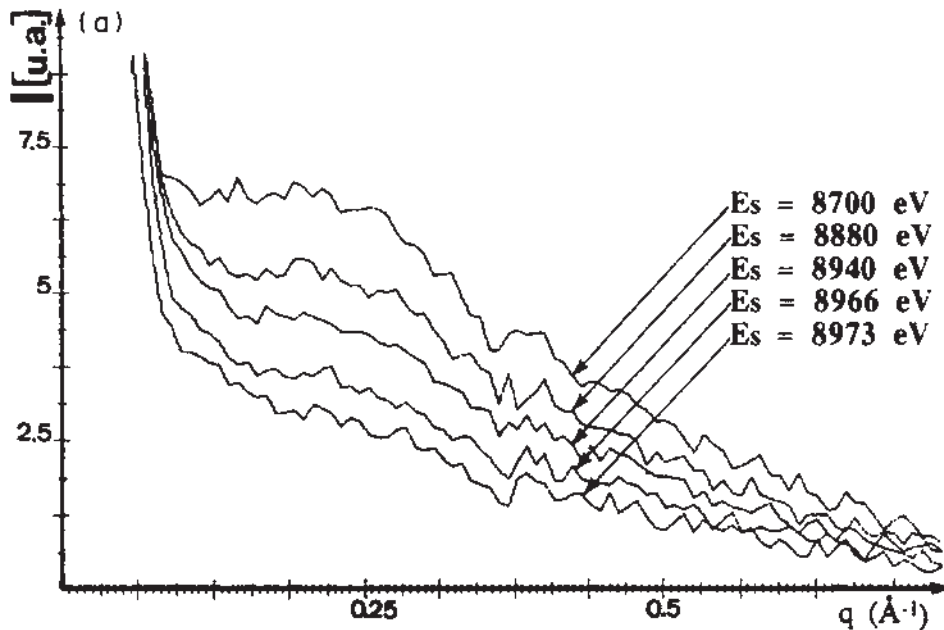
Figure 13 Anomalous x-ray scattering performed on an Al-Li-Cu-Mg alloy. The spectrum, representative of GPB zones, is clearly dependent on the energy of x-rays when approaching the $CuK\alpha$ edge. This enables to characterize the presence of copper in the precipitates (From Ref. 32).

monochromator. This is best achieved with synchrotron radiation sources, which provide a high flux of x-rays. Finally, one should note that this technique is not applicable to any elements: it works very well for elements such as Cu, Zn, Fe, bt is almost impossible to carry out on Mg or Si: at the edge energies for these elements, the x-rays do not traverse more than a few nanometers of aluminum. Thus it is not reasonable to use ASAXS for characterizing an Al-Mg-Si alloy, for instance: in this case, one should use neutrons scattering.

1.4.3 Tomographic Atom Probe

The ideal method for determining the precipitates composition if the recently developed Tomographic Atom Probe. This apparatus enables the full characterization of the microstructure with atom-scale precision, and thus can provide the composition of individual precipitates [26]. This powerful technique is of course extremely difficult to carry out, and can at this date only be performed in a few places.

TAP measurements provide an individual precipitate composition. This means that an idea of the variability of this composition from one precipitate to an other can be obtained, whereas this is very difficult to obtain by other means. In turn, it is delicate to obtain a good statistics with this technique: at best, a few dozens of precipitates can be measured for a given state in a reasonable time (see Fig. 14).

The main problem with the Tomographic Atom Probe is its uniqueness: indeed, it is the only technique which can provide such local results, and needs to be checked with other more indirect methods before one can be sure of its results. Measurements have been carried out on a 7000 series aluminum alloy designed for favoring the precipitation of the T′ phase [34]. For precipitates of an average diameter of 9 nm, the composition of this phase was fount by TAP to be 38% Mg, 37% Al and 25% Zn. This composition is very close to the equilibrium concentration of the T phase ($Mg_{32}(Al,Zn)_{49}$), which contains 39.5% Mg. This results suggests that the precision of this technique is very good even for precipitates having a small size.
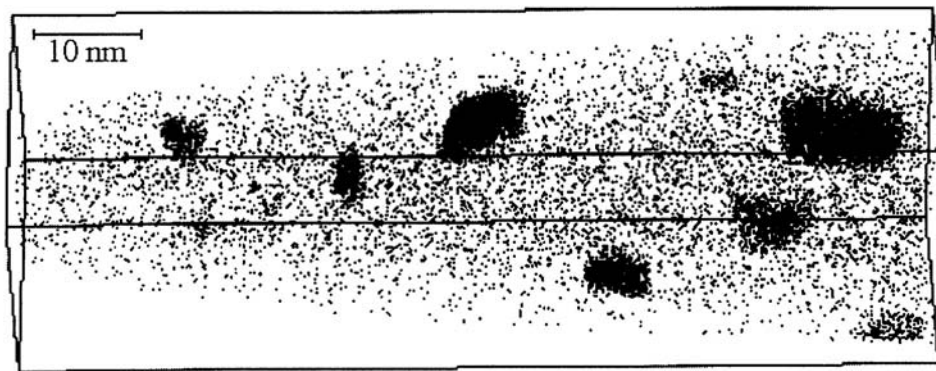
Figure 14 3D reconstruction of a Al-Zn-Mg needle analyzed with the Tomographic Atom Probe. Black dots represent solute atoms, precipitates are clearly identified and can then be analyzed in terms of size, chemistry and volume fraction (Courtesy of A. Bigot, University of Rouen, France).

However, an other study on a ternary Al-Zn-Mg, this time promoting the precipitation of the η' and η phase, has lead to less clear-cut results. First, the measured compositions in η' were found to be very far from the equilibrium composition of η ($MgZn_2$): more than 50% of aluminum was found in the early stages of aging (precipitates 3 nm in diameter) [35]. Secondly, a parallel study in Small Angle X-Ray Scattering was in disagreement with the TAP results in these early stages, whereas the two studies were consistent in later stages where precipitates were larger and did not contain more than 15% Al.

The possible artifacts of this method are numerous, perhaps the most relevant in this case is the local magnification effect: at the end of a TAP needle, the radius of curvature of a precipitate can be substantially different than the radius of curvature of the matrix, due to capillary effects. In this case, there can be some mixing between the composition measurements in the matrix and in the precipitates.

In summary, Tomographic Atom Probe results for the determination of precipitates composition are extremely useful but should still be checked with other experimental techniques when the size of the measured objects is very small (less than 10 nm in diameter). However, although there is still a doubt about the high aluminum concentrations which are sometimes found in TAP studies, the measurement of the ratios between various solute species is thought to be very reliable.

1.5 Precipitation Sequence

Precipitation hardening aluminum alloys are characterized by very complex precipitation sequences, which include several metastable phases. During usual thermomechanical sequences, several of these phases appear sequentially. The way these phases nucleate in the presence of the precedent phase is complicated: in Al-Zn-Mg alloys, for instance, the presence of GP zones promotes the nucleation of η', whereas in Al-Mg-Si alloys, the presence of GP zones or solute clusters inhibits the precipitation of β''. The privileged technique to study these precipitation sequences is Differential Scanning Calorimetry, which measures the heat exchange between a sample and its environment during a thermal sequence.

DSC measurements can be used in several ways, which will be detailed below. However, one should keep in mind that the outcome of it can usually only be analyzed in a qualitative manner, and the validity of the interpretations should be checked carefully with other experimental techniques (e.g. the type of phases present at one stage of the thermal sequence can be checked by Transmission Electron Microscopy).

DSC can first be used as a probe of the microstructure present at one stage of a thermomechanical sequence. In this case, the thermal sequence will be a fast heating ramp ($>50°C/min$), so that no phase transformation can occur during the heating ramp except the dissolution of the phases present at the beginning of the test. The type of phases present at the beginning can be then analyzed from the dissolution temperature. This method is illustrated in Fig. 15 in 6 Al-Zn-Mg alloys containing 5 wt% Zn, 2 wt% Mg and increasing amounts of copper ranging from 0% to 1.6 wt% [36]. One dissolution peak occurs at constant temperature for all alloys and is interpreted as corresponding to Al-Zn-Mg GP zones. A second dissolution peak occurs at higher temperature in the high copper containing alloys and corresponds to a second family of GP zones containing copper.
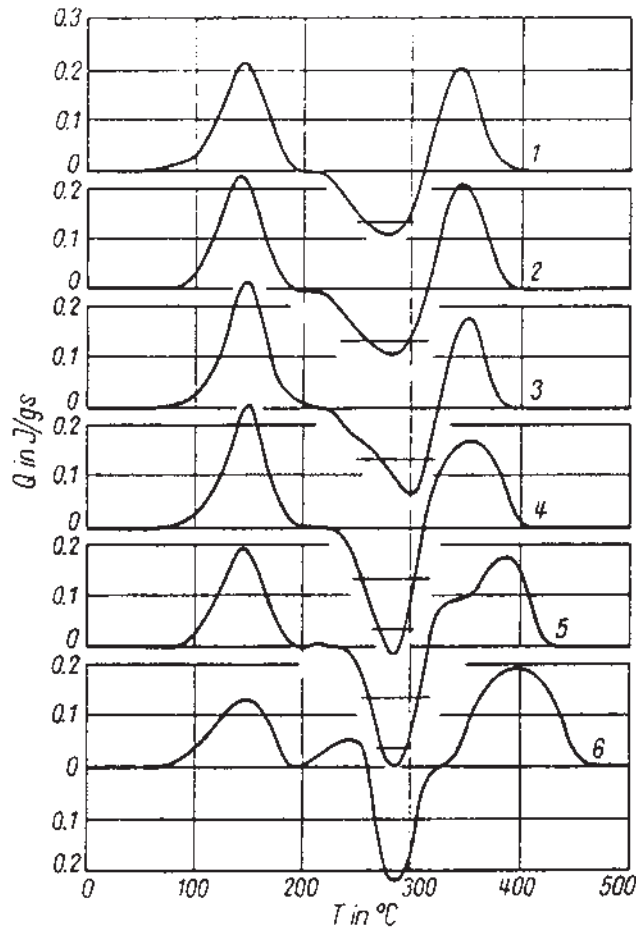
Figure 15 DSC thermograms for Al-Zn-Mg alloys containing an increasing amount of copper, performed after a quench and 40 days at room temperature. The heating rate is 80°C/min. One can see clearly the transition from one low-temperature dissolution peak for alloys 1–5 to two dissolution peaks for alloy 6 (From Ref. 36).

A second way of using DSC is to simulate the heat treatment which is used for obtaining the desired precipitation structure. In this case, one can vary the experimental conditions of the initial conditions and of the heating ramp to observe the induced changes in phase transformations. For this purpose, slow heating ramps need to be used, since the precipitation treatments in aluminum alloys are usually carried out with slow ramps. Depending on the equipment used, reasonable data can be obtained with heating ramps down to 0.3–1°C/min. Of course, ideally one would like to carry out isothermal treatments in the DSC (even possibly after a heating ramp), in order to simulate perfectly the actual thermal history of the material studied. However, extremely sensitive equipment need to be used in order to do so, since after a few hours of heat treatment, the heat exchange signal between

the sample and its environment becomes ridiculously small. Such equipment exists, and has been used to study very low-signal transformations, such as grain growth. However it is not very widely spread such as conventional DSC. An example of such an experimental method is shown in Fig. 16, where the thermograms of an Al-Zn-Mg alloys are represented [14]. The heating ramp is varied from 0.5°C/min, C/min, representing the industrial heat treatment, and 5°C/min, in order to characterize the effect of the heating ramp on phase transformations. One can observe that the low-temperature dissolution peak of GP zones is much more extensive if the heating ramp is faster, resulting in a less well-defined η' precipitation peak at higher temperature. Thus this experiment enables to state the detrimental effect of increasing the heating ramp on nucleation of the hardening η' phase.

Finally, DSC can be used to study the kinetics of phase transformations, and notably to determine activation energies for transformations. However, in aluminum alloys this is often delicate, since a reasonable estimate of such quantities can only be obtained in the case of an isolated DSC peak, whereas in most practical cases DSC scans show interrelated peaks, exotherms following endotherms, etc. Moreover, such studies should be carried out on transformations which nature do not depend on the heating ramp: this is rarely true in aluminum alloys! One example of this methodology is shown in Fig. 17. Figure 17(a) shows DSC thermograms performed on alloy 2219-T31 at rates between 1 and 20K/min [37]. These thermograms show two reasonably separated peaks, one corresponding to dissolution of GP zones, and the second to precipitation of the θ' phase. Integration of the second peak with temperature enables to plot the evolution of transformed
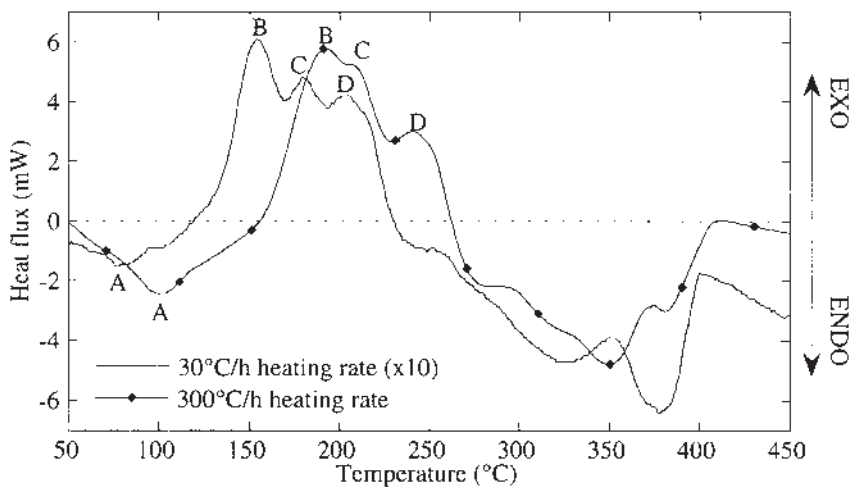
Figure 16 DSC thermograms performed on an Al-Zn-Mg alloy after quench and 3 days at room temperature. The faster heating ramp increases the magnitude of the low-temperature GP zones dissolution peak (A), resulting in a less well-defined precipitation peak of the hardening phase η' (B) (From Ref. 14).
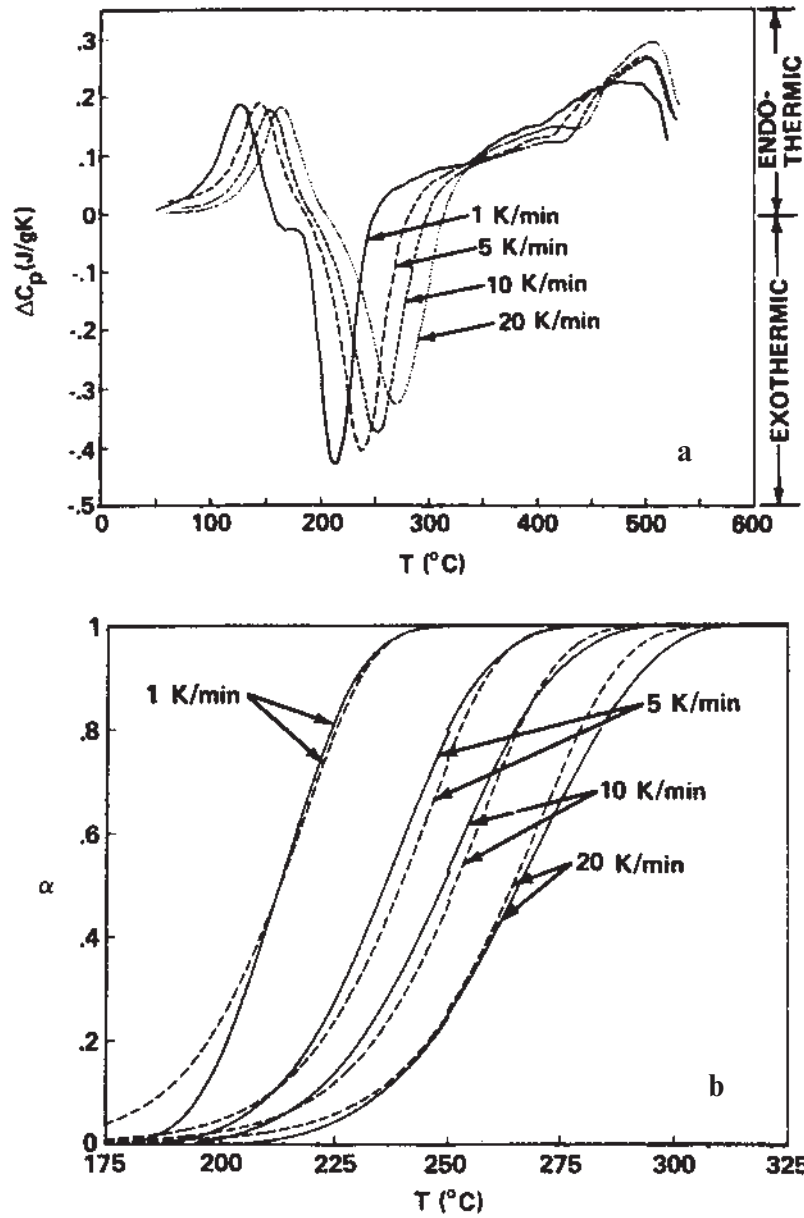
Figure 17 (a) DSC thermograms of alloy 2219-T31 as a function of heating rate. (b) Interpretation of the transformed fraction in terms of Avrami formalism, resulting in an activation neergy of 117 kJ/mol and $n = 1.1$ (From Ref. 37).

fraction α with temperature for the different heating rates. The starting point of the analysis is a general relationship between the transformed fraction α, the reaction rate coefficient $k(T)$ and time t:

$$f(\alpha) = k(T)\,t$$

Assuming that the rate coefficient follows an Arrhenius equation, and that the heating rate $\phi = \mathrm{d}T/\mathrm{d}t$ is constant, one obtains:

$$f(\alpha) = \frac{AE}{R\phi}p(x)$$

where A is the pre-exponential term in the Arrhenius equation, E is the activation energy, R the universal gas constant and

$$p(x) = \int_0^\infty \frac{\exp(-x)}{x^2}\,\mathrm{d}x$$

$f(\alpha)$ is finally often fitted to an Avrami law:

$$f(\alpha) = [-1n(-\alpha)]^{1/n}$$

Such a fit for the precipitation of θ' is shown in Fig. 17(b), and results in the following value for the parameters: $E = 117$ kJ/mol, $A = 3.10^9$ s-1 and $n = 1.1$. The activation energy is close to the one of copper diffusion in aluminum, and the value $n = 1.1$ 1.1 indicates some sort of growth-controlled reaction.

1.6 Other Problems

A number of other characteristics of precipitates often need to be characterized; however, it is not the purpose of this chapter to be completely exhaustive, so we will only give a brief survey of how less widespread precipitate characteristics can be characterized.

- Coherency strains around coherent or semi-coherent precipitate can be analyzed by image analysis of TEM images taken in well controlled imaging conditions.
- The study of the interface between the precipitates and the matrix is best studied by High Resolution Electron Microscopy (HREM).
- The internal chemical order in the precipitates can be studied by some x-ray techniques such as EXAFS (Extended X-ray Absorption Fine Structure) and XANES (X-ray Absorption Near Edge Spectroscopy), which both study the local dependence of the transmission to x-rays of a sample as a function of wavelength near an absorption edge.

2 CHARACTERIZATION OF SECOND PHASE PARTICLES

The characterization of large second-phase particles in aluminum alloys needs to meet a number or requirements, which are dictated by the aspects of the process and properties which may be controlled by their presence. For instance, second phase

particles play an important role in the occurrence of recrystallization and on the resulting texture, and determine significantly the resistance of the material to fatigue by promoting the nucleation of microcracks in the material.

First, the type of phase needs to be determined (crystallography and composition) for these particles, in order to control the temperature domain where they may appear in the microstructure. Another important parameter is obviously the volume fraction of the various phases, and their size. However, more important is their geometry, since a number of them (e.g. Al_7Cu_2Fe) have very complicated shapes. It is well known that the aspect ratio of hard particles can influence very significantly their role on the microstructure formation. Finally, one other characteristic of these phases is that they are not distributed homogeneously in space, and it is important to characterize both their location as compared to other defects (and particularly grain and sub-grain boundaries), as well as their degree of clustering.

2.1   Phase Type, Crystallographic Structure, Composition

2.1.1   Optical Microscopy

Optical microscopy offers an easy and efficient way of characterizing the type of precipitates. Notably, in well known alloy systems, the various phases can be chemically etched to different colors, and thus identified. Fig. 18 shows for instance the use of the Keller etch to reveal the intermetallic phase Mg_2Si in an Al-Mg-Mn alloy.
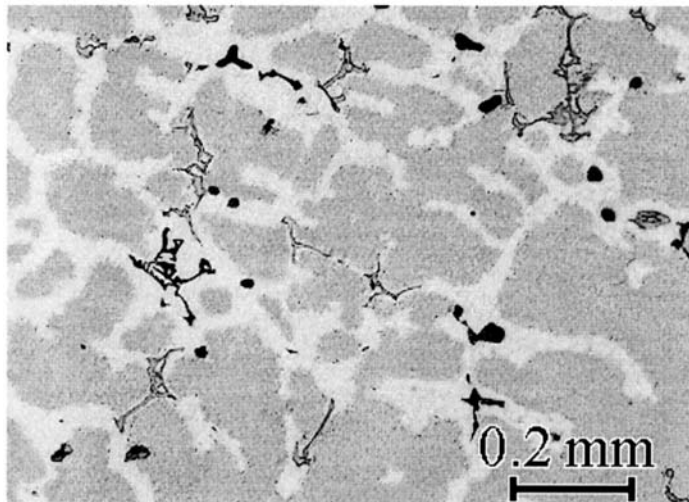
Figure 18   Optical micrograph of an Al-Mg-Mn alloy (5383) after Keller etching (HNO_3, HCl, HF, H_2O). The Mg_2Si phase appears in gray and the micro-porosity in black (Courtesy of S. Péron, LTPCM, France).

2.1.2   Scanning Electron Microscopy

The SEM is probably the most efficient way to characterize the type of second phase particles in classical materials where the phases have been identified in terms of composition and crystallographic structure. The sample preparation is easy and fast, and the apparatus is well adapted to the size of the particles (a few microns).

The type of second-phase particle is determined from its composition through an x-ray analysis (Energy Dispersive X-ray detector, or Wavelength Dispersive X-ray detector for better resolution in energy). The volume of the material which participates to the emission of x-rays under the electron beam is about $1\mu m$ in diameter for aluminum under standard operating conditions (20 kV). Thus it is not reasonable to characterize quantitatively the composition of second phase particles which smallest size is smaller than $1\ \mu m$.

For particles which are large enough, truly quantitative analysis of the composition can only be obtained by comparison of the x-ray signal with spectra of pure materials analyzed under the same experimental conditions. In this case, good precision on the compositions can be obtained ($\pm 1\%$ in absolute values).

However, a qualitative analysis is often sufficient to enable the determination of the phase type, in systems where these are well known and characterized. For instance, if one finds Fe, Cu and Al in a second phase particle in a 7000 series aluminum alloy, as shown in Fig. 19, the particle is definitely Al_7Cu_2Fe. In a 3000 series alloy, a qualitative analysis is often sufficient to determine if a particle is Al_6Mn
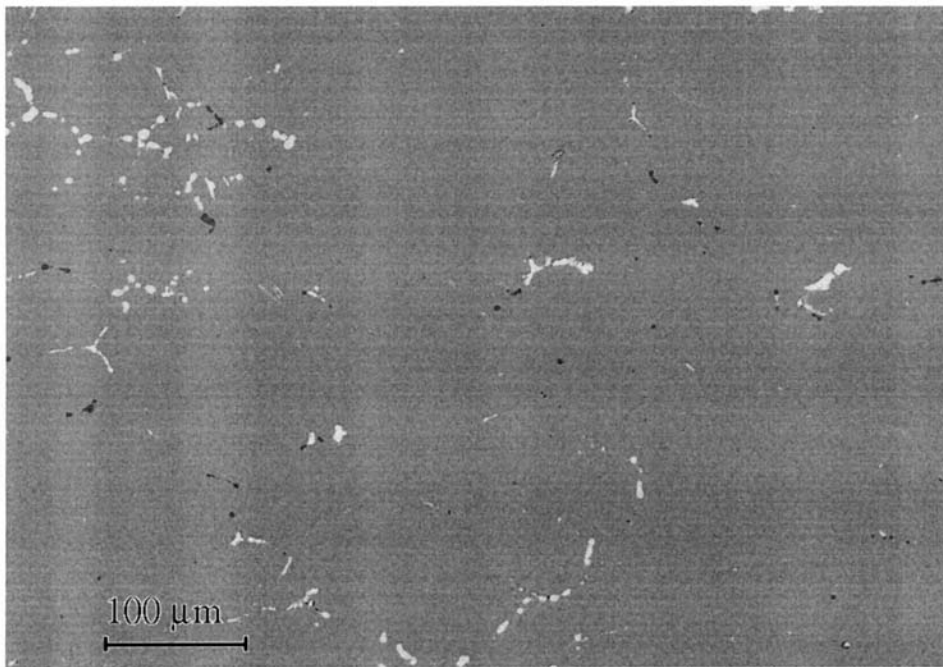
Figure 19   Intermetallic particles in a 7050 alloy. White particles consist in Al_7Cu_2Fe and $AlCu_2Mg$ particles; black particles in Mg_2Si particles (Courtesy of D. Dumont, LTPCM, France).

or $Al_{12}Mn$. Qualitative analysis in the SEM is simply carried out by fitting the measured spectrum with peaks calculated from standards stored in the computer. One point which is important in a qualitative chemical analysis is that the computer system will always complete the composition to 100%, even if the user forgets an important element in the analysis. Therefore, this method is best to measure roughly the ratio between the compositions in different elements, and not for self-consistent absolute measurements.

The SEM can also be used in x-ray mapping mode in order to make a cartography of the different atomic species present in the material. For this purpose some areas of the x-ray spectrum, corresponding to the different species, are recorded for each point of the map. Figure 20 shows an example of such maps for Al, Si, Cu and Fe in a 7000 series aluminum alloy. One can clearly distinguish Al_7Cu_2Fe and Mg_2Si phases. This result was obtained with a Field-Emission Gun SEM, which provides much higher resolution than a conventional SEM.
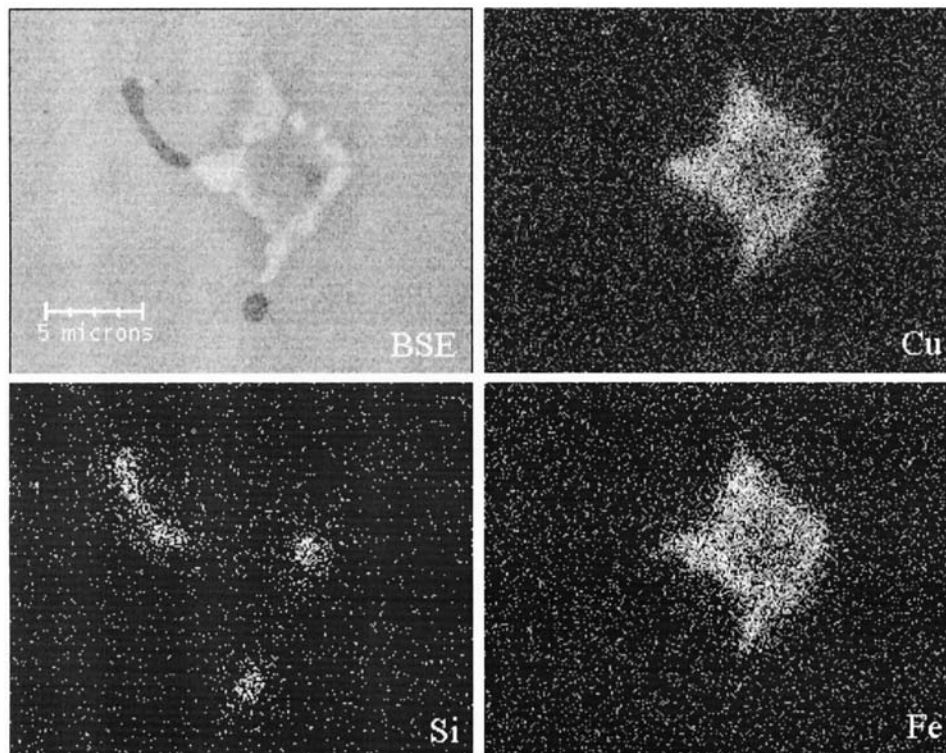
Figure 20   Chemical characterization of a second phase particle in AA7050 alloy. The backscattered electrons image shows that the particle is composed of two phases. The various maps of composition show that the central part of the particle contains Cu and Fe (composition: Al_7Cu_2Fe). The particle at the periphery contains Si and is Mg_2Si (Courtesy of D. Dumont, LTPCM, France. Images obtained with the help of J.-L. Pastol, CECM, France).

2.1.3 Transmission Electron Microscopy

The use of Transmission Electron Microscopy becomes necessary in some specific cases: when precipitates are too small to be detected by the SEM (e.g. dispersoïds), when the crystallographic structure of the precipitate is important to know, or when one wants to characterize quantitatively the composition of sub-micron sized particles.

The crystallographic nature of the precipitates can be inferred from diffraction patterns. However, a good sample preparation is necessary for this purpose, since second phase particles, of a different chemical nature than the matrix, are often not thinned as efficiently as the matrix, and therefore are very thick and difficult to observe in the microscope.

The chemistry of large particles is very straightforward to obtain in a TEM equipped with an EDX detector. For really quantitative measurements, however, two conditions need to be fulfilled: the particle needs to be separated from the matrix, i.e. the electron beam should not cross any matrix but only the precipitate, and the intensities of the x-ray emission of the different atomic species have to be calibrated with reference samples.

2.2 Geometrical Information: Experiments

Second phase particles in aluminum alloys have some specific geometrical characteristics which make their analysis particularly difficult to formalize. They are often non-symmetrical and complicated in shape with a high surface-to-volume ratio, they are not evenly distributed in space, they are clustered, and their volume fraction is very small.

Several experimental methods exist which can provide many data with all this information. However, the most difficult is sometimes to find the relevant mathematical analysis which can provide a simple parameter which is well related to a given property (e.g. the degree of clustering of particles and the fatigue life).

2.2.1 Optical Microscopy

Optical microscopy offers god potential for the observation of large areas of the material. One useful technique is to observe on the same sample several features of the microstructure by applying sequentially different etching procedures. One can then determine if the particles are lying mostly on the grain or subgrain boundaries, for instance.

The main problem with optical microscopy is the difficulty to carry out quantitative measurements by direct image analysis. This is largely due to the non-uniformity of the light in the observation area of the microscopy, which results in non-uniform light on the photographs.

2.2.2 Scanning Electron Microscopy

The most useful technique is probably the SEM. It offers good resolution, uniform light and digital ready-to-analyze data.

The usual observation mode is of course back-scattered electrons. They offer high contrast between the particles and matrix if they have a significantly different atomic weight. Moreover, in this mode the contrast is extremely uniform both

for the matrix and the particles (if their composition is constant, of course), and therefore it is very easy to perform automatic image analysis. Figure 19 shows for instance intermetallic particles of Al_7Cu_2Fe in a 7000 series alloy.

Recently, the development of Field Emission Guns for SEMs has increased drastically the resolution of this technique. It is now possible to image in Back-Scattered mode precipitates smaller than 100 nm.

2.2.3 From 2D to 3D: Reconstruction Techniques

One important problem in utilizing the precedent experimental techniques to characterize second phase particles is the fact that they provide only a 2-dimensions section of the microstructure. One good example of the mistakes that can be done is the microstructure of an eutectic Sr-modified Al-Si alloy. This alloy consists in a network of fibrous eutectic, as shown in Fig. 21(b), where the aluminum matrix has been etched away. However, an observation of the polished surface in the SEM shows a microstructure which looks very similar to a distribution of globules (Fig. 21(a)).

Through this example, we can see that due to the nature of second phase particles, which have a complicated, non-uniform, high aspect ratio geometry, important mistakes can be made in the characterization of precipitates if one keeps the analysis to 2 dimensions.

A method for obtaining data on the microstructure from 2 dimensions sections consists in taking sequential images at various depths, the sample being polished down to the desired thickness between each photograph. Software can then be used to reconstruct the volume from the sequence of images.

This technique is obviously very cumbersome, and offers discontinuous reconstruction of the volume of the material. Therefore it is used only relatively rarely, when 3D information is absolutely needed.

2.2.4 True 3D: the High Resolution X-ray Tomography

Recently, the development of high brilliance x-ray sources such as the European Synchrotron Radiation Facility (ESRF) in Grenoble, as well as the development of powerful computers, has lead to a new characterization technique of the 3D structure of bulk materials. High resolution computed tomography is a non-destructive technique that can be used to obtain images of the interior of micro-heterogeneous materials such as metal matrix composites or alloys containing micron-size heterogeneous particles. The principle is described briefly as follows (more details can be found in [38]). A beam of x-rays fired at a rotating sample produces a 2-dimensional projection of the sample which is recorded on a charge coupled device (CCD) detector set behind the sample. In classical tomography (attenuation tomography), the contrast on 2-dimensional projections arises from the attenuation coefficients of each phase present in the material. Once recorded, the different 2-dimensional projections are combined mathematically through a filtered back projection software to re-create a 3-dimensional map of the distribution of attenuation coefficients in the material. $\mu(x,y,z)$. The final result is a numerical 3-dimensional grey level image divided into volume elements called voxels. When the attenuation coefficients of the phases in the material are very similar, (Al/SiC, for example), the phase contrast tomography obtained with coherent synchrotron x-ray sources provides reconstructed images with improved contrast. The experimental set up is similar to the attenuation mode, only the detector is located
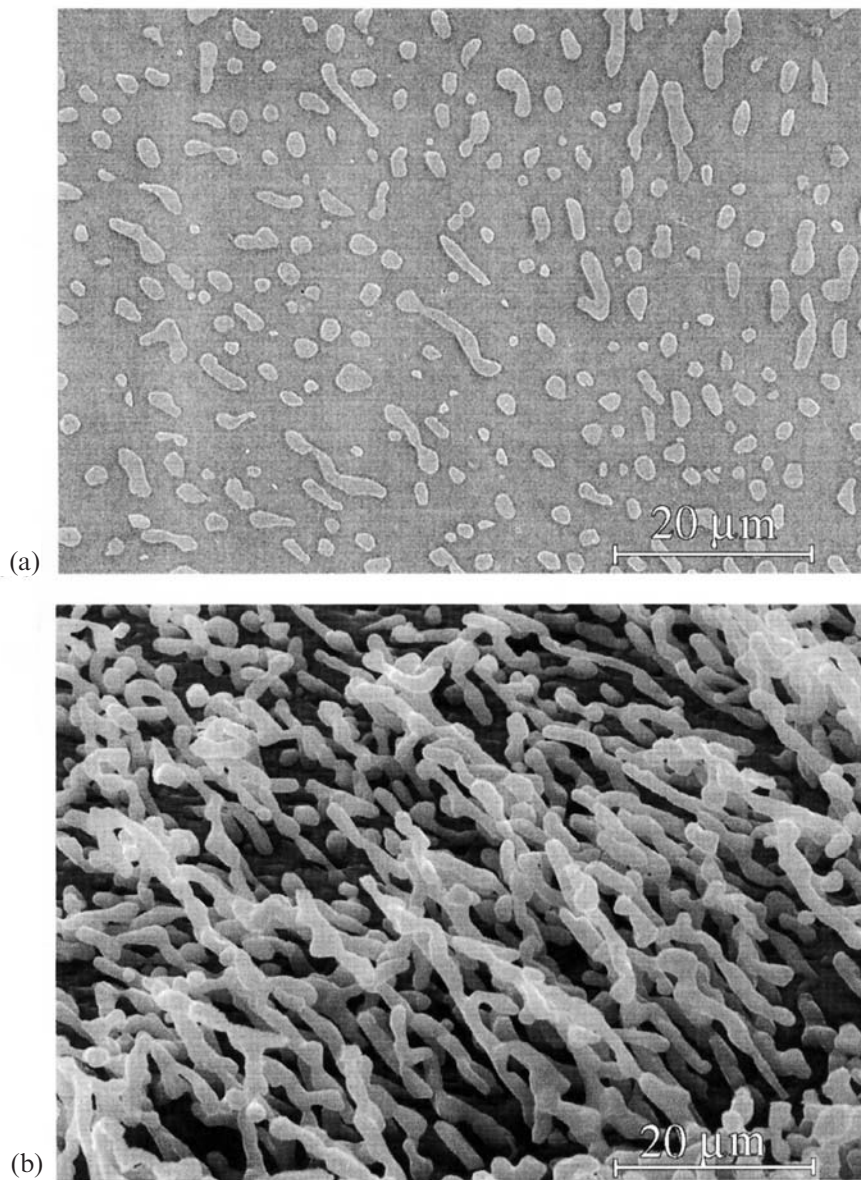
Figure 21 Eutectic structure in a Sr-modified Al-Si alloy of eutectic composition. (a) The backscattered electrons image of the polished surface gives no indication of a continuous fibrous structure of the Si phase; (b) this fibrous structure is revealed by deep etching of the aluminum phase and subsequent observation in the SEM (Courtesy of G. Guiglionda, University of British Columbia, Canada).

further away from the object at a distance which depends on the x-ray energy and on the size of the microheterogeneity to be imaged. Diffraction occurs whenever the x-ray beam encounters an interface. The deflected and non-deflected waves interfere giving Fraunhofer and Fresnel fringes and an "edge contrast" outline image [39]. A

classical algorithm is then used to computationally extract a high quality image from the 2-dimensional projections. An example of such a reconstruction is shown in Fig. 22, in the case of secondary phase particles (mainly Fe-containing phases) present in an Al-Zn-Mg-Cu alloy. From such a 3D reconstruction, it is now possible to compute the detailed shape of the particles and to determine the degree of clustering in the various directions of the sample.

2.3   Geometrical Information: Analysis

2.3.1   Standard Output of Image Analysis

When image analysis is carried out with a standard software, the output is the characteristics of each single particle detected, and then the statistical analysis of the complete distribution. Some useful parameters can be detailed as follows:

- the Ferret diameter is the diameter of a disk of equivalent area
- the elongation is the aspect ratio (the ratio between the major axis size and the minor axis size)
- the perimeter which can describe the complexity of the shape of the particle

300 μm
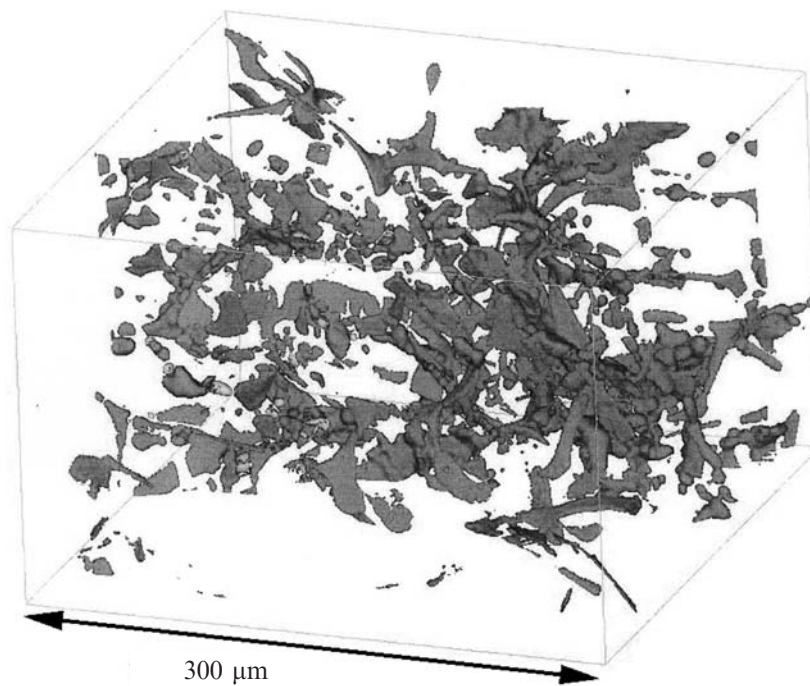
Figure 22   Characterization of the 3-D structure of second phase particles in a 7000 alloy using high resolution x-ray tomography. Second phase particles appear in red and porosity appears in green (Courtesy of E. Maire and J.-Y. Buffière, GEMPPM, France, and C. Sigli, Pechiney – CRV, France. Image obtained with the help of J. Baruchel, P. Cloetens, and W. Ludwig at the ID19 beamline of ESRF, France).
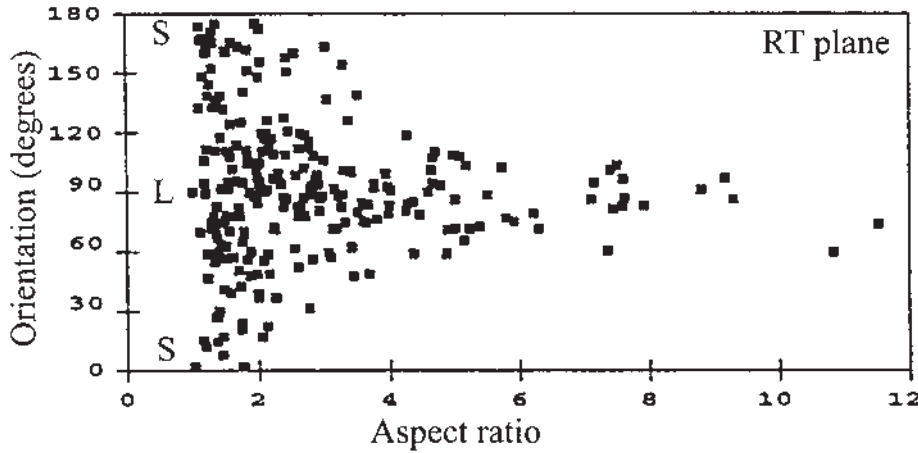
Figure 23 Correlation between the orientation of the major axis of second phase particles and their aspect ratio in a 7010 aluminum alloy. It appears clearly that elongated particles exist only parallel to the rolling plane. (Courtesy Denis Solas, GPM2, France)

For a more detailed description of the possibilities one can refer to the free image analysis software ImageTool developed by the University of Texas Health Science Center (http://www.ddsdx.uthscsa.edu/dig/download.html).

The area fraction is simply calculated as the ratio between the number of pixels inside the detected particles divided by the total number of pixels.

These parameters enable the determination of average values for the particles geometry (average size, elongation, interface with the matrix). An interesting analysis is for instance the degree of correlation between the orientation of the main axis of the particle and the geometry of the sample in a rolled material, as shown in Fig. 23. This analysis shows very clearly that elongated particles can only survive if they are parallel to the rolling plane. If they have a large angle with the rolling plane at the beginning of the rolling process, they are broken down in smaller, less elongated particles. This anisotropy of the particles shapes can have important consequences on crack initiation in fatigue, for instance.

However these parameters do not give any information on the distribution of the particles, which can also be an important parameter for the property control.

2.3.2 Voronoi Cell Analysis

A useful method for characterizing the spatial distribution of second-phase particles is to construct Voronoi cells around the center of gravity of each particle. This is achieved by constructing the median between each two neighboring particles. The Voronoi cell around one particle is then given by the intersection of all the medians which are nearest to this particle, as shown in Fig. 24 for intermetallic particles in a 7000 series aluminum alloy. The degree of clustering of the particles can be detected by considering the ratio between the standard deviation and the average value of the surface of the Voronoi cells. A random distribution of particles leads to a ratio of 0.5, and the degree of clustering increases when this ratio increases.
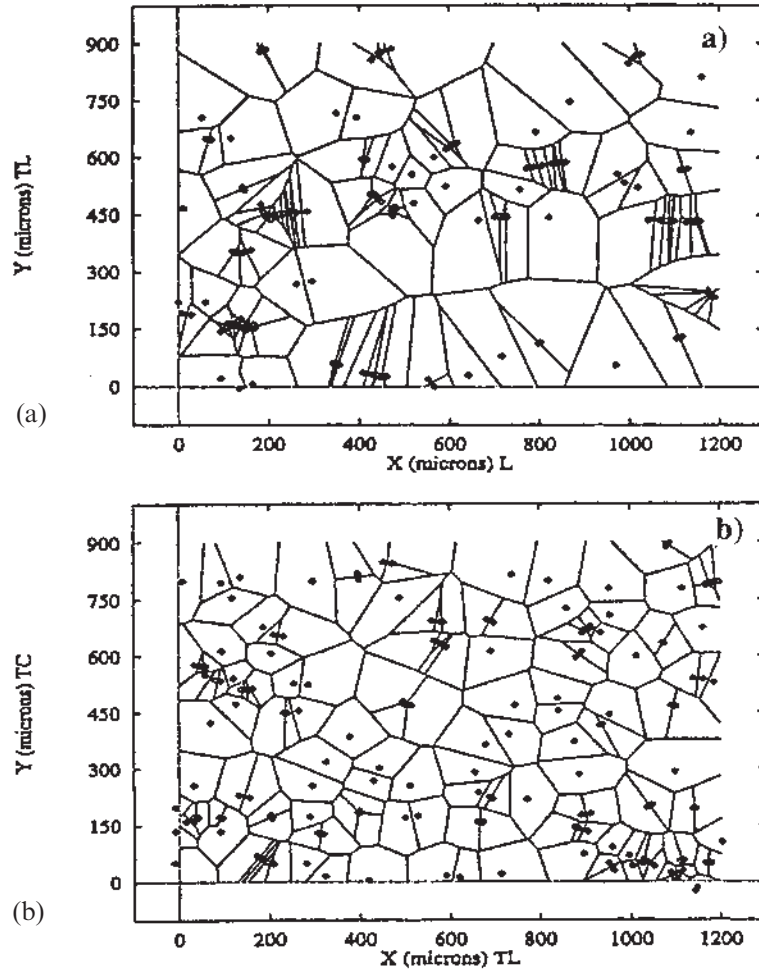
Figure 24 Voronoi cells constructed after SEM images (in Back-scattered mode) of a 7010 alloy. Each point represents a second phase particle. (a) Shows a section of the rolling plane; and (b) A section of the Short-Transverse plane. (c) Shows the statistical analysis of the Voronoi cells. The ratio between the standard deviation and the average values of the surface area gives an indication of the degree of clustering, which is much higher in the rolling plane. (Courtesy Denis Solas, GPM2, France)

For instance, the microstructure represented in Fig. 24(a) is highly clustered with a ratio of 1.4, whereas the structure in Fig. 24(b) is moderately clustered with a ratio of 0.7.

REFERENCES

1. P. Villars and L. D. Calvert, Eds., *Pearson's Handbook of Crystallographic Data for Intermetallic Phases*, ASM, 1985.
2. X. Z. Li, V. Hansen, J. Gjønnes, and L. R. Wallenberg, Acta Mater., 1999, 47(9), p. 2651.
3. A. Deschamps, Y. Bréchet, P. Guyot, and F. Livet, Zeit. für Metallk., 1997, 88(8), p. 601.
4. H. P. Degischer, W. Lacom, A. Zahra, and C. Y. Zahra, Zeit. für Metallk., 1980, 71, p. 231.
5. P. Donnadieu, F. Carsughi, A. Redjaimia, C. Diot, and G. Lapasset, J. Appl. Cryst., 1998, 31, p. 212.
6. M. H. Jacobs, Phil. Mag., 1972, 26, p. 1.
7. S. P. Ringer, B. C. Muddle, and I. J. Polmear, Metall. Trans., 1995, 26A, p. 1659.
8. W. A. Cassada, G. J. Shiflet, and E. A. Starke Jr., Metall. Trans., 1991, A22, p. 299.
9. J. D. Kim and J. K. Park, Metall. Trans., 1993, A24, p. 2613.
10. G. A. Edwards, G. L. Dunlop, and M. J. Couper, in: Proceedings of the 4th ICAA, Atlanta, Georgia Institute of Technology, 644, 1994.
11. O. Glatter and O. Kratky, *Small Angle x-ray Scattering*, Academic Press, London, 1982.
12. D. Dumont, A. Deschamps, Y. Bréchet, and C. Sigli, Proceedings of Euromat 99, Munchen, 1999.
13. S. Abis, A. Bœuf, R. Caciuffo, P. Fiorini, M. Magnani, S. Melone, F. Rustichelli, and M. Stefanon, J. of Nuclear Mat. 1985, 135, p. 181.
14. A. Deschamps, F. Livet, and Y. Bréchet, Acta Mater., 1999, 47(1), p. 281.
15. A. Deschamps and Y. Bréchet, Scripta Mater., 1998, 39(11), p. 1517.
16. A. K. Vasudevan and R. D. Doherty, Acta Metall. 1987, 35(6), p. 1193.
17. R. T. De Hoff and F. N. Rhines, *Quantitative Microscopy*, McGraw-Hill, New York, 1968.
18. F. R. Fickett, Cryogenics, 1971, 11, p. 349.
19. N. Luiggi, J. P. Simon, and P. Guyot, J. Phys. Fr., 1980, 10, p. 865.
20. D. Godard, Ph.D. Thesis, Institut National Polytechnique de Lorraine, France, 1999.
21. A. Deschamps, Y. Brechet, and F. Livet, Mat. Sci. Techn., 1999, 15(9), p. 993.
22. P. Donnadieu, M. Roux-Michollet, and V. Chastagnier, Phil Mag., A 1999, 79(6), p. 1347.
23. A. Tekin, A. Uguz, and J. W. Martin, Mater. Characterization, 1990, 25(1), p. 99.
24. S. M. Allen, Phil. Mag., 1981, 43, p. 325.
25. P. M. Kelly, A. Jostons, R. G. Blake, and J. G. Napier, Phys. Stat. Sol. (a), 1975, 31, p. 771.
26. D. Blavette, A. Bostel, J. M. Sarran, B. Deconihaut, and A. Menand, Nature, 1993, 363, p. 432.
27. D. B. Williams, J. I. Goldstein, and D. E. Newbury, Eds., *X-ray Spectrometry in Electron Beam Instruments*, Plenum, New York, 1995.
28. D. B. Williams and C. B. Carter, *Transmission Electron Microscopy, a Textbook for Materials Science*, Plenum, New York, 1996.
29. R. F. Egerton, *Electron Energy-loss Spectroscopy in the Electron Microscopy*, Plenum, New-York, 2nd Edn., 1996.
30. M. M. Disko, C. C. Ahn, and B. Fultz, Eds., *Transmission Electron Energy Loss Spectrometry in Materials Sciences*, Minerals, Metals and Metarials Society, Warrendale, PA, 1992.

31. D. T. Cromer and D. A. Liberman, Acta Cristall., 1981, A37, p. 267.
32. P. Gomiero, F. Livet, O. Lyon, and J. P. Simon, Acta Metall. Mater., 1991, 39(12), p. 3007.
33. J. P. Simon and O. Lyon, ''Annales de Physique,'' Colloque C3, 1995, 20, p. 127.
34. A. Bigot, Ph.D. Thesis, Université de Rouen, France, 1998.
35. A. Bigot, F. Danoix, P. Auger, D. Blavette, and A. Reeves, Mat. Sci. Forum, 1996, 217–222, p. 695.
36. P. Nagy, E. Kovacs-Csetenyi, T. Turmezey, G. Honyek, I. Kovacs, and J. Lendvai, Aluminium, 1984, 84, p. 368.
37. J. M. Papazian, Metall. Trans., 1982, 13, p. 761.
38. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, 1988.
39. P. Cloetens, M. Pateyron-Salomé, J. Y. Buffière, G. Peix, J. Baruchel, F. Peyrin, and M. Schlenker, J. Appl. Phys. 1999, 81(9), p. 1.

6

Work Hardening, Recovery, Recrystallization, and Grain Growth

ANGELO FERNANDO PADILHA and RONALD LESLEY PLAUT

University of São Paulo, São Paulo, Brazil

SUMMARY

This chapter reviews the main aspects of the physical metallurgy related to the work hardening, recovery, recrystallization, and grain growth of aluminum and its alloys. Initially, it is attempted to classify the main alloying elements and related microstructures. The main alloying elements and impurities are classified in terms of their solubility and solidification types (eutectic, peritectic, and monotectic) that occur. Second-phase particles that occur in these alloys are classified into five major types, according to their nature, origin, and main characteristics. In the following, we deal with the effects of solute atoms both in solid solution as well as in the form of particles, related to the amount and distribution of dislocations after the plastic deformation. The importance of the recovery processes on the softening that occurs during annealing after plastic deformation is also emphasized. Later on, the recrystallization is analyzed in terms of nucleation and growth, giving major attention to the role played by second-phase particles. The effect of second-phase particles on grain growth and secondary recrystallization as well as the resulting textures due to deformation and annealing are also analyzed. Finally, some brief comments are made related to the hot working and softening processes.

1   INTRODUCTION

The major part of aluminum production is related to plane products (plate, sheet, and foils) and extrusions, with a strong predominance of the former [1]. Over 85% of the fabricated aluminum products are rolled as sheet and foil. The understanding of the phenomena such as work hardening, recovery, recrystallization, and grain growth is essential not only to obtain these products but also to control their microstructures, textures and orientation relationships between the phases and, hence, optimize their properties and performance. The phenomena above mentioned, in turn, strongly depend on the alloy composition and on the quantity, size, morphology, and distribution of phases present in the microstructure. In the first part, the alloying elements and the phases present in the aluminum alloys shall be classified into characteristic types. In the following will be discussed, separately, the phenomena of work hardening, recovery, recrystallization, grain growth, and texture control.

The large-scale production processes for aluminum (Bayer and Hall-Héroult) have been available for more than 100 years, since the last decades of the XIX century, and the production of aluminum (about 17 million tons) is only surpassed, nowadays, among the metallic materials, by steel production (about 760 million tons). Despite this, the number of most employed commercial aluminum alloys is relatively small, if compared to steels. Table 1 presents the chemical composition of some of the main aluminum wrought alloys.

Table 2 illustrates that the solubility of the main alloying elements is relatively limited, mainly if we consider that iron and silicon are coming from the processing and are not (depending on the alloy) properly alloying elements. Baumann and Sanders [2] prefer to classify the alloying elements, in accordance to their solubility, into three classes:

(i)   High solubility solutes (>10 at.%): Zn; Ag; Mg; Li.
(ii) Intermediate solubility solutes (>1 and < 10 at.%): Ga; Ge; Cu; Si.
(iii) Low solubility solutes (< 1 at.%): all others.

The aluminum alloying elements can also be classified, in relation to their solidification type, into three other classes:

Table 1   Chemical Composition (wt%) of the Main Aluminum Wrought Alloys

Alloy	Si	Fe	Cu	Mn	Mg	Zn	Cr	Ti	Al
1050	0.25	0.40	0.05	0.05	0.05	0.05	—	0.03	99.50
1100	Si+Fe = 1.0		0.05–0.20	0.05	—	0.10	—	—	99.00
2024	0.50	0.50	3.8–4.9	0.30–0.9	1.20–1.8	0.25	0.10	0.15	Rem.
3003	0.6	0.7	0.05–0.20	1.0–1.5	—	0.10	—	—	Rem.
3004	0.3	0.7	0.25	1.0–1.5	0.8–1.3	0.25			Rem.
5052	0.25	0.40	0.10	0.10	2.2–2.8	0.10	0.15–0.35	—	Rem.
5456	0.25	0.40	0.10	0.50–1.0	4.7–5.5	0.25	0.05–0.25	0.20	Rem.
6063	0.20–0.60	0.35	0.10	0.10	0.45–0.9	0.10	0.10	0.10	Rem.
7075	0.40	0.50	1.2–2.0	0.30	2.1–2.9	5.1–6.1	0.18–0.28	0.20	Rem.
8011	0.50–0.9	0.6–1.0	0.10	0.20	0.05	0.10	0.05	0.08	Rem.

Table 2 Solubility (wt%) of the Main Elements of Aluminum Alloys

Alloying element	Maximum solid state solubility	Solubility at lower temperature	Atomic radius difference (%)	Lattice structure
Si	1.65 (577°C)	0.05 (250°C)	−6.3	Cubic; A4
Fe	0.052 (655°C)	0.001 (400°C)	−11.2	BCC/FCC
Cu	5.65 (548°C)	0.2 (200°C)	−11.2	FCC
Mn	1.82 (659°C)	0.36 (500°C)	−8.4	Cubic; A12
Mg	14.9 (451°C)	2.95 (150°C)	+11.9	CPH
Zn	82.8 (382°C)	4.4 (100°C)	−4.2	CPH

(a) Eutectic forming: Au; Ba; Be; Ca; Co; Cu; Fe; Ga; Ge; Mn: Ni; Sb; Si; Sn; Zn.
(b) Peritectic forming: Cr; Hf; Mo; Nb; Ta; Ti; V; W; Zr.
(c) Monotectic forming: Bi; Cd; In; Na; Pb; Tl.

The as-cast microstructure of aluminum alloys generally presents dendrites with central enrichment in aluminum and peritectic-forming elements ($k>1$), while the eutectic formers ($k < 1$) concentrate on the interdendritic regions. The partition coefficient (k) is given by the ratio between the concentration of solute in the solid (C_S) in relation to the concentration of solute in the liquid (C_L). The peritectic-forming elements act in addition as grain refiners. The quantity and size of the intermetallic compounds formed in the interdendritic regions depend mainly on the solute quantity and on the solidification conditions. In general, their size varies from 1 μm up to dozens of micrometers.

After solidification, some alloying elements, such as Cr, Mn, V, Zr, and Fe, are in solid solution in concentration higher than equilibrium (given by their phase diagram). These elements precipitate in the form of intermetallic compounds during the alloy thermomechanical processing. The particles formed are fine, generally smaller than 0.1 μm, and close to each other. They are called dispersoids in the technical literature and have an accentuated effect on the control of the recrystallization kinetics, final grain size, and resulting texture.

Schematically, the main microstructural constituents that are present in aluminum alloys can be classified [2] into five types:

(1) Intermetallic compounds formed during solidification and that remain practically insoluble during the thermomechanical processing, such as Al_3Fe; $Al_{12}(Fe,Mn)_3Si$; $Al_6(Fe,Mn)$; Al_9Fe_2Si; Al_7Cu_2Fe; and Si.
(2) Intermetallic compounds formed during solidification and that can be dissolved during thermomechanical processing, such as Mg_2Si; Al_2Cu; $Al_4CuMg_5Si_4$; Al_2CuMg; $Mg(Zn,Al,Cu)_2$; and $Mg_{32}(Al,Cu,Zn)_{49}$.
(3) Intermetallic particles formed during the first thermal or thermomechanical treatment starting from the supersaturated solution produced during solidification. After being formed, these particles (dispersoids) are practically insoluble. Examples: $Al_{12}Mn_3Si$; $Al_{20}Cu_2Mn_3$; Al_3Zr; and $Al_{12}Mg_{12}Cr$.

(4) Precipitates formed during precipitation hardening heat treatment, at relatively low temperatures. They can easily be dissolved and reprecipitate through heat treatments. Examples: $CuAl_2$; Al_2CuMg; $Mg(Zn,Al,Cu)_2$; $Al_4CuMg_5Si_4$; and a variety of Guinier-Preston zones and metastable precipitates.

(5) Undesirable non-metallic inclusions. Examples: Al_2O_3; MgO; $MgAl_2O_4$.

In addition to the elements and compounds mentioned above, mention should be made of hydrogen as being the only gas that presents measurable solubility in aluminum. It causes porosity, surface defects (blisters), reduction in fatigue resistance, and loss in toughness.

2 WORK HARDENING

The addition of alloying elements or the presence of impurities can lead to three types of alloys, from the point of view of work hardening and recrystallization: (i) the impurities or alloying elements are in solid solution; (ii) the alloying elements or impurities are in the form of second-phase precipitates; and (iii) the alloying elements are in the form of second phase and the volume fraction of this second phase is high. In the alloys of type (ii), the recrystallization is studied in the matrix phase, generally more ductile, considering the precipitate undeformable. Typical examples of this type are the alloys that can be hardened by coherent precipitation in the overaged condition (incoherent precipitates), the commercial aluminum (Al–Fe–Si) and the 3003 (Al–Mn–Fe–Si) alloy. In the alloys of type (iii), it is considered that the two phases deform generally in a different way and recrystallize in accordance to different kinetics for each phase. Microstructures of the type (iii) are called duplex and practically do not occur in commercial aluminum alloys. Typical examples of type (iii) alloys are the α–β brasses and the ferritic–austenitic stainless steels. In the following, we discuss the work hardening of pure aluminum and aluminum alloys of the (i) and (ii) types.

Pure aluminum has a stacking fault energy (SFE) higher than $150 \, mJ/m^2$, i.e., fairly high when compared [3] to other metals and alloys with FCC lattice, such as Ag, Cu, Au, Pb, Pt, brasses, bronzes, and austenitic stainless steels. After coldworking, a small part (between 2% and 10%) of the energy employed in the deformation is stored in the metal [4], in the form of lattice defects. The density and distribution of these defects, mainly the dislocations, may be distributed non-homogeneously, forming a cellular substructure [5,6], which is the case of aluminum alloys. The stored energy and the dislocation density (and its distribution) depend on various factors such as the metal in itself, its purity, grain size, and deformation: amount, mode, temperature, and rate [7]. Apart from the cellular substructure, aluminum and its alloys can present, after deformation, other and more accentuated heterogeneities in the distribution of lattice defects, here generically called "deformation bands" [8,9]. These heterogeneities present a high concentration of lattice defects and considerable variations in the orientation difference (dozens of degrees) between regions, which are about some micrometers apart, within the same grain. They are, in general, preferred locations for recrystallization start (nucleation). Adjacent regions to the grain boundaries also present a differentiated behavior during plastic deformation [10,11].

For a tensile test, the flow curve of annealed and recrystallized polycrystalline aluminum alloys, in the region of uniform plastic deformation, can be expressed by the Ludwik equation [12]

$$\sigma = \sigma_0 + K\varepsilon^n \qquad (1)$$

where σ_0 is the yield stress, K is the strength coefficient, ε is the true strain, and n is the strain-hardening exponent. Typical n values for aluminum alloys are situated between 0.24 and 0.18. As n decreases, there is an increase in K.

As mentioned previously, in Sec. 1, the solid solution hardening (in comparison to precipitation hardening) mechanism has minor significance in the case of aluminum alloys. Two causes, closely related, contribute to this behavior. The main one is that aluminum cannot dissolve, in the solid state, other atoms or solutes in significant quantities. Zinc and magnesium are alloying elements that present higher solubility in commercial aluminum alloys. The second cause is that the pure aluminum SFE is very high and does not lower significantly due to the presence of solutes in solid solution, which is the case, for example, of copper alloys. Despite this fact, the solute atoms in solid solution interact with the dislocations and decrease their mobility, requesting higher densities of mobile dislocation in order to allow plastic deformation to occur, resulting in a higher stored energy (driving force for recrystallization). Figure 1 presents [2] strain-hardening curves for alloys containing different levels of magnesium, as compared to pure aluminum.

Plastic deformation theories of alloys containing non-deformable particles have taken into account two fundamental aspects: (i) variation of the yield strength (YS) with the quantity and size of precipitates; and (ii) influence of these parameters on work hardening. The influence of the precipitates on the YS has been discussed in a classic work by Orowan in 1948, the main conclusions accepted
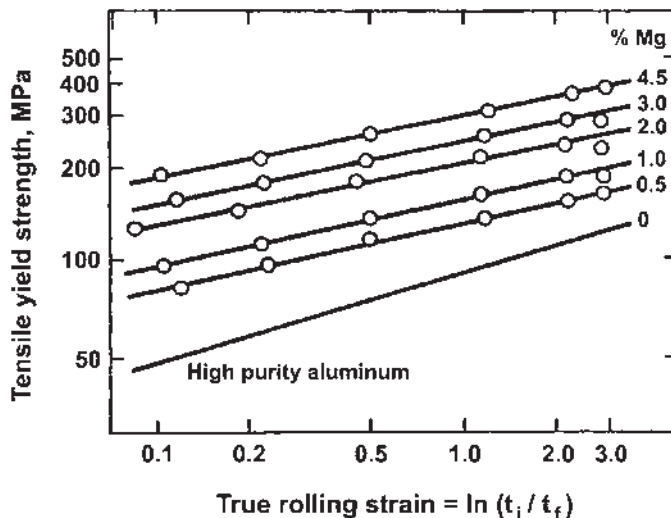
Figure 1   Strain-hardening response of high purity Al and Al–Mg binary alloys during cold rolling. (From Ref. 12.)

until nowadays. The main conclusion is that the YS is inversely proportional to the interparticle spacing. In the following, we discuss work hardening by itself, which is more closely linked to the stored energy by deformation, hence, to recrystallization.

One of the first attempts to explain work hardening due to dispersed obstacles has been suggested by Fisher et al. [13]. They considered that during deformation, successive dislocation loops would be formed around the precipitates, on the slip plane; the "back stress" corresponding to these rings requesting increasing stresses from the dislocation source. The expression that has been developed to describe the hardening is given by

$$\tau_{h} = \{[3NG(F_{v})^{3/2}] \div r\} + \tau_{h}' \tag{2}$$

where τ_{h} is the deformation shear stress, N is the number of dislocation loops, F_{v} is the precipitate volume fraction, r is the precipitate radius, τ_{h}' is the strain hardening due to other causes, and G is the shear modulus.

Dew-Hughes and Robertson [14], using Eq. (2) for 7% shear deformation in overaged Al–Cu single crystals, have shown that it was not adequate and, according to their experimental results, suggested to use simply the factor F_{v} instead of $(F_{v})^{3/2}$.

Ashby [15] has shown that the Fisher et al. [13] theory was reasonable only for deformations smaller than 2% in shear and discussed [16], in a general manner, the formation of dislocations in the cases in which the matrix and precipitate do not deform equally. It has been observed that if a dispersed phase, in the form of small particles, deforms less than the matrix (or does not deform), there are deformation gradients within the material. In the non-homogeneous microstructure, in order to accommodate the deformation gradients that occurred during plastic deformation, "geometrically necessary" dislocations will be stored in the matrix, next to the precipitates. Only for small deformations, conditions in which the density of geometrically necessary dislocations is smaller [16], the theory developed by Fisher et al. [13] would be applicable.

Two different types of geometrically necessary dislocations are schematically presented in Fig. 2. Figure 2a shows a volume element of the non-deformed crystal containing the non-deformable second-phase inclusion. In Fig. 2b, the inclusion is hypothetically removed and the crystal is sheared uniformly; as the inclusion is non-deformable, the hole has to be restored in order to accommodate the inclusion. Figures 2c and d show the restoration to the original form through shear rings, respectively, in the primary and secondary systems. It should be noted that this process leads to the curvature of the lattice in the matrix in this region. Local lattice curvature is a factor that favors the formation of recrystallized regions [17]. In Figs. 2e and f, the restoration is made through the formation of prismatic dislocation loops, that do not lead to a local lattice curvature, but are susceptible to rearrangements of the recovery type.

During deformation process, the particle form and size is important in the determination of the types of dislocation arrangements formed around the precipitates. In dispersions with equiaxed particles smaller than 3000 Å, the arrangements of the 2e and 2f types are preferred and the types 2c and 2d are
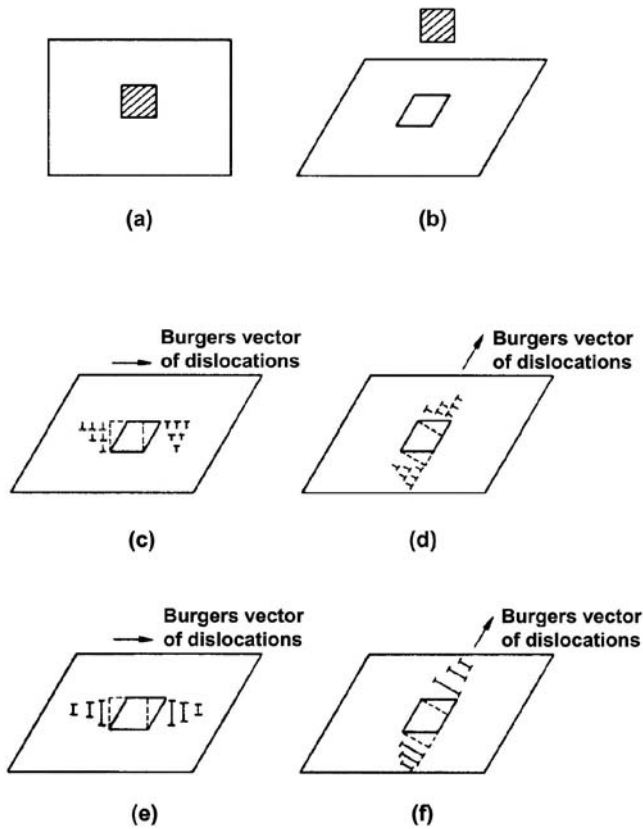
Figure 2 (a) Shows an element of crystal containing a second-phase hard inclusion (shown prior to deformation). (b) The inclusion has been removed and the crystal sheared uniformly. (c) and (d) show the hole restored to its original shape by the emission of glide dislocation on (c) the primary system and (d) a secondary system. (e) and (f) show the hole restored to its original shape by prismatic glide. (From Refs. 16 and 17.)

preferential with non-equiaxial particles larger than 5000 Å, for example, in plates [17].

Related to the dislocation density, Russel and Ashby [18], working with aged Al-Cu single crystals, showed that the dislocation density of the alloys containing dispersed precipitates increased with the decrease in the precipitate spacing. The effect accentuated as the deformation increased, due to the geometrically necessary dislocations.

Related to the deformation substructure, it has been shown in aluminum [19] and in other systems [20–22] that the finer the precipitate distribution, the larger would be the required strain for the cell formation, and smaller the cell size and its misorientation.

By using the equation proposed by Dew-Hughes and Robertson [14], for a given strain, the stored energy from cold working and, therefore, the driving force (ΔP) for recrystallization, can be given by the expression

$$\Delta P = C' + C''(F_v \div r) \tag{3}$$

where C' and C'' are the constants, F_v is the precipitate volume fraction, and r is their radius. According to this expression, for a given strain, the finer the precipitate dispersion, the larger is the stored deformation energy and, therefore, larger is the driving force for recrystallization.

Figure 3 schematically presents in an idealized way the dislocation substructure in an aluminum alloy containing primary particles (larger ones) of intermetallic phases and dispersoids (smaller ones). It is interesting to observe the large strain concentration around the larger particles and the role of the smaller particles (dispersoids) in blocking the cell walls that will transform into subgrain boundaries or sub-boundaries during subsequent annealing.

3 SOME DEFINITIONS

In a cold-worked metal, the microstructural changes that occur after annealing happen in such a manner to diminish the stored energy due to straining. This energy decrease is provided through mechanisms of reduction and rearrangement of lattice defects. The major microstructural changes are the following [24]:

(1) point defect reactions leading to the decrease of their number;
(2) annihilation of dislocations of opposite sign and shrinking of dislocation loops;
(3) dislocation rearrangement to attain smaller energy configurations (for example, small angle boundaries);
(4) formation of high angle boundaries;
(5) absorption of point defects and dislocations by migration of high angle boundaries; and
(6) reduction of total grain boundary area.

These processes can superpose each other locally and chronologically during annealing, but as the lattice defect distribution is generally heterogeneous, each
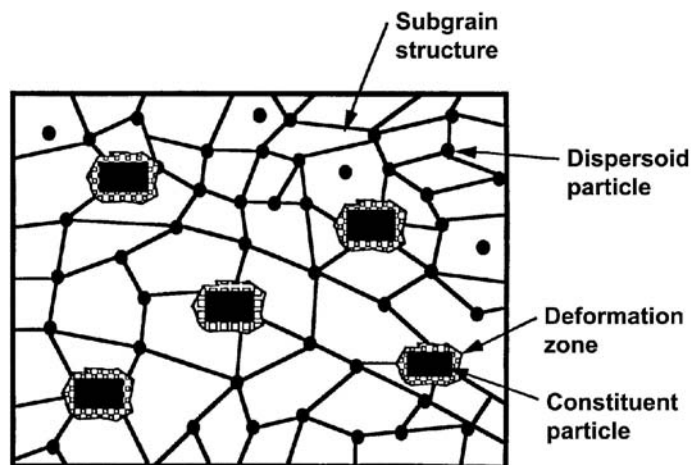
Figure 3   Idealized cold-worked substructure in an aluminum alloy containing both coarse and fine (dispersoids) particles. (From Refs. 2 and 23.)

region within the material can go, in different times, through the sequence of microstructural changes mentioned above. The processes (1), (2), (3), and (4) are defined as recovery and the processes (5) and (6) are defined as recrystallization and grain growth, respectively.

Some cold-worked alloys can present considerable softening, in some cases, more than 50%, without formation and/or migration of high angle boundaries. This accentuated recovery is called *continuous recrystallization* by Haessner [24].

4 RECOVERY

Before and even during the recrystallization of a cold-worked metal, the driving force for the migration of the high angle boundaries is diminishing continuously due to recovery. In general, the recovery and recrystallization processes overlap chronologically in the same sample [25]. In other words, as the lattice defect distribution in the same sample is heterogeneous, a given microregion (more deformed) goes through the recrystallization process, some other neighboring microregion (less deformed), goes through the recovery process. This means that the regions that are not swept by the migration of the high angle boundaries show a decrease in the dislocation density due to recovery. The decrease in the stored energy in these non-recrystallized regions can also occur through dislocation rearrangements.

There are some factors that influence the competition between recovery and recrystallization. They are SFE, strain, annealing temperature, heating-up speed, deformation temperature, and applied stress. In the following, each factor will be briefly analyzed.

Dislocation mobility in a metal is directly proportional to its SFE. Therefore, in metals having a high SFE, recovery occurs more easily, diminishing the quantity of lattice defects (the main driving force for recrystallization) in the non-recrystallized regions. On the other hand, in metals having a low SFE, recovery is sluggish, not competing in an accentuated manner with recrystallization. Figure 4 illustrates the relation between the fractions of softened and recrystallized material for two materials with different SFEs. The softened fraction (A) is defined [26] by Eq. (4):

$$A = [(H_e - H_x) \div (H_e - H_r)] \tag{4}$$

where H_e is the hardness of the work hardened material, H_r is the hardness of the recrystallized material, and H_x is the sample hardness. One can clearly observe in Fig. 4 that copper softens nearly exclusively through recrystallization, while aluminum presents considerable softening through recovery.

Higher strains increase the quantity of recrystallization nuclei as well as the driving force for recrystallization. On the other hand, smaller deformations hinder recrystallization, leaving room for recovery to occur. Furthermore, these factors can be analyzed in the following manner: the number of nuclei formed in highly strained materials migrate to smaller distances and with higher speed in order to complete recrystallization. The opposite occurs for materials with little strain — while the reaction fronts have to migrate to larger distances with smaller speed, the occurrence of recovery decreases the driving force for recrystallization [27,28].
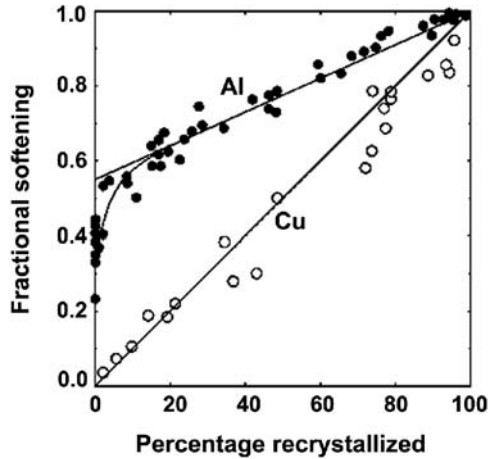
Figure 4 Fractional softening versus recrystallization volumes in copper (lower SFE) and aluminum (higher SFE). (From Ref. 26.)

The lower the annealing temperature, the higher will be the participation of recovery in the global softening process and the explanation to this is that the recovery mechanisms have, in general, smaller activation energies than those associated with the recrystallization mechanisms. If both processes are thermally activated and compete between themselves, lower temperatures favor the lower activation energy, i.e., recovery.

Recovery occurs through various mechanisms, in a large temperature range, starting at 0.2 T_f (T_f is the absolute metal melting temperature). Recrystallization, on the other hand, occurs generally in the range of 0.3 to 0.6 T_f. Therefore, when a metal is slowly heated up, the residence time at lower temperatures is greater, where exclusively recovery prevails. Consequently, the driving force for recrystallization will diminish due to the decrease in the quantity of lattice defects and to their rearrangement, hence delaying recrystallization [29].

As the deformation temperature increases, recovery occurs with greater intensity, diminishing the driving force for recrystallization. The recovery can occur either in the *static* manner, due exclusively to the heating-up of the metal, or it can occur in the *dynamic* manner, under stress, during deformation at higher temperatures. Thorton and Cahn [30] studied the effect of small applied loads on the recovery and recrystallization of pure aluminum (99.99%) and copper with various purity levels, initially work hardened. Recovery in aluminum was accentuated and much faster, the effect in copper was much smaller. The larger and faster recovery in aluminum, due to the applied stress, caused also a delay in the recrystallization kinetics. The movement (and rearrangement) of dislocations is thermally activated, but can be helped by an applied stress field. The differences in behavior between aluminum and copper can be attributed to the differences in their SFE and the different dislocation arrangements in both the cold-worked materials.

The studies related to the annealing of alloys with a dispersion of precipitates have given little attention to the recovery process, which can occur competitively with recrystallization. In these alloys, apart from the normal recovery processes

common in pure metals and in solid solutions, recovery processes can occur involving geometrically necessary dislocation arrangements next to the precipitates.

Work on plastic deformation [31] with internally oxidized alloys has shown that work hardening presents time-dependent relaxation due to copper oxide particles, even at room temperature. Ashby [16] proposes that, probably, this effect is typical of dispersion-hardened alloys. This relaxation presents itself in two ways: (i) through a significantly different strain-hardening rate when identical crystals are tested at different temperatures; or (ii) when the test is interrupted, there is a lowering of the required stress to continue the deformation, once load is reapplied.

Hirsch and Humphreys [32], using transmission electron microscopy, showed that the decrease in the strain-hardening rate is associated with a large reduction in the density of dislocation loops. The observations indicate that the geometrically necessary loops are removed, by a diffusion processes, easier than the statistically stored dislocations. The arrangement of the geometrically necessary dislocation loops, either shear or prismatic, are different to the dislocation arrangements found in work hardened solid solutions without particles. As can be observed from Fig. 5, both sides of the shear loops, or from both prismatic loops with opposing signs, are separated by a small and well-defined distance, which is the particle diameter. The atom diffusion from one side to the other of the particle causes dislocation climb and annihilation shear loops by slip (Fig. 5a) or concomitant contraction of the two prismatic dislocation loops, situated on opposing sides of the particle (Fig. 5b).

Ashby [16] reminds us that this process can occur at low temperatures because the driving force for the diffusion is high, the diffusion distance is small and that there exist high diffusivity paths through the matrix/precipitate interface.

The particle size and shape have an important role in these recovery processes. As to the size, Ebeling and Ashby [33] have shown that at room temperature, the work hardening of copper single crystals with a silica dispersion of about $1000\,\text{Å}$ in diameter was stable. On the other side, Jones [31] showed that in copper single crystals containing beryllium oxide particles with about $100\,\text{Å}$, work hardening was unstable at room temperature. As to the form, arrangements of geometrically necessary dislocations around particles, whose geometry imposes a longer diffusion path, are generally more resistant to recovery.

5   RECRYSTALLIZATION

The major part of the energy accumulated due to straining is, in the majority of the cases, eliminated through the recrystallization process. As has been shown in Sec. 3, recrystallization is defined as being associated with the removal of defects through the migration of high angle grain boundaries. These boundaries, once they migrate, causing recrystallization, can be pre-existing boundaries before the deformation, or even new ones formed during annealing [7]. Both generation and migration of high angle boundaries are thermally activated processes. The generation of new high angle boundaries can occur by migration of small angle boundaries, also known as sub-boundaries, or by the coalescence or rotation of subgrains [7]. In the following, the three mechanisms associated with recrystallization nucleation mentioned above will be discussed, without taking into account the presence of solutes or precipitates. The effect of solute atoms in solid solution and, mainly, the effect of second-phase particles, will be discussed later.
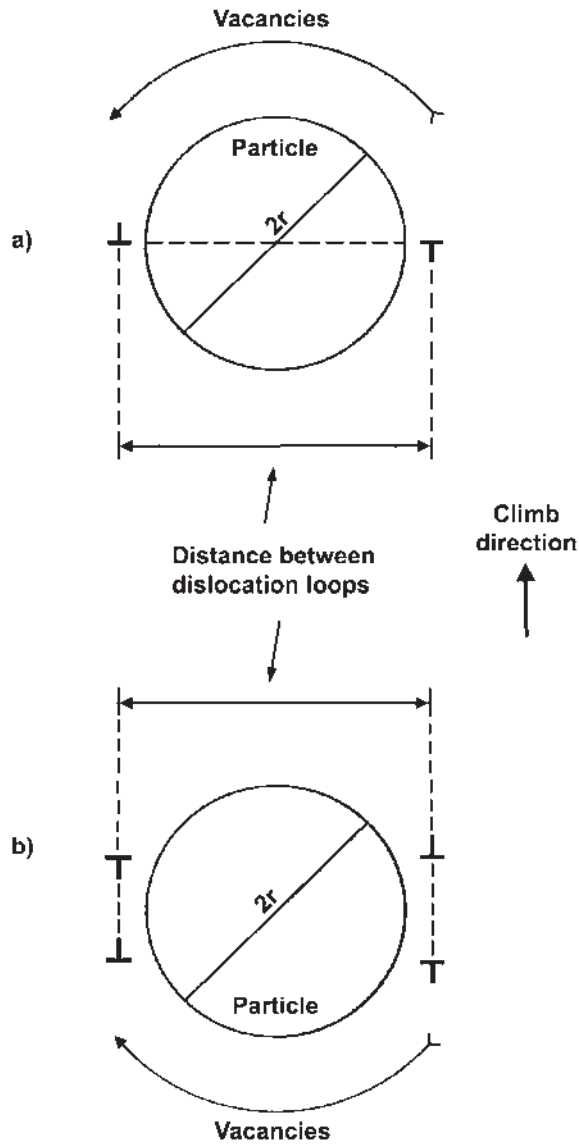
Figure 5   Flow of vacancies from right to left causes dislocation climb and annihilation of shear loops by slip (a) or concomitant contraction of two prismatic dislocation loops (b), situated on opposite sides of the particle. (From Ref. 16.)

The model for the recrystallization nucleation, through migration of small angle grain boundaries, has been proposed independently by Cahn [34] and Beck [35], later enhanced by Cottrell [36] and is generally quoted in the literature as the Cahn-Cottrell model. The model is based on the polygonization phenomenon where regions of small dislocation densities are formed, surrounded by sub-boundaries. Once a subgrain is formed, it is capable to grow over its neighbors by migration of its thermally assisted sub-boundaries. In this

manner, there will be a decrease in the stored energy by the removal and rearrangement of crystalline defects. The migrating sub-boundary absorbs dislocations, increasing its difference in orientation between subgrains, its energy and mobility until it transforms itself into high angle boundaries, characterizing the nucleation of recrystallization. Sandström et al. [37] observed subgrain growth in pure aluminum at temperatures higher than 0.6 T_f. Varma and Willits [38] made the same observation. Walter and Koch [39] observed with electron transmission microscopy the occurrence of the nucleation mechanism by sub-boundary migration, associated with transition bands in an Fe–3% Si alloy. There has been recent experimental evidence of low angle boundary migration during hot working in pure aluminum and for Al–5% Mg alloy [40].

The recrystallization nucleation model via coalescence or rotation of subgrains has been postulated in 1948 by Smith [41] and later studied by Hu [42,43] in Fe–3% Si single crystals. Li [44] analyzed the process from the kinetic and thermodynamic point of view and concluded that it is a slow process. Later, Doherty and Szpunar [45] showed that the subgrain coalescence at temperatures lower than 0.65 T_f, could lead to a kinetically viable process. Jones et al. [46–48] suggested also, from observations made with transmission electron microscopy, the occurrence of subgrain coalescence in aluminum. The mechanism is linked to the coalescence of two neighboring subgrains, which is equivalent to a rotation, making their crystal lattices coincide. This rotation makes that the lattices have no deviation, removing the sub-boundary due to the transference of dislocations from the sub-boundary that is vanishing to the neighboring sub-boundaries. Therefore, coalescence is a mechanism that promotes subgrain growth, sub-boundary removal (diminishing the stored energy) and changing the orientation differences between the group that went through coalescence and its neighboring subgrains. The increase of these orientation differences leads to the appearance of a high angle grain boundary capable of migrating with high speed, constituting the recrystallization nucleus.

The start of the recrystallization through the migration of high angle boundaries, pre-existing prior to deformation, has been originally suggested by Beck and Sperry [49] based on observations with optical microscopy in aluminum. Later, Bailey [50,52] and Bailey and Hirsch [51] studied it in Cu, Ni, Au, Ag, and Al, using transmission electron microscopy. In these studies, the strain was smaller than 40%.

This mechanism (strain-induced grain boundary migration; SIGBM) involves the migration of a pre-existing grain boundary to the inner part of the more strained grain. The condition for this process to occur is the favorable energy balance between the reduction in stored energy due to the removal of defects and the increase in total surface of the grain boundary due to bulging of this boundary over the more strained grain. The migration of existing boundaries prior to deformation is induced by the difference in the crystal defect density (mainly dislocations) between neighboring grains. Migration occurs from the less deformed grain to the more deformed one. This leads to a reduction in internal energy, which is equivalent to the difference between the energy liberated due to the lowering of the defect density in the region swept by the grain boundary and the energy associated with the increase in boundary surface. The condition for growth for a segment with a length of $2L$ in grain boundary is

$$L > (2\gamma \div \Delta E) \tag{5}$$

where γ is the grain boundary surface energy per unit area and ΔE is the liberated energy due to lowering of the number of crystal defects.

The mechanism of recrystallization start by migration of existing boundaries before straining is favored by small deformations, because under these conditions, the differences in lattice defect densities between neighboring grains are generally more accentuated.

Before discussing the role of the impurities and alloying elements on aluminum recrystallization, it should be mentioned that the recrystallization nucleation in deformation heterogeneities, such as deformation bands, is well documented in the literature [53–55].

Solid solution atoms exert conflicting effects on the recrystallization. As discussed in Sec. 2 (work hardening), they diminish the dislocation mobility and increase the dislocation density and the driving force for recrystallization, in comparison with the pure metal. On the other hand, solid solution atoms hinder the dislocation arrangement and cause an anchoring force both on the low angle boundaries (recrystallization nucleation) and on the high angle boundaries (recrystallization growth), making difficult their movement. Considering the global effect of the solute in solution, on work hardening, at the beginning (nucleation) and growth of the recrystallized regions, we may state that, in general, the higher the solute content in solid solution, greater will be the recrystallization delay. It must be pointed out, nevertheless, that the effect of small additions of solute to metals of high purity (dilute solutions) is more accentuated than the effect of additions in solutions which already present high solute content, as illustrated in Table 3.

Analyzing the literature on aluminum alloys containing precipitate dispersions published in the last 50 years, there are both evidences related to cases where the presence of dispersed precipitates delayed recrystallization [57–60], as well as cases where recrystallization has been accelerated [61,62], in relation to the solid solution without precipitates.

Doherty and Martin [63–65] working with Al–Cu alloys with different levels of Cu and various precipitate dispersions, both in single crystals and in polycrystals, showed that: (a) in a matrix containing a coarse precipitate, recrystallization

Table 3 Influence of Impurities on the Recrystallization Temperature of Some Metals (From Ref. 56.)

Metal	Recrystallization Temperature (°C)	
	Commercial purity	Zone refined
Aluminum	200	−50
Copper	180	80
Iron	480	300
Nickel	600	300
Zircon	450	170

occurred faster than in the case of a solid solution without precipitates; (b) in fine precipitate dispersions, recrystallization was increasingly delayed until it became even slower than the recrystallization of the solid solution; and (c) a fourfold decrease in the mean free path between precipitates raised 10^5 times the 50% recrystallization time and increased the final grain size by a factor of 10^3.

These results can be explained considering that the dispersion influences three parameters: (i) the work hardened state (see Sec. 2); (ii) the recrystallization nucleation; and (iii) the speed of boundary migration.

There is substantial experimental evidence [63–68] of the recrystallization nucleation next to precipitates or inclusions. Three conditions have to be fulfilled by a dispersion in order to favor recrystallization. (i) Size and shape of favorable precipitates. Dislocation arrangements in the work hardened matrix, that favor nucleation, occur preferably next to non-equiaxial particles larger than 5000 Å [17]. (ii) Large interparticle distance. There is a minimum interparticle distance below which the sub-boundaries are blocked before they migrate sufficiently in order to transform themselves into high energy boundaries with high mobility. (iii) Strain − there is a minimum critical strain [68,69], which depends on particle size (see Fig. 6), and above which the particles induce recrystallization nucleation.

Doherty and Martin [63–65], explaining their results only considered condition (ii), i.e., the interparticle spacing. However, on analyzing the dispersions in which the recrystallization has been accelerated, in these works, we may observe that the three conditions have been satisfied.

In dispersion containing alloys, there are two factors, acting in opposition, that influence the growth speed of the recrystallized regions. On one side, there is the stored energy by deformation, driving force for recrystallization, to which the growth speed is directly proportional. On the other side, we have the resistance imposed by the precipitates to the grain boundary migration, making it sluggish. The
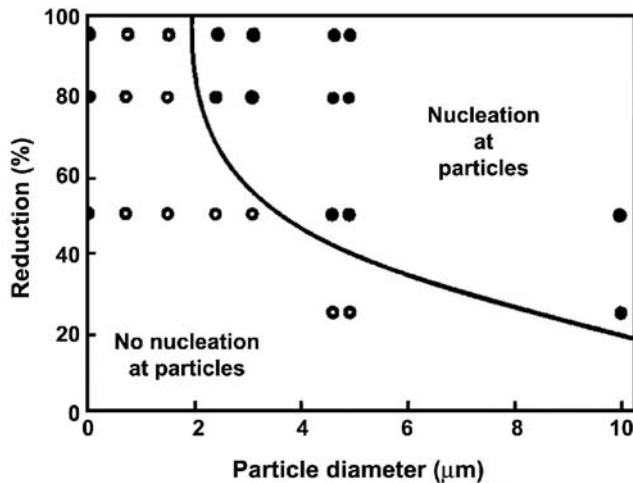
Figure 6 Effect of rolling reduction and particle size on the occurrence of particle-stimulated nucleation (PSN) on recrystallization. (From Ref. 68.)

growth speed depends on the balance between these two factors; in extreme cases, the resistance imposed by precipitates may stop completely the boundary migration. The influence of precipitate dispersion on the driving force has been discussed in the topic related to work hardening. In the following, we discuss the role of the precipitate dispersion on the resistance to boundary migration.

According to Zener, as quoted by Smith [70], the maximum total resisting force per unit boundary area that spherical precipitates exert on a boundary, retarding its movement, is directly proportional to the precipitate volume fraction (F_v) and inversely proportional to the particle radius (r):

$$F_{\max} = \frac{3}{4}[\gamma(F_v \div r)] \tag{6}$$

Therefore, the resistance to boundary movement, due to the particles, is proportional to the factor ($F_v \div r$).

As discussed in Sec. 2, the finer the precipitate distribution, the larger will be the stored energy by deformation, the larger will be the driving force for recrystallization and, hence, easier will be the growth, once the growth speed is directly proportional to the driving force. On the other hand, the resistance to boundary mobility due to precipitates is also larger for finer precipitate dispersions. Starting from these opposing effects, the variation in the precipitate dispersion exerts on the growth speed, Doherty and Martin [63–65] suggested that the growth stage of the recrystallization would not be greatly affected by the precipitate dispersion variation, leaving to nucleation the major role in the kinetic of recrystallization of alloys containing precipitate dispersions.

Figure 7 summarizes the effect of strain, interparticle spacing and particle size on the recrystallization kinetics [71,72].

From what has been discussed up to now, it may be concluded that in aluminum alloys, the larger intermetallic compounds formed during solidification can stimulate recrystallization nucleation, accelerating it and cause grain refinement, while dispersoids cause recrystallization delay. Figure 8 presents
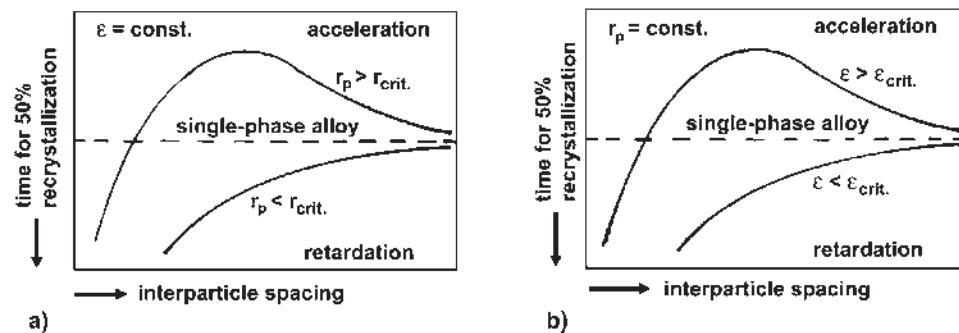
Figure 7 Schematic diagram of the influence of dispersed particles on recrystallization [72]: (a) Recrystallization acceleration occurs only if the particle size is large enough to give rise to local lattice curvature ($r_p > r_{crit}$). (b) For a given particle size, a critical strain is necessary for particle-stimulated nucleation ($\varepsilon > \varepsilon_{crit}$).
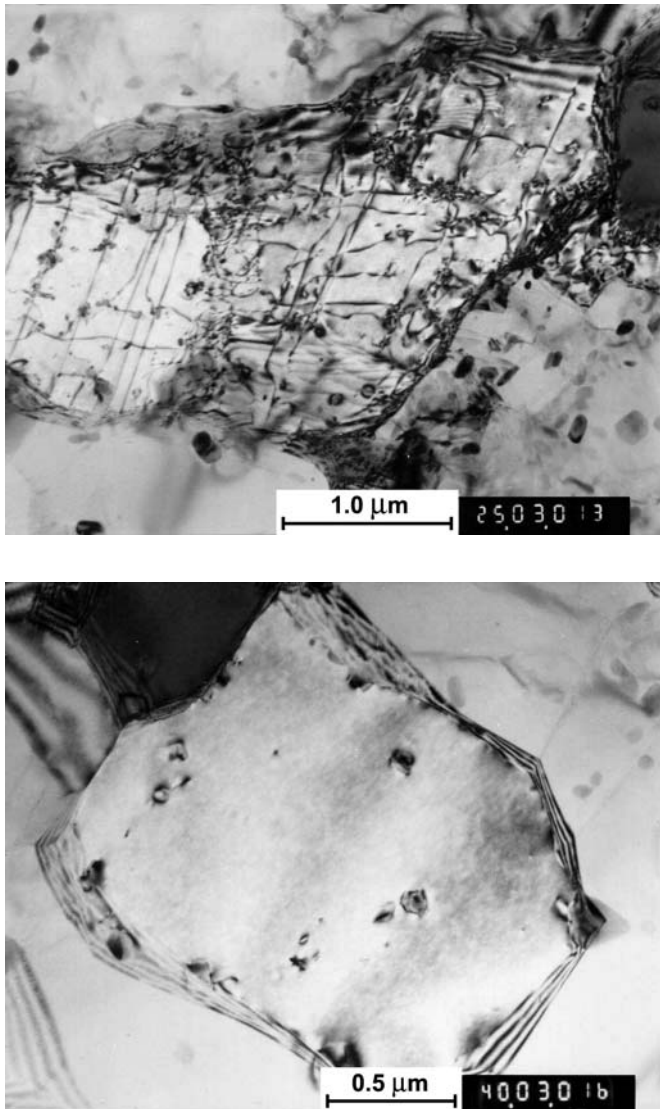
Figure 8 Transmission electron micrographs of specimens (alloy 3003) after solution annealing at 630°C for 2 h, cold rolled to $\varepsilon = 0.7$ and annealed at 370°C for 11 h. $\varepsilon = \ln(t_i/t_f)$, where t_i and t_f are the initial and final thickness, respectively. (From Ref. 28.)

transmission electron micrographs of specimens (alloy 3003) after solution annealing at 630°C for 2 h, cold rolled to $\varepsilon = 0.7$ (50% reduction) and annealed at 370°C for 11 h. The dislocation rearrangement, the subgrain migration and thereby the nucleation of new recrystallized grains were inhibited by the reprecipitation of fine particles.

6 GRAIN GROWTH

After recrystallization is completed, the microstructure of recrystallized grains still is not the most stable. The grain boundary energy now acts as a driving force for the grain growth in order to decrease the number of grains per unit volume or, in other words, diminish the total area of these boundaries. This may occur in two ways: (i) continuously, i.e., with the continuous growth of the average grain size; and (ii) discontinuously, with the accelerated growth of only some grains. The first one is called grain growth and the second one is called secondary recrystallization.

Secondary recrystallization is also known by the following denominations: abnormal grain growth, discontinuous grain growth, selective grain growth, germination, selective grain growth, fast grain growth, preferential grain growth, grain coarsening, and anomalous grain growth.

As in primary recrystallization, grain growth is thermally activated. However, the driving force for grain growth is about two orders of magnitude smaller than the driving force for primary recrystallization [73].

The average energy per unit volume due to grain boundaries in a polycrystal, can be obtained by the product $S_v \gamma$ (where S_v is the grain boundary area per unit volume and γ is the average energy of the grain boundaries in a given material per unit area). Therefore, the driving force (as a function of the average grain diameter (D) for grain growth (F_{cg}) can be given by

$$F_{cg} = 2(\gamma \div D) \tag{7}$$

Either during recrystallization or even during grain growth, there is a migration of large angle boundaries. There are, however, two basic differences between both the cases. First, the boundary migration speed is much higher during recrystallization than during grain growth. Further, during recrystallization, grain boundaries migrate in the opposite sense to its center of curvature, while in grain growth, migration occurs in the sense of the center of curvature of the boundary.

When the grains grow uniformly, i.e., the majority have at all instants a "diameter" close to a mean-diameter, we define the phenomenon as a normal grain growth or simply grain growth. Another possibility is the preferential growth of some grains, leading to a bimodal statistical distribution of the grain size. This process is known as abnormal grain growth or secondary recrystallization. The kinetics of secondary recrystallization follow a sigmoidal-type equation. The similarity with primary recrystallization kinetics is in its origin; that is why it has been called "secondary recrystallization".

Secondary recrystallization is associated with factors that limit normal grain growth. In other words, for secondary recrystallization to occur, it is necessary that normal grain growth is inhibited or hindered. In its intermediate stages, secondary recrystallization leads to a bimodal grain size distribution, which finalizes at the end of the process.

Among the factors that restrain the normal grain growth, that are impurities in solid solution, presence of particles, presence of strong texture, and limitations due to sample thickness, only the last three ones can promote secondary recrystallization. On the other hand, the retarding effect due to impurities in solid solution (in metals)

is not sufficient to stabilize the matrix grain structure and leads to secondary recrystallization. However, there are other causes that can promote secondary recrystallization in solid solutions and in pure metals.

Secondary recrystallization is frequently associated with the presence of fine particles, as shown by Beck et al. [74], in 1949, in an Al–Mn alloy. Abnormal grain growth occurred when particles of Al_6Mn coalesced and/or dissolved at the imposed annealing temperature. In these regions, the grain boundaries did not suffer any migration restriction. Figure 9 shows the occurrence of secondary recrystallization in an Al–Mn alloy at 625°C and 650°C.
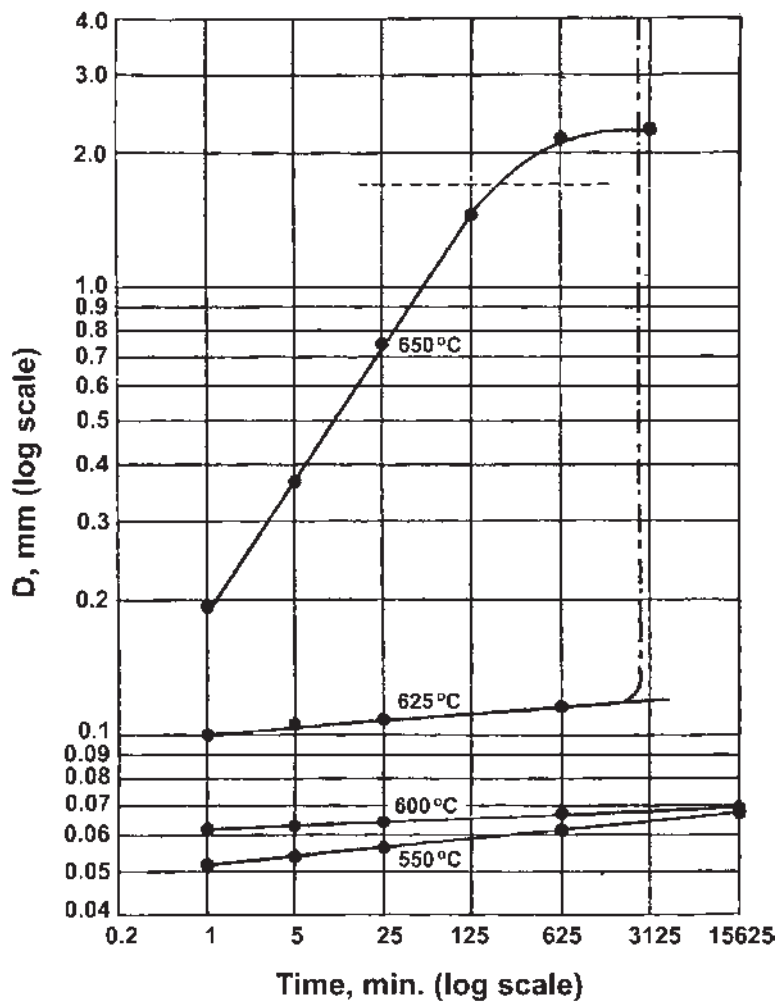
Figure 9 Effect of second-phase particles on grain growth in a manganese–aluminum alloy (1.1% Mn). Grain growth is severely inhibited at temperatures below 650°C because of the presence of second-phase precipitate particles Al_6Mn [74]. Vertical dash-dot line indicates an approximate start of secondary recrystallization (coarsening). Horizontal dotted line corresponds to specimen thickness.

7 TEXTURE

Polycrystalline aggregates exempt from preferential orientation are rarely found among natural materials as well as in industrially processed materials. Rocks and crystalline minerals frequently develop a texture during solidification or during its metamorphosis. Natural fibers and artificial ones present orientation in their molecular chains during growth or manufacture.

Texture can develop in a metal or alloy during one or more processing stages, such as foundry, mechanical working and annealing. Metallic films deposited by sputtering, electro-deposition or chemical vapor deposition (CVD) frequently present also a texture; however, their mechanism of formation is complex and little clarified. In castings, the longer grains in the columnar zone present strong preferential orientation. Yet, the grains in the equiaxed central zone present a more random orientation.

During mechanical working, plastic deformation occurs mainly by slip of certain crystal planes. In this process, the crystal lattice suffers rotation, by slip or twinning, to more favorable orientations, establishing the so-called deformation texture [75]. The final deformation texture of a piece or test sample depends mainly on the change of shape given by mechanical working and by the temperature at which the material has been deformed, and, for small deformations, depends on the initial grain orientation before deformation. Cold rolled deformation textures most frequently found in aluminum and its alloys [76] present predominantly the $\{112\}\langle 11\bar{1}\rangle$ components (known as copper-type texture) and $\{110\}\langle\bar{1}12\rangle$ (known as brass-type texture) and, eventually, the presence of the components $\{011\}\langle100\rangle$ (known as Goss-type texture) and the $\{123\}\langle634\rangle$ (known as S-type texture).

When the deformed metal is annealed, recovery and/or recrystallization may occur, depending on the deformation strain, time and mainly annealing temperature. Generally, annealing at lower temperatures of materials with small deformation causes only recovery and small or nearly no texture modification [75].

Annealing at higher temperatures frequently leads to recrystallization. The occurrence of recrystallization can generate a preferential crystallographic orientation (texture) completely different to the one generated by deformation. In reality, recrystallization can lead to a complete absence of texture, as well as develop extremely intense texture components or, simply, not alter the deformation texture. For example, high purity aluminum, submitted to high reductions by rolling and annealing, can present an extremely intense $\{100\}\langle001\rangle$-type texture (known as cube-type texture), very close to a single crystal texture [77]. This modification in the deformation texture, during recrystallization, is consistent with the idea that the phenomenon of recrystallization is the result of a group of alterations in the microstructure of a material, due to heat treatment, with new grains nucleating and growing. These new grains can have a different orientation to the one given by the microstructure of deformed grains. The presence of precipitates, especially the large ones that stimulate recrystallization nucleation, favor the presence of random texture, or even better, no texture at all. In general, the recrystallization texture of the cube type is strengthened by high levels of cold working and by high recrystallization temperatures. Cube-type texture is in general undesirable, because it causes "earing" during deep drawing. The cube texture results in the formation of ears at $0°$ and $90°$ to the former rolling direction. On the other hand, rolled sheets

display ears at $\pm 45°$ to the rolling direction. The controlled mixing of rolling and cube textures can be utilized to control earing [78,79].

Figure 10 illustrates the appearance of recrystallization texture replacing the deformation texture for the 1050 commercial aluminum and for the 3003 Al-Mn-Fe-Si alloy, as a function of annealing temperature after 70% cold rolling.

It may happen that, at even higher temperatures, during annealing and after recrystallization, both continuous or normal grain growth as well as abnormal or secondary recrystallization may occur. In this case, few grains will grow consuming their smaller neighbors. The resulting texture, also in this case, is generally different from the deformation and recrystallization textures. In the case of aluminum, the resulting texture from secondary recrystallization most frequently found [80] is of the (1 0 0) [0 1 1] type.

8 FINAL REMARKS

Probably more than 90% of all metal production is hot rolled, while cold rolling is performed merely on about 30% of the world metal production. On the other hand, the great majority of the recovery and recrystallization studies available in the literature were accomplished after cold deformation (static recrystallization). The majority of existent theories and models contemplate predominantly static recovery and static recrystallization when compared with dynamic recovery and dynamic recrystallization that occur during hot deformation [normally at $T > 0.5$ T_m (T_m, melting temperature, K)]. It must be pointed out that, although dynamic restoration processes are of great industrial significance, they are not so well understood because they are very difficult to study experimentally and model theoretically [7].

In metals and alloys in which recovery processes are slow, such as those with low or medium SFE (copper, brasses, and austenitic stainless steels), dynamic recrystallization may take place when a critical deformation level (ε_c) is reached [7]. The occurrence of dynamic recrystallization does not result in the elimination of substructure as it does in static recrystallization. Dynamic recrystallization is initiated at a critical strain that is a little less than the strain to peak. As recrystallization spreads, the average flow stress decreases as a result of the flow stress of those restored regions. However, before recrystallization is complete, the first grains to nucleate may have reached a critical strain and new grains nucleate within them. The result of this sequence of events is a uniform distribution of small regions with a variety of strains between zero and the critical value. The concurrent straining establishes a substructure in the new grains, which diminishes the driving force, thus altering the critical strain and nucleus size, and constrains growth so that it ceases before completion by impingement [81–84].

In high SFE (aluminum and bcc metals) present very fast recovery kinetics (static and dynamic). In these metals, dislocation climb and cross-slip occur readily. The recovery occurrence (static or dynamic) decreases the density of crystalline defects and the driving force for recrystallization. Dynamic recovery, therefore, being rapid and extensive at high temperatures is usually considered as being the sole restoration mechanism. At the initial stages, for small deformations, dislocations interact and multiply. As dislocation density increases with deformation, so will the driving force and, hence, the recovery rate, with the development of low angle boundaries and subgrains. For strains higher than a
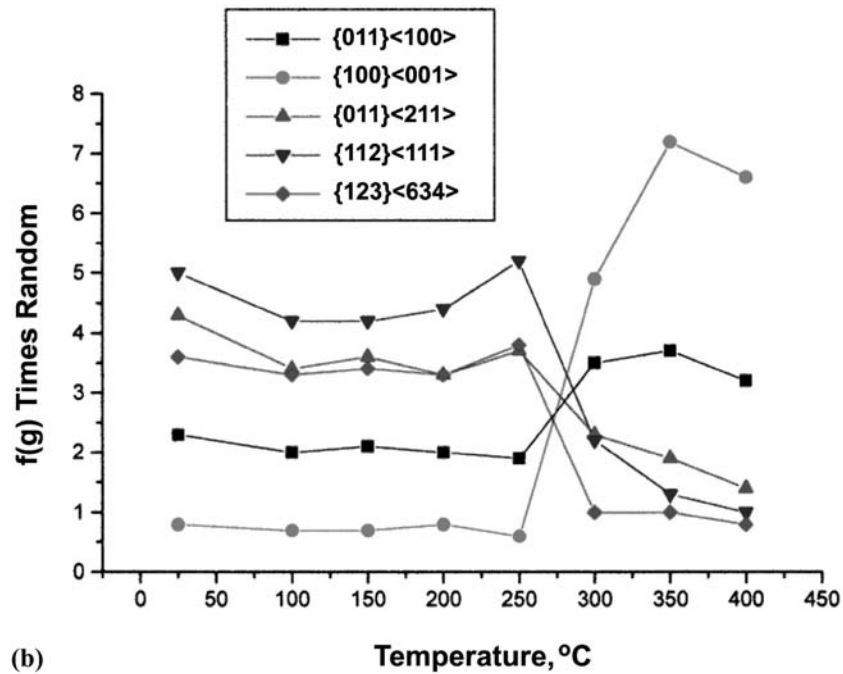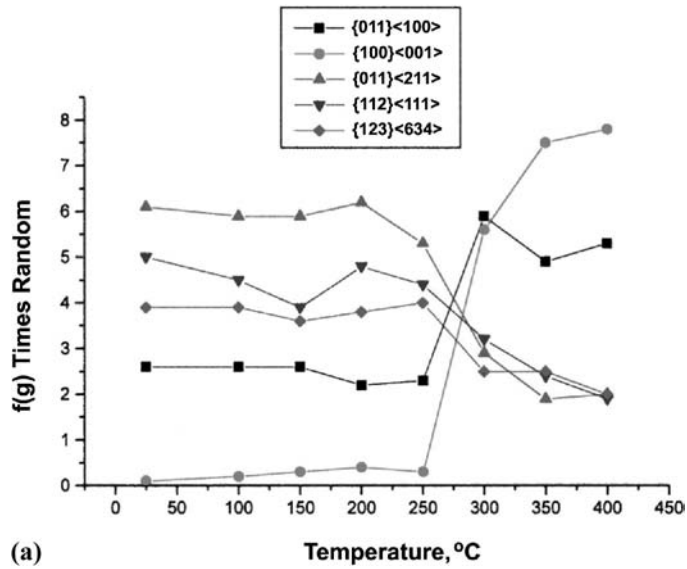
Figure 10 Development of recrystallization texture as a function of 1 h isochronous annealing temperature after 70% cold rolling: (a) for 1050 commercial aluminum and (b) for the 3003 alloy. (Courtesy of Marjorie Löw, IPEN, Brazil.)

critical strain ($\varepsilon > \varepsilon_c$), the rate of work hardening and recovery will reach a dynamic equilibrium in which the subgrain remains equiaxed, with a constant mean size and misorientation (steady state), which is a function of temperature and strain rate. This contrasts with the grain structure, which becomes progressively more elongated with increasing strain [85].

For commercial purity aluminum and aluminum alloys, several well-established constitutive equations have been intensively developed in the last 40 years that relate flow stress with initial microstructure, strain, strain rate, and temperature of deformation [86–90]. Grain boundaries with serrations in several Al alloys that have been hot worked and did not recrystallize have been studied by several authors [91,92].

It is commonly believed that pure aluminum does not undergo dynamic recrystallization. The generally accepted explanation for the non-occurrence of dynamic recrystallization in aluminum is that, due to the high SFE, a stable dislocation arrangement is established by extensive dynamic recovery, so that the value of the stored energy is too low to cause dynamic recrystallization [93].

Some recent work, however, reported that dynamic recrystallization takes place in very high purity aluminum single crystals during compression testing [93–95].

Recent observations in Al-Mg-Mn (AA5456) alloys revealed the presence of small equiaxed grains formed at original grain boundaries and at coarse non-deformed intermetallic particles (PSN). The grains showed extensive networks within them, typical of dynamic recrystallization [90,96]. Their growth is restricted, under dynamic conditions, due to the development of the dislocation substructure within the new grains. However, when the intermetallic particles are small, the only softening mechanism is dynamic recovery [90].

In spite of the fundamental discussion whether dynamic recrystallization is occurring or not during the thermomechanical processing of pure aluminum and in its alloys, two basic aspects seem to be applicable for the dynamic recrystallization of aluminum alloys: (i) the tendency for dynamic recrystallization increases with the increase in the concentration of alloying elements; and (ii) dynamic recrystallization usually nucleates at the original grain boundaries and at coarse intermetallic particles.

REFERENCES

1. R. E. Sanders, Jr., "Technology Innovation in Aluminium Products," JOM (The Journal of the Minerals, Metals and Materials Society) 2001, 53, pp. 21–25.
2. S. F. Baumann and R. E. Sanders, Jr., *Metallurgy of Wrought Aluminium Products*:*Principles and Applications*, Alcoa Aluminum Metallurgy Seminar, Itapissuma, Brazil, December 1996.
3. L. E. Murr, *Interfacial Phenomena in Metals and Alloys*, *Interfacial Free Energy*, Addison-Wesley, Reading, 1975, pp. 87–164 (Chapter 3) (ISBN 0-201-04884-1).
4. M. B. Bever, D. L. Holt, and A. L. Titchener, "The Stored Energy of Cold Work," Progress in Materials Science 1973, 17, pp. 1–190.
5. D. Kuhlmann-Wilsdorf and H. G. S. Wilsdorf, "Origin of Dislocations Tangles and Loops in Deformed Crystals," in: *Electron Microscopy and Strength of Crystals* (G. Thomas and J. Washburn, eds.), Interscience, New York, 1963, pp. 575–604.

6. P. R. Swan, "Dislocations Arrangements in Face Centered Cubic Metals," in: *Electron Microscopy and Strength of Crystals* (G. Thomas and J. Washburn, eds.), Interscience, New York, 1963, pp. 131–181.

7. F. J. Humphreys and M. Hatherly, *Recrystallization and Related Annealing Phenomena*, Pergamon/Elsevier Science Ltd., Oxford, 1996 (ISBN 008 0418848).

8. B. Bay, N. Hansen, and D. Kuhlmann-Wilsdorf, "Microstructural Evolution in Rolled Aluminum," Materials Science and Engineering A 1992, A158, pp. 139–146.

9. O. Engler, "An EBSD Local Texture Study on the Nucleation of Recrystallization at Shear Bands in the Alloy Al–3%Mg," Scripta Materialia 2001, 44, pp. 229–236.

10. Q. Liu and N. Hansen, "Microstructural Study of Deformation in Grain Boundary Region during Plastic Deformation of Polycrystalline Aluminium," Materials Science and Engineering A 1997, A234–236, pp. 672–675.

11. V. Randle, N. Hansen, and D. Juul Jensen, "The Deformation Behaviour of Grain Boundary Regions in Polycrystalline Aluminium," Philosophical Magazine A 1996, 73, pp. 265–282.

12. J. E. Hatch, ed., "Work Hardening, Recovery, Recrystallization, and Grain Growth," in: *Aluminum Properties and Physical Metallurgy*, ASM, Ohio, 1984, pp. 105–133 (Chapter 4).

13. J. C. Fisher, E. W. Hart, and R. H. Pry, "The Hardening of Metal Crystal by Precipitate Particles," Acta Metallurgica 1953, 1, pp. 336–339.

14. D. Dew-Hughes and W. D. Robertson, "Dispersed Particle Hardening of Aluminum–Copper Alloy Single Crystals," Acta Metallurgica 1960, 8, pp. 147–155.

15. M. F. Ashby, "Work Hardening in Dispersion Hardened Crystals," Philosophical Magazine 1966, 14, pp. 1157–1178.

16. M. F. Ashby, "The Deformation of Plastically Non-Homogeneous Alloys," in: *Strengthening Methods in Crystals* (A. Kelly and R. B. Nicholson, eds.), Applied Science Publ., London, 1971, pp. 137–192 (Chapter 3).

17. T. C. Rollason and J. W. Martin, "The Effect of Particle Size upon the Annealing Behavior of Plastically Deformed Two Phase Crystals," Journal of Materials Science 1970, 5, pp. 127–132.

18. K. C. Russel and M. F. Ashby, "Slip in Aluminum Crystals Containing Strong, Plate-Like Particles," Acta Metallurgica 1970, 18, pp. 891–901.

19. R. S. Goodrich, Jr. and G. S. Ansell, "Observation of Deformation Induced Substructure in Al–Al_2O_3 SAP-Type Dispersion Strengthened Alloys," Acta Metallurgica 1964, 12, pp. 1096–1100.

20. J. L. Brimhall and R. A. Huggins, "Electron-Microscopic Observations of Deformed Internally Oxidized Alloys," Transactions of the Metallurgical Society AIME 1965, 233, pp. 1076–1084.

21. F. J. Humphreys and J. W. Martin, "The Effect of Dispersed Phase upon Dislocation Distribution in Plastically Deformed Copper Crystals," Philosophical Magazine 1967, 16, pp. 927–957.

22. M. J. Klein and R. A. Huggins, "The Structure of Cold Worked Silver and Silver-Magnesium Oxide Alloys," Acta Metallurgica 1962, 10, pp. 55–62.

23. E. Nes, "The Effect of a Fine Particle Dispersion on Heterogeneous Recrystallization," Acta Metallurgica 1976, 24, pp. 391–398.

24. F. Haessner, "Systematic Survey and Basic Problems of Recrystallization," in: *Recrystallization of Metallic Materials* (F. Haessner, ed.), Dr. Riederer Verlag, Stuttgart, 1978, pp. 1–10.

25. R. A. Vandermeer and P. Gordon, "The Influence of Recovery on Recrystallization of Aluminum," in: *Recovery and Recrystallization of Metals* (L. Himmel, ed.), AIME, New York, 1963, pp. 211–240.

26. H. P. Stüwe, *Einführung in die Werkstoffkunde*. 2. *Auflage*, BI − Wissenschaftsverlag, Mannheim, 1991, pp. 117–130 (ISBN 3-411-05467-0).

27. A. F. Padilha and I. G. S. Falleiros, "Recristalização em uma liga de alumínio contendo dispersão de precipitados," Metalurgia-ABM (Brazil), 1979, 35, pp. 587–590 (in Portuguese).

28. F. C. Pimenta, Jr., A. C. F. Arruda, and A. F. Padilha, "Resistance to Recrystallization in Al–1%Mn Alloys," Zeitschrift für Metallkunde 1986, 77, pp. 522–528.

29. E. A. Simielli, R. L. Plaut, and A. F. Padilha, "Influence of Heating Time on the Recrystallization of Two Aluminium Alloys," Zeitschrift für Metallkunde 1987, 78, pp. 770–776.

30. P. H. Thornton and R. W. Cahn, "The Recovery and Recrystallization of Copper and Aluminum Under Stress," Journal of the Institute of Metals 1960–1961, 89, pp. 455–463.

31. R. L. Jones, "The Tensile Deformation of Copper Single Crystal Containing BeO Particles," Acta Metallurgica 1969, 17, pp. 229–235.

32. P. B. Hirsch and F. J. Humphreys, "Plastic Deformation of Two-Phase Alloys Containing Small Non-Deformable Particles," in: *Physics of Strength and Plasticity* (A. S. Argon, ed.), MIT Press, Cambridge, 1969, pp. 189–216 (Chapter 15).

33. R. Ebeling and M. F. Ashby, "Dispersion Hardening of Copper Single Crystals," Philosophical Magazine 1966, 13, pp. 805–834.

34. R. W. Cahn, "A New Theory of Recrystallization Nuclei," Proceedings of the Physical Society of London 1950, 63, pp. 323–336.

35. P. A. Beck, "The Formation of Recrystallization Nuclei," Journal of Applied Physics 1949, 20, pp. 633–634.

36. A. H. Cottrell, "Theory of Dislocations," in: *Progress in Metal Physics* (B. Chalmers, ed.), Vol. 4, Pergamon Press, London, 1953, pp. 251–255.

37. R. Sandström, B. Lehtinen, E. Hedman, I. Groza, and S. Karlsson, "Subgrain Growth in Al and Al–1% Mn during Annealing," Journal of Materials Science 1978, 13, pp. 1229–1242.

38. S. K. Varma and B. L. Willitis, "Subgrain Growth in Aluminum during Static Annealing," Metallurgical Transactions A 1984, 15A, pp. 1502–1503.

39. J. L. Walter and E. F. Koch, "Substructures and Recrystallization of Deformed (1 0 0) [0 0 1]-Oriented Crystals of High Purity Silicon–Iron," Acta Metallurgica 1963, 11, pp. 923–938.

40. W. Blum, C. Schlögl, and M. Meier, "Subgrain Formation and Subgrain Boundary Migration in Al–5Mg during High Temperature Deformation in the Range of Class a Behaviour in Comparison with Pure Aluminium," Zeitschrift für Metallkunde 1995, 86, pp. 631–637.

41. C. S. Smith, Discussion on Paper by Mr. J. S. Bowles and Dr. W. Boas: The Effect of Crystal Arrangements on 'Secondary Recrystallization' in Metals, Journal of the Institute of Metals 1947–1948, 74, pp. 742–758.

42. H. Hu, "Annealing of Silicon–Iron Single Crystals," in: *Recovery and Recrystallization in Metals* (L. Himmel, ed.), Interscience, New York, 1963, pp. 311–362.

43. H. Hu, "Recrystallization by Subgrain Coalescence," in: *Electron Microscopy and Strength of Crystals* (G. Thomas and J. Washburn, eds.), Interscience, New York, 1963, pp. 564–573.

44. J. C. M. Li, "Possibility of Subgrain Rotation during Recrystallization," Journal of Applied Physics 1962, 33, pp. 2958–2965.

45. R. D. Doherty and J. A. Szpunar, "Kinetics of Subgrain Coalescence−A Reconsideration of the Theory," Acta Metallurgica 1984, 32, pp. 1789–1798.

46. A. R. Jones, B. Ralph, and N. Hansen, "Subgrain Coalescence and the Nucleation of Recrystallization at Grain Boundaries in Aluminum," Proceedings of the Royal Society of London 1979, 368A, pp. 345–357.

47. A. R. Jones, B. Ralph, and N. Hansen, "Nucleation of Recrystallization in Aluminium Containing Dispersions of Alumina," Metal Science 1979, 13, pp. 149–154.

48. A. R. Jones, "Grain Boundary Phenomena during the Nucleation of Recrystallization," in: *Grain-Boundary Structure and Kinetics*, ASM, Metals Park, OH, 1980, pp. 379–425.

49. P. A. Beck and P. R. Sperry, "Strain Induced Grain Boundary Migration in High Purity Aluminum," Journal of Applied Physics 1950, 21, pp. 150–152.

50. J. E. Bailey, "Electron Microscope Observations on the Annealing Processes Occurring in Cold Worked Silver," Philosophical Magazine 1960, 5, pp. 485–497.

51. J. E. Bailey and P. B. Hirsch, "The Recrystallization Process in Some Polycrystalline Metals," Philosophical Magazine 1962, 267, pp. 11–30.

52. J. E. Bailey, "Electron Microscope Observations on Recovery and Recrystallization Processes in Cold Worked Metals," in: *Electron Microscopy and Strength of Crystals* (G. Thomas and J. Washburn, eds.), Interscience, New York, 1963, pp. 535–564.

53. R. D. Doherty, "Nucleation," in: *Recrystallization in Metallic Materials* (F. Haessner, ed.), Dr. Riederer Verlag, Stuttgart, 1978, pp. 23–61.

54. A. Kreisler and R. D. Doherty, "Structure of Well Defined Deformation Bands and Formation of Recrystallization Nuclei in Aluminium," Metal Science 1978, 12, pp. 551–560.

55. P. Fraive and R. D. Doherty, "Nucleation of Recrystallization in Compressed Aluminium: Studies by Electron Microscopy and Kikuchi Diffraction," Journal of Materials Science 1979, 14, pp. 897–919.

56. O. Dimitrov, R. Fromageau, and C. Dimitrov, "Effect of Trace Impurities On Recrystallization Phenomena," in: *Recrystallization of Metallic Materials* (F. Haessner, ed.), 2nd Edn., Dr. Riederer Verlag, Stuttgart, 1978, pp. 137–157.

57. E. J. Westerman and F. V. Lenel, "Recrystallization of Cold-Drawn Sintered Aluminum Powder," Transactions of the Metallurgical Society AIME 1960, 218, pp. 1010–1014.

58. E. Gregory and G. C. Smith, "The Effects of Internal Oxidation on the Tensile Properties of Some Silver Alloys at Room and Elevated Temperature," Journal of the Institute of Metals 1956–1957, 85, pp. 81–87.

59. M. Adashi and N. J. Grant, "The Effects of Stored Energy and Recrystallization on the Creep Rupture Properties of Internally Oxidized Copper–Alumina and Copper–Silica Alloys," Transactions of the Metallurgical Society AIME 1960, 218, pp. 881–887.

60. O. Preston and N. J. Grant, "Dispersion Strengthening of Copper by Internal Oxidation," Transactions of the Metallurgical Society AIME 1961, 221, pp. 164–173.

61. T. L. Richards and S. E. Pugh, "The Influence of Prior Heat Treatment on Recrystallization of Commercial Aluminum," Journal of the Institute of Metals 1959–1960, 88, pp. 141–143.

62. W. M. Williams and R. Eborall, "Critical Strain Effects in Cold Worked Wrought Aluminum and its Alloys," Journal of the Institute of Metals 1952–1953, 81, pp. 501–512.

63. R. D. Doherty and J. W. Martin, "The Effect of a Dispersed Second Phase on the Recrystallization of Aluminum–Copper Alloys," Journal of the Institute of Metals 1962–1963, 91, pp. 332–338.

64. R. D. Doherty and J. W. Martin, "The Effect of Dispersed Second Phase on Recrystallization of Al–Cu Alloys," Journal of the Institute of Metals 1963, 92, pp. 124–125 (discussion).

65. R. D. Doherty and J. W. Martin, "Recrystallization in Two Phase Aluminum Copper Alloys," Transactions Quarterly ASM 1964, 57, pp. 874–881.

66. A. T. English and W. A. Backofen, "Recrystallization in Hot-Worked Silicon–Iron," Transactions of the Metallurgical Society AIME 1964, 230, pp. 396–407.

67. P. R. Mould and P. Cotterill, "The Effect of Particle Content and Matrix Grain Size on the Recrystallization of Two Phase Aluminum–Iron Alloys," Journal of Materials Science 1967, 2, pp. 241–255.

68. F. J. Humphreys, "The Nucleation of Recrystallization at Second Phase Particles in Deformed Aluminium," Acta Metallurgica 1977, 25, pp. 1323–1344.

69. J. D. Embury, "Plastic Flow in Dispersion Hardened Materials," Metallurgical Transactions A 1985, 16, pp. 2191–2200.

70. A. S. Smith, "Grain, Phases and Interfaces: An Interpretation of Microstructure," Transactions of the Metallurgical Society AIME 1948, 175, pp. 15–51.

71. U. Köster, "Recrystallization Involving a Second Phase," Metal Science 1974, 8, pp. 151–160.

72. E. Hornbogen and U. Köster, "Recrystallization of Two-Phase Alloys," in: *Recrystallization in Metallic Materials* (F. Haessner, ed.), Dr. Riederer Verlag, Stuttgart, 1978, pp. 159–194.

73. H. P. Stüwe, "Driving and Dragging Forces in Recrystallization," in: *Recrystallization in Metallic Materials* (F. Haessner, ed.), Dr. Riederer Verlag, Stuttgart, 1978, pp. 11–21.

74. P. A. Beck, M. L. Holzworth, and P. R. Sperry, "Effect of a Dispersed Second Phase on Grain Growth in Al–Mn Alloys," Transactions of the Metallurgical Society AIME 1949, 180, pp. 163–192.

75. H. P. Stüwe and J. Faustmann, *Introduccion a las texturas de los materiales metalicos*, Editorial Montecorvo, Madrid, 1969.

76. O. Engler, "Nucleation and Growth during Recrystallization of Aluminium Alloys Investigated by Local Texture Analysis," Materials Science and Technology 1996, 12, pp. 859–872.

77. J. Hjelen, R. Orsund, and E. Nes, "On the Origin of Recrystallization Textures in Aluminum," Acta Metallurgica e Materialia 1991, 39, pp. 1377–1404.

78. V. Randle and O. Engler, *Introduction to Texture Analysis:Macrotexture,Microtexture and Orientation Mapping* (Case study no. 12: Texture and Earing in Aluminum Sheets), Gordon and Breach Science Publishers, Amsterdam, 2000, pp. 339–342 (ISBN: 90-5699-224-4).

79. J. Hirsch and O. Engler, "Recrystallization and Texture in Industrial Aluminium Sheet," in: *Recrystallization and Grain Growth* (G. Gottstein and D. A. Molodov, eds.), Vol. 2, Springer-Verlag, Berlin, 2001, pp. 731–740 (ISBN: 3-540-41837-7).

80. F. A. Underwood, *Texture in Metal Sheets*, McDonald, London, 1961, pp. 90–111.

81. H. J. McQueen and J. J. Jonas, "Recovery and Recrystallization during High Temperature Deformation," in: *Treatise on Materials Science and Technology* (R. J. Arsenault, ed.), Vol. 6, Academic Press, New York, 1975, pp. 393–493.

82. H. J. McQueen, "The Production and Utility of Recovered Dislocation Substructures," Metallurgical Transactions A 1977, 8A, pp. 807–823.

83. H. J. McQueen and D. L. Bourell, "Hot Workability of Metals and Alloys," Journal of Metals 1987, 9, pp. 28–34.

84. H. J. McQueen, E. Evangelista, and N. D. Ryan, "Dynamic Recrystallization and Recovery: Mechanical and Kinetic Behavior: Nucleation and Growth Mechanisms," in: *Recrystallization'90* (T. Chandra, ed.), TMS-AIME Publ., Warrendale, P.A., 1990, pp. 89–99.

85. C. M. Sellars, "Modelling Microstructural Development during Hot Rolling," Materials Science and Technology 1990, 6, pp. 1072–1081.

86. I. Gutierrez, F. Castro, J. J. Urcola, and M. Fuentes, "Recrystallization during Deformation of a Commercial Aluminum Alloy," Zeitschrift für Metallkunde 1990, 81, pp. 136–143.

87. B. Davenport, N. J. Silk, C. N. Sparks, and C. M. Sellars, "Development of Constitutive Equations for Modelling of Hot Rolling," Materials Science and Technology 2000, 6, pp. 539–547.

88. H. Shi, A. J. Mc Laren, C. M. Sellars, R. Shahani, and R. Boligbroke, "Constitutive Equations for High Temperature Flow Stress of Aluminium Alloys," Materials Science and Technology 1997, 13, pp. 210–216.

89. H. J. McQueen and J. Belling, "Constitutive Constants for Hot Working of Al–4.5Mg–0.35Mn (AA5182)," Canadian Metallurgical Quarterly 2000, 39, pp. 483–492.

90. F. R. Castro-Fernandez and C. M. Sellars, "Static Recrystallization and Recrystallization during Hot Deformation of Al–1Mg–1Mn Alloy," Materials Science and Technology 1988, 4, pp. 612–627.

91. G. A. Henshall, M. E. Kassner, and H. J. McQueen, "Dynamic Restoration Mechanisms in Al–5.8 at. pct. Mg Deformed to Large Strains in Solute Drag Regime," Metallurgical Transactions A 1992, 23A, pp. 881–889.

92. H. J. McQueen, N. D. Ryan, E. V. Konopleva, and X. Xia, "Formation and Application of Grain Boundary Serrations," Canadian Metallurgical Quarterly 1995, 34, pp. 219–229.

93. P. Ponge, M. Bredehöft, and G. Gottstein, "Dynamic Recrystallization in High Purity Aluminum," Scripta Materialia 1997, 37, pp. 1769–1775.

94. H. Yamagata, "Dynamic Recrystallization and Dynamic Recovery in Pure Aluminum at 583 K," Acta Metallurgica et Materialia 1995, 43, pp. 723–729.

95. H. Yamagata, Y. Ohuchida, N. Saito, and M. Otsuka, "Nucleation of New Grains during Discontinuous Dynamic Recrystallization of 99.998 mass% Aluminum at 453 K," Scripta Materialia 2001, 45, pp. 1055–1061.

96. T. Sheppard and M. G. Tutcher, "Effect of Process Parameters on Structure and Properties of Al–5Mg–0.8Mn alloy (AA5456)," Metals Technology 1981, 8, pp. 319–327.

7

Modeling of Microstructural Evolution During Processing of Aluminum Alloys

BALA RADHAKRISHNAN and GORTI SARMA

Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.

CHRIS H. J. DAVIES

Monash University, Victoria, Australia

Significant microstructural changes occur during commercial thermomechanical processing of aluminum alloys. In the so-called non-heat-treatable alloys comprising the $1xxx$, $3xxx$, and $5xxx$ alloys, microstructural control essentially involves obtaining the right grain size and crystallographic texture, since there is no scope for hardening these alloys through precipitation. In these alloys, strength is mainly controlled through grain size, and texture plays a major role in controlling the formability of sheet products. In the heat-treatable alloys that include the $2xxx$, $6xxx$, and $7xxx$ series, strengthening is mainly due to precipitation phenomena. However, the crystallographic texture and grain size are strongly influenced by the precipitation processes that occur concurrently during recrystallization. Therefore, there is a greater flexibility in processing these alloys to produce a variety of microstructures and properties. In the past, significant progress has been achieved in the process control and product optimization of specific industrial alloys, through the development of empirical models that are either alloy-specific or process-specific or both. The need for developing physically based models that are more fundamental in nature and have predictive capability over a range of alloy chemistries and processing conditions has resulted in the development of several new modeling and simulation tools, which, along with the availability of increased computational power, are beginning to provide fundamental insights into the evolution of

microstructure during the processing of aluminum alloys. The goal of this chapter is first to provide a brief description of the empirical and analytical models that were developed in the past to understand the microstructure and texture development during deformation, recrystallization, and precipitation in aluminum alloys. This is followed by a description of the more recent developments in microstructure modeling with regard to each of the above phenomena.

1   DEFORMATION

Deformation of aluminum alloys to produce desirable microstructure and texture in sheet products is particularly important in the thermomechanical processing of 1*xxx*, 3*xxx*, and 5*xxx* aluminum alloys that are non-heat treatable. In the 2*xxx*, 6*xxx*, and 7*xxx* alloys, the final properties are controlled more by natural or artificial aging heat treatments. Therefore, the section on heat treatment will deal mainly with these alloy types. The section on deformation processing will focus primarily on the non-heat-treatable alloys. Also, aluminum metal matrix composites and deformation processing of composites through semisolid processing are not considered. Also not included in this section are superplastic forming and semisolid forming of aluminum alloys and aluminum metal matrix composites. Only those models that deal in some way with the development of microstructure during processing and its consequence to alloy properties are considered.

1.1   Analytical Models

One of the earliest models developed that provides an empirical relationship between hot-working parameters and flow stress of the material during hot deformation is the one by Sellars and Tegart [1] based on the equation

$$\sigma_s = \frac{1}{\alpha'} \operatorname{Arcsinh} \left(\frac{Z}{A'} \right)^{1/n} \tag{1}$$

where A', α', and n are temperature-independent material constants, and Z is the Zener–Hollomon parameter given by $Z = \dot{\varepsilon} \exp(U/kT)$, where $\dot{\varepsilon}$ is the applied strain rate, T is the temperature, k is the Boltzmann constant, and U is the activation energy. It has to be noted that Eq. (1) is purely empirical, and it does not provide a mechanistic understanding of hot deformation. Nevertheless, it provided a good fit to experimental data relating the flow stress of aluminum alloys to the Zener–Hollomon parameter. An interesting observation is that the above equation applies only to high temperatures, since it erroneously predicts an infinite saturation stress at 0K. Therefore, there was a strong motivation for developing physically based theories to predict the strength parameters of metals during hot working.

The earliest models that dealt with some aspect of material microstructure during deformation are those that provided an analytical description of the dislocation density and its evolution during deformation. These so-called single-parameter models date back to the works of Seeger and coworkers on work hardening and the works of Kuhlmann-Wilsdorf. Detailed descriptions of these models and their applicability to understanding the work hardening characteristics of aluminum can be found in an excellent review of this topic [2]. The focus of these

works was the analytical description of the dislocation mechanisms responsible for stage I, stage II, and stage III hardening in metallic systems. The bases for these models that have the average dislocation density as the single parameter are the following equations:

$$\sigma \frac{d\sigma}{d\varepsilon} = \frac{(\alpha\mu b)^2}{2} \frac{d\rho}{d\varepsilon} \tag{2}$$

and

$$\frac{d\rho}{d\varepsilon} = \frac{1}{\Lambda b} - L_r N_r \frac{v_r}{\dot{\varepsilon}} \tag{3}$$

where ρ is the average dislocation density, μ is the shear modulus, b is the Burgers vector of the dislocation, ε is the plastic strain, α is a constant, Λ is the mean free path of the dislocations, L_r is the mean length of the dislocation segment that recovers, N_r is the number of dislocation segments, and v_r is the annihilation rate of the dislocation segment, which is a function of the stress, temperature, and strain rate. Equation (2) expresses the work hardening rate as a function of the rate of change of the dislocation density with strain. The evolution of the dislocation density with strain as shown in Eq. (3) has two parts: an "athermal" part that contributes to the linear hardening in stage II and a temperature- and strain-rate-dependent term that is responsible for the deviation from stage II and the onset of stage III. Mecking and Kocks [3] have modified the above model by considering that the rate of change of dislocation density with strain is a function of not only of strain but also the strain rate and temperature. The mechanism of dislocation rearrangement has been historically attributed to cross-slip, although more recent theories suggest that dynamic recovery may be rate-controlled by dislocation climb mediated by vacancies produced during plastic deformation [4].

Following the above one-parameter model, two-parameter models for describing the dislocation density in deformed metals were developed [5]. These models form the first "microstructural" models that explicitly considered the deformation microstructure in the form of a cell/subgrain boundary structure. The total dislocation density ρ was assumed to be made up of ρ_b, the boundary dislocation density, and ρ_i, the dislocation density in the interior of the cells. Dynamic recovery was then modeled as a subgrain growth problem for the migration of the cell boundaries, and the evolution of the dislocation network inside the cells was modeled as before, based on cross-slip or climb aided by vacancy production during deformation.

In the two-parameter models, the flow stress of the material during hot working is defined by

$$\tau = \tau_i + \alpha_1 \mu b \sqrt{\rho_i} + \alpha_2 \mu b \delta_s^{-1} \tag{4}$$

where τ_i is the lattice friction stress, α_1 and α_2 are constants, b is the Burgers vector, and δ_s is the steady-state cell size. Based on the observation of Sellars and coworkers [6] that the separation of the dislocations within the cells, $1/\sqrt{\rho_i}$, was proportional to δ_s, Eq. (4) can be rewritten using δ_s as the only independent variable as follows:

$$\tau = \tau_i + \alpha_3 \mu b \delta_s^{-1} \tag{5}$$

where α_3 is a constant. Since δ remains constant, the following equation holds:

$$\left[\frac{\mathrm{d}\delta^-}{\mathrm{d}t}\right]_{\dot{\varepsilon},T} + \left[\frac{\mathrm{d}\delta^+}{\mathrm{d}t}\right]_{\dot{\varepsilon},T} = 0 \tag{6}$$

where the first term on the left-hand side represents the dislocation accumulation term, and the second term represents the dislocation annihilation due to dynamic recovery. The first term is calculated based on the observation by Gil Sevillano et al. [7] that during cold deformation, the cell size decreases linearly with strain, γ, in stage IV of work hardening according to

$$\frac{1}{\delta} = \frac{(0.7 + 0.09\gamma)}{\delta_{\varepsilon=1.5}} \tag{7}$$

Differentiating Eq. (7) with respect to time and recognizing that $\delta_{\varepsilon=1.5}$ is proportional to the saturation cell size δ_s, we have

$$\left[\frac{\mathrm{d}\delta^-}{\mathrm{d}t}\right]_{\dot{\varepsilon},T} = \alpha_4 \frac{\dot{\gamma}\delta^2}{\delta_s} \tag{8}$$

where α_4 is a constant.

The second term on the left-hand side of Eq. (6) is calculated by assuming that subgrain growth is controlled by solute drag, which gives an expression of the form [8]

$$\frac{\mathrm{d}\delta^+}{\mathrm{d}t} = \alpha_5 \exp\left(-\frac{U_i^b}{kT}\right) \sinh\left(\frac{PV_a^b}{kT}\right) \tag{9}$$

where α_5 is a constant, P is the driving pressure for subgrain growth given by $2\gamma_{SB}/\delta$, γ_{SB} is the misorientation-dependent boundary energy, and U_i^b is the interaction energy between the dislocation core and the solute. Equating Eqs. (8) and (9) leads to the steady-state value of δ_s. The steady-state value of δ_s obtained using the above theory provides a physically based model for the steady-state cell size as well as the flow stress during hot working of aluminum alloys rather than the empirical relationship of the form $1/\delta_s = A \ln Z - B$ [6] obtained by fitting to experimental data. The theory also shows that there is no unique relationship between steady-state cell size and Z. The microstructural model has been applied to hot deformation data based on plane strain compression of AA3004 alloy and hot torsion of AA1050, and a very good agreement between steady-state flow stress and $1/\delta_s$ was obtained for the deformation conditions used [9].

The above microstructural model has been further improved by Nes [10], who developed a statistical approach to the problem of dislocation storage, and also extended the formulation to alloys that contain non-deformable dispersoids and solutes that cause solid-solution strengthening [11]. The ALFLOW model thus developed provides excellent agreement between experimental and predicted stress–strain behavior for solid-solution- and particle-containing aluminum alloys for any combination of constant strain rate and temperature [12].

1.2 Mesoscale Models

The analytical models described thus far assume the material constitutive response to be isotropic. This assumption is valid only for initially isotropic materials that are deformed to moderate plastic strains. Most metal forming operations impart fairly large plastic strain, leading to considerable anisotropy in the material. The main cause for anisotropy is the change in grain orientation and grain shape with deformation, with grain orientation playing the dominant role up to moderately large strains. In this section, we discuss models based on crystal plasticity that have been developed to include the effects of texture on the material response during deformation processing. These models are based on the assumption that plastic deformation occurs predominantly by crystallographic slip, which involves movement by shear of specific planes of atoms along specific directions. Models incorporating other modes of deformation, such as twinning or grain boundary sliding, are not considered, since they are not typically observed under the deformation conditions of interest here.

The deformation of single crystals is modeled using Schmid's law [13], which states that slip occurs when the resolved shear stress along the direction of slip on the slip plane reaches a critical value τ_{cr}^{α}. The relation between the applied stress $\boldsymbol{\sigma}$ and the resolved shear stress τ^{α} on a given slip system can be written using the Schmid orientation tensor (dyadic product of the slip direction $\boldsymbol{s}$ and slip plane normal vector $\boldsymbol{n}$) as

$$\tau^{\alpha} = \sigma_{ij} n_j^{\alpha} s_i^{\alpha} = \boldsymbol{\sigma} \cdot (\boldsymbol{s}^{\alpha} \otimes \boldsymbol{n}^{\alpha}) \tag{10}$$

Various assumptions have been used to extend the model for single crystal deformation to that of polycrystalline aggregates, and some of these are discussed in this section.

The early work in developing polycrystal models based on single crystal deformation was focused on averaging schemes for predicting the stress–strain behavior assuming random texture in the material. Even though the resulting models were restricted to isotropic materials, they formed the basis for most of the polycrystal models in use today. One of the first such models was proposed by Sachs [14], who assumed that each grain of a polycrystalline aggregate under uniaxial loading was subjected to the same stress state, and deformed by single slip on the most highly stressed slip system. Assuming a random distribution of grain orientations, Sachs computed the uniaxial yield stress to be $2.2\tau_{cr}$. Since the Sachs model imposes the same stress state on each grain (although not the same stress magnitude) without regard to compatibility among the grains, it satisfies neither equilibrium nor compatibility. The application of this model (and its various extensions) to the processing of aluminum alloys has been limited [15–19], and has been mainly to compare the predictions of deformation textures during rolling with those of Taylor-type theories discussed below.

In addition to violating compatibility and equilibrium, the Sachs model also considered deformation to occur only on a single slip system with the highest stress. As pointed out by von Mises [20], in the absence of any volume change, an arbitrary shape change in a body requires five independent strain components. For a crystalline material, this implies five independent slip systems to accommodate a general deformation. The polycrystal models most commonly used for simulating

deformation textures are based on a concept originally proposed by Taylor [21] using this criterion. In this approach, all grains of a polycrystalline aggregate are assumed to undergo the same strain, which is the macroscopic average strain. The stress in each grain needed to accommodate the imposed strain is determined using crystal plasticity, and a weighted average of the grain stresses is used to compute the macroscopic stress. The Taylor model satisfies compatibility but violates equilibrium across grain boundaries, thus providing an upper bound for the stress. In Taylor's original work based on analyzing uniaxial deformation of single-phase fcc poly-crystals, the criterion for determining the proper combination of slip systems to accommodate the strain in each grain was based on minimizing the net internal work. Assuming a random texture and equal strain hardening for all grains, Taylor found the tensile yield stress to be $3.06\tau_{cr}$, which was much closer to the experimental data than Sachs' calculation. Bishop and Hill [22,23] employed an alternate approach, using the principle of maximum work to compute single crystal and polycrystal yield surfaces, and, in the process, came up with a different scheme to determine the active slip systems under Taylor's uniform strain assumption. They also extended the model to include general deformation conditions.

The Taylor–Bishop–Hill model had the limitation that the selection of active slip systems was not unique, and this problem was overcome by Asaro and Needleman [24] through the introduction of rate dependence into the Taylor model. A power law relates the resolved shear stress τ^{α} to the rate of shearing $\dot{\gamma}^{\alpha}$ on each slip system:

$$\tau^{\alpha} = \tau^{\alpha}_{cr} \frac{\dot{\gamma}^{\alpha}}{\dot{\gamma}_0} \left| \frac{\dot{\gamma}^{\alpha}}{\dot{\gamma}_0} \right|^{m-1} \tag{11}$$

where $\dot{\gamma}_0$ is the reference rate of shearing, and m is the strain rate sensitivity parameter. If the formulation is simplified to neglect elastic deformations, the crystal rate of deformation, which is the symmetric portion of the velocity gradient, can be written as a linear combination of the slip system shear rates,

$$\mathbf{D}_c = \sum_{\alpha} \mathbf{P}^{\alpha} \dot{\gamma}^{\alpha} \tag{12}$$

where $\mathbf{P}^{\alpha}$ is the symmetric part of the Schmid orientation tensor. Using Eq. (11) for the shear rate and Eq. (10) for the resolved shear stress, the above equation can be rewritten to develop a linearized constitutive relation for the crystal,

$$\mathbf{D}_c = \left[\sum_{\alpha} \left| \frac{\boldsymbol{\sigma} \cdot \mathbf{P}^{\alpha}}{\tau^{\alpha}_{cr}} \right|^{(1/m)-1} \frac{\dot{\gamma}_0}{\tau^{\alpha}_{cr}} \mathbf{P}^{\alpha} \otimes \mathbf{P}^{\alpha} \right] \boldsymbol{\sigma} \tag{13}$$

which can be solved to obtain the crystal stress for a given rate of deformation. The crystal spin $\mathbf{W}_c$ is a combination of the plastic spin due to slip system shearing and any rigid rotation $\mathbf{R}^*$ required to maintain compatibility with other crystals,

$$\mathbf{W}_c = \dot{\mathbf{R}}^* \mathbf{R}^{*^T} + \sum_{\alpha} \mathbf{Q}^{\alpha} \dot{\gamma}^{\alpha} \tag{14}$$

where $\mathbf{Q}^\alpha$ is the skew-symmetric part of the Schmid orientation tensor. Equation (14) can be rewritten to obtain the rate of reorientation of the crystal, which in turn determines the texture evolution in the aggregate,

$$\dot{\mathbf{R}}^* = \left(\mathbf{W}_c - \sum_\alpha \mathbf{Q}^\alpha \dot{\gamma}^\alpha\right)\mathbf{R}^* \tag{15}$$

Strain hardening is modeled through an evolution equation for the critical resolved shear stress based on the accumulated shear rate on all slip systems,

$$\dot{\tau}^\alpha_{cr} = \sum_\beta H_{\alpha\beta}\left|\dot{\gamma}^\beta\right| \tag{16}$$

where $H_{\alpha\beta}$ is a hardening matrix.

The models discussed thus far assume equiaxed grains of uniform size in the aggregate, and effects of grain shape are neglected. However, this assumption may not be valid for materials subjected to large plastic strains. For instance, the large reductions in thickness during rolling can lead to flat and elongated grains, and for such deformation conditions, the assumption of uniform strain has been modified by "relaxing" the strict compatibility requirements of the Taylor model [25–28]. In the relaxed constraints models, only some of the strain components are required to be strictly compatible, while for the remaining components, traction continuity is enforced by relaxing the compatibility requirement on the strains, and requiring the average strain over the grains to equal the macroscopic strain.

In addition to incorporation of grain shape effects, the models based on the relaxed constraints theory introduce some deviation in the strain components of individual grains from the macroscopic strain. In reality, interactions with neighboring grains might be expected to cause deviations both among and within individual grains from the average macroscopic strain (and stress). There have been several attempts to introduce the effects of grain interactions into the polycrystal averaging schemes. The most widely used among these is the "self-consistent" model, based on extending the ideas of Kroner [29], Budiansky and Wu [30], and Hill [31]. In these models, each grain (or a small cluster of grains) is treated as an inclusion embedded in a homogeneous effective medium (HEM) whose properties are given by the average properties over all the grains in the aggregate. The HEM forms the local neighborhood with which each grain (or cluster) is assumed to interact, and the partitioning of the deformation is accomplished such that the average value over all the grains or clusters equals the macroscopic value. The self-consistent model has been mainly applied to model the deformation of materials with lower symmetry, such as hcp crystals, and its use for fcc and bcc polycrystals has been rather limited [32,33]. The effects of a local neighborhood on the inhomogeneous deformation of the aggregate are considered in the "neighborhood compliance" method discussed in Sarma and Dawson [34], wherein the compliance of each grain relative to its immediate neighbors is used to compute its rate of deformation and spin. While these methods introduce variability in deformation between different grains, they fall short of considering variations within individual grains, leaving such treatment to detailed finite element models, as described later in the section.

The discussion has thus far focused on application of polycrystal models to analyze deformation processing through approximations of homogeneous macroscopic strains applied on a polycrystalline aggregate. We now consider implementations of these polycrystal models as constitutive laws within the framework of finite element formulations, to solve initial-boundary value problems associated with bulk deformation processing. In these types of simulations, the discretization of the deforming workpiece is carried out at the macroscopic or continuum length scale using the finite element method (FEM). The constitutive response of the material in each finite element, typically at each integration point, is obtained by averaging the response of a polycrystalline aggregate associated with the element or integration point. The anisotropy in the material response thus directly influences the heterogeneous deformation field in the workpiece. One of the first examples of such an implementation using the Taylor model was reported by Mathur and Dawson [35], who analyzed texture evolution during steady-state cold rolling of aluminum. This formulation was subsequently modified to include grain shape effects using the relaxed constraints model [36]. More recently, Aretz et al. [37] used this methodology to study rolling of aluminum alloy AA3104. A similar approach was used to investigate the effect of anisotropy on the formability of rolled sheet by Beaudoin et al. [38]. Several authors have used the Taylor model within finite element formulations to investigate the phenomena related to earing during deep drawing [39–42] and hydroforming [43]. The influence of plastic anisotropy on damage due to localized deformation during deep drawing of rolled aluminum sheet has been analyzed by Hu et al. [44].

Figure 1 shows the effective plastic strain contours from a simulation of sheet stretching using a single layer of 500 eight-node brick elements. Each element was assumed to contain 256 crystals, whose initial orientations were generated to match the texture of a rolled sheet based on experimental measurements. The simulations predict a non-uniform strain distribution due to the texture-driven anisotropy, with maximum plastic strain along the x-axis. The sheet has been observed to fail along the x-axis, which corresponds to the rolling direction in the original sheet, during limiting dome height tests.

In addition to the direct inclusion of polycrystal models within the finite element framework, several authors have also reported on a different approach to modeling bulk deformation textures. In these efforts, the FEM is used to model the deformation process using simpler phenomenological constitutive laws, and the results are used to provide the deformation history for certain locations of interest within the workpiece. Polycrystal models are then used to evolve the texture by applying the deformation history to an aggregate of grains, to study the effect of the heterogeneous deformation fields on the texture evolution and its variation within the workpiece. Examples include simulations of rolling [45] and extrusion [46,47] of aluminum alloys using the Taylor and relaxed constraints models.

The use of constitutive models based on crystal plasticity within finite element simulations of bulk deformation processing led to the ability to model the effects of anisotropy due to crystallographic texture on the material response. However, these constitutive laws still require the adoption of some sort of averaging scheme (such as the Taylor hypothesis) in order to relate the single crystal response to the behavior of the aggregate. In addition, although some of the models incorporate grain-to-grain variations in deformation, they assume uniform deformation within each grain. In
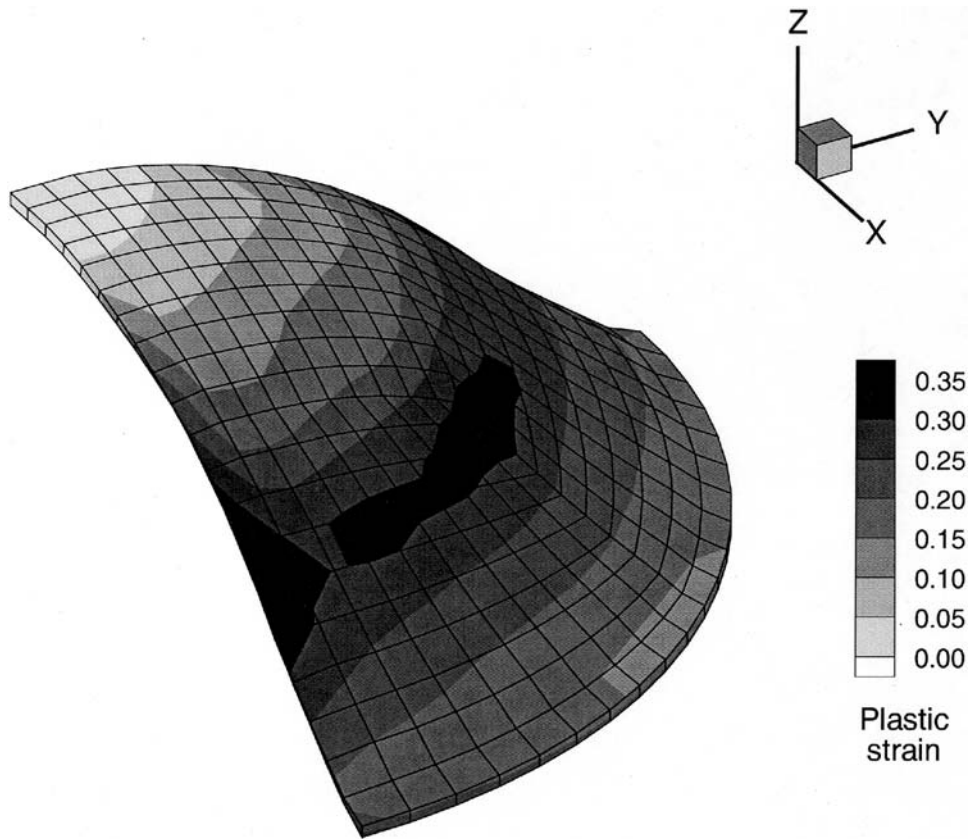
Figure 1 FE simulation of sheet stretching process showing non-uniform distribution of effective plastic strain due to texture-driven plastic anisotropy.

recent years, another type of coupling between the finite element method and crystal plasticity models has been used to gain greater insights into the evolution of material microstructures during deformation processing. The finite element discretization is applied directly to a polycrystalline aggregate, with many elements used to discretize each grain, so that topological effects are directly incorporated into the model. The microstructure is typically deformed by application of boundary conditions corresponding to a homogeneous deformation, and the constitutive response in each element is obtained based on crystal plasticity. No assumptions are required for relating the deformation of each crystal to that of the polycrystal. Instead, the deformation of each element is computed from the discretized deformation or velocity field, based on its position within the domain and on its material state (characterized by its crystallographic orientation and its slip system strength). The simulations permit the development of non-uniform deformations within each grain, depending on its size, orientation, and interactions with neighboring grains. Several investigators have used this approach to study the deformation of single crystals and polycrystals of aluminum, using both two-dimensional (2-D) and three-dimensional (3-D) discretizations. Early applications of this approach were limited to planar

meshes [48–52]. With the recent availability and advances in parallel computing technology, it has become possible to handle fairly large three-dimensional meshes [53–58]. Two examples of such simulations are briefly discussed here. Figure 2 shows the initial and deformed microstructure from plane strain compression of aluminum to 50% reduction in height. The initial microstructure was generated using a Monte Carlo grain growth model. The grains in the microstructure are shown using different colors. The simulations were able to capture the localized deformations, with considerable shearing seen in some elements. As discussed later, these simulations were used to provide initial data for Monte Carlo simulations of recovery and recrystallization. Figure 3 shows the texture evolution in the form of ⟨1 1 1⟩ pole figures from simulations of a single crystal initially close to the cube orientation deformed in plane strain compression. The simulations were carried out without and with the non-octahedral {1 1 0} ⟨1 1 0⟩ slip systems, in addition to the usual {1 1 1} ⟨1 1 0⟩ slip systems. As seen from the textures, the presence of the {1 1 0} ⟨1 1 0⟩ slip systems enhances the stability of the cube orientation, and has been used to explain the stability of the cube texture component during hot deformation of aluminum. This approach holds a lot of promise in providing greater knowledge and understanding of the microstructure and microtexture evolution in metals as a consequence of different processing conditions.
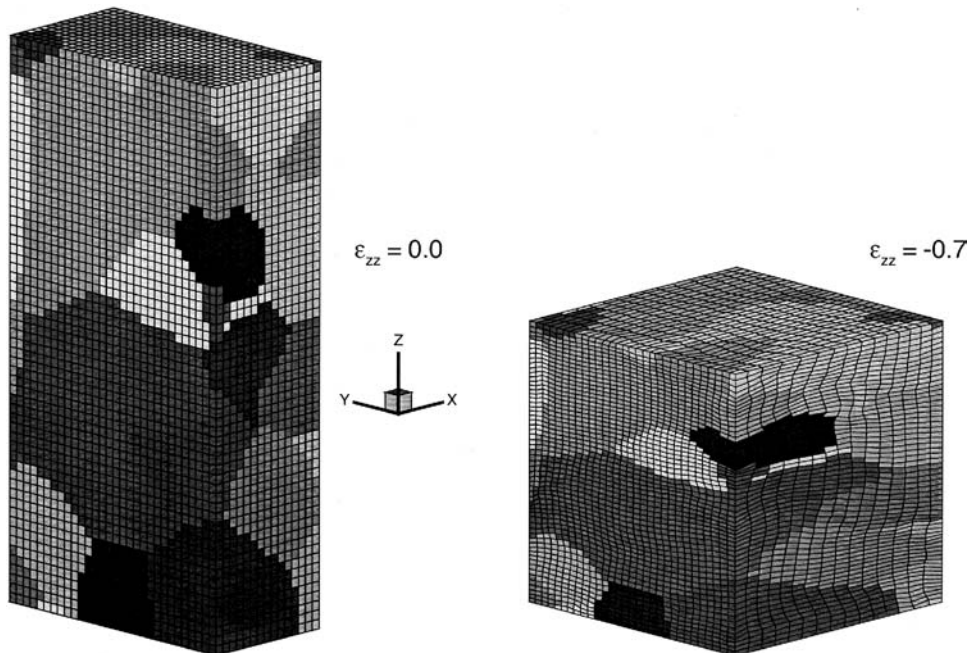
Figure 2   FE simulation of microstructural deformation in polycrystalline aluminum using crystal plasticity approach, showing non-uniform deformation at grain level. (From Ref. [55], reprinted with permission from Institute of Physics Publishing.)
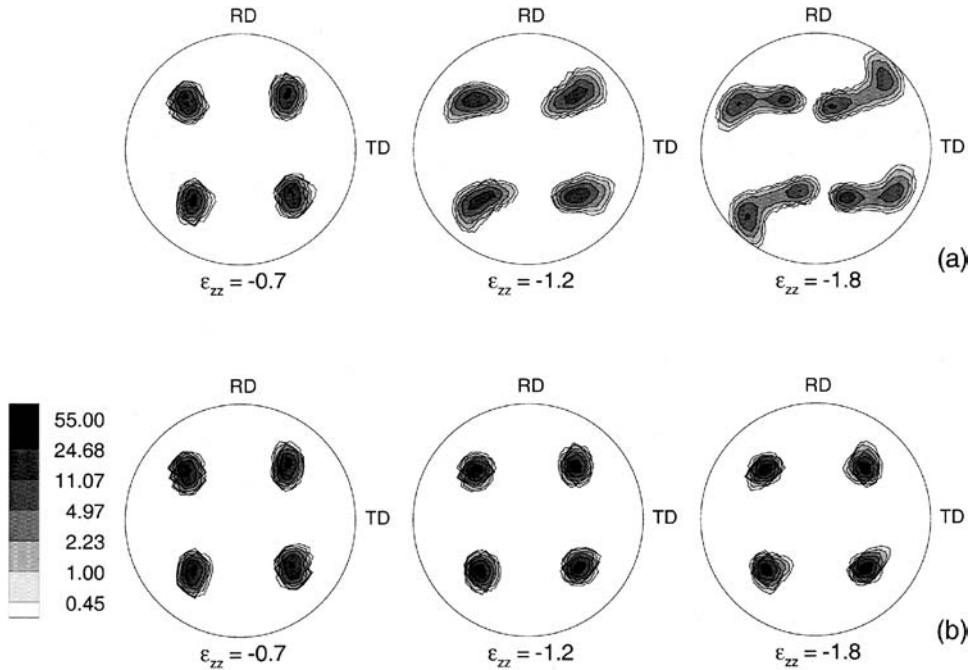
Figure 3 $\langle 1\ 0\ 0 \rangle$ pole figures of deformed single crystal initially close to cube orientation deformed using (a) $\{1\ 1\ 1\}\ \langle 1\ 1\ 0\rangle$ slip systems and (b) $\{1\ 1\ 1\}\ \langle 1\ 1\ 0\rangle$ and $\{1\ 1\ 0\}\ \langle 1\ 1\ 0\rangle$ slip systems. Stability of cube orientation increases when slip occurs on both $\{1\ 1\ 1\}\ \langle 1\ 1\ 0\rangle$ and $\{1\ 1\ 0\}\ \langle 1\ 1\ 0\rangle$ slip systems. (From Ref. [55], reprinted with permission from Institute of Physics Publishing.)

2 RECRYSTALLIZATION

The following section deals with the analytical models of dynamic and static recrystallization in aluminum and its alloys. The conventional form of dynamic recrystallization, also known as discontinuous dynamic recrystallization (DDRX), that involves the nucleation and growth of new grains during deformation rarely occurs in aluminum and its alloys. The only exceptions are high-purity aluminum (99.999% purity), where the grain boundary mobilities are extremely high [59], and also particle-containing alloys, such as Al–Mg–Mn alloys, where localized deformation occurs in the vicinity of Al_6Mn particles that promotes particle-stimulated nucleation (PSN) [60]. However, in general, DDRX plays an insignificant role in the hot deformation of commercial aluminum alloys.

A more common form of dynamic recrystallization in aluminum alloys is the so-called continuous dynamic recrystallization process (CDRX), where high-angle grain boundaries form from low-angle subgrain boundaries through a continuous process. One form of CDRX that is reported to occur in aluminum alloys after severe plastic deformation is one where the misorientation between the initial subgrains continues to increase with the accumulation of grain boundary dislocations to the point where they all become high-angle boundaries. Such a phenomenon has been reported [61] for AA1200, AA5052, and single crystals of pure aluminum during hot deformation under uniaxial compression and torsion.

However, the same experimental results have been explained largely on the basis of dynamic recovery [62]. Therefore, there seems to be a controversy in the literature over this issue. However, there is another mechanism of dynamic recrystallization called geometric dynamic recrystallization (GDRX), where the original grain boundary segments form serrations during subgrain formation, and these serrations either recombine or pinch off to form nearly equiaxed grains with high misorientations [63]. Such a mechanism is operative only under certain conditions of hot working where the subgrain size is equal to the grain size in the thickness direction. Geometric dynamic recrystallization has been shown to occur in Al–5% Mg and Al–5% Mg–0.6% Mn alloys. The mechanism does not occur if the alloy contains a large volume fraction of small precipitates that can pin down the grain boundaries and does not allow the formation of serrations. A detailed discussion of the modeling of such dynamic recrystallization phenomena is beyond the scope of this review. We will focus mainly on discontinuous static recrystallization as it occurs between deformation steps in multi-stage hot rolling and also annealing of cold deformed microstructures.

2.1   Analytical Models

These models are mostly empirical and require the experimental determination of several constants that are alloy-specific. Nevertheless, these empirical equations are used in the industry to model recrystallization in multi-stage hot deformation processes. The kinetics of recrystallization is modeled based on the Johnson–Mehl–Avrami–Kolmogorov (JMAK) equation given by [64]

$$X_{\mathrm{v}} = 1 - \exp\left[k\left(\frac{t}{t_{0.5}}\right)^{n}\right] \tag{17}$$

where X_{v} is the volume fraction of recrystallized material, k and n are constants, and $t_{0.5}$ is the time for 50% recrystallization. $t_{0.5}$ is related to the hot deformation conditions through the following relationship [65]:

$$t_{0.5} = c_1 D_0^c \varepsilon^{-n} Z^{-k} \exp\left(\frac{Q_{\mathrm{rex}}}{RT}\right) \tag{18}$$

where c_1, c, n, and k are constants, D_0 is the initial grain size, and Q_{rex} is the apparent activation energy for recrystallization. Such empirical models have been used to study the extent of static recrystallization in hot deformed aluminum alloys AA3104 [65], AA5056 [66], AA5182 [67], and AA5052 [67]. The empirical microstructural model was combined with a finite element model of hot deformation to predict the variation of recrystallized fraction from the surface to center of the hot band. One of the features of these analytical models is the use of a temperature-compensated time parameter, t, as follows [67]:

$$t' = \sum \delta t_i \exp\left(\frac{-Q_{\mathrm{rex}}}{RT_i}\right) \tag{19}$$

where a thermal cycle can be broken down as a combination of several isothermal steps of duration δt_i at temperature T_i. This is analogous to the additivity approach

adopted in modeling precipitation, which will be described later in the chapter. The recrystallized grain size in these empirical models is given by

$$D_R = c_2 D_0^{c'} \varepsilon^{-n'} Z^{-k'} \tag{20}$$

where c_2, c, n, and k are empirical constants.

Figures 4 and 5 show the results using such an analytical model for the inter-pass recrystallization during commercial hot-strip rolling of AA5182 and AA5052 [67], respectively. The model shows that in AA5182, the surface recrystallizes much more quickly than the interior, because the surface experiences deformation at lower temperatures, higher strain rates, and higher strains, and therefore has a higher stored energy for recrystallization. There is a significant difference between the recrystallization kinetics of AA5182 and AA5052, which is attributed to the higher Mg content in 5182. Higher Mg content reduces the rate of dynamic recovery and increases the rate of work hardening, and therefore results in higher stored energy after deformation.

Analytical models do not provide a physical basis for microstructure evolution during recrystallization. In non-heat-treatable aluminum alloys, such as the 3xxx and 5xxx alloys, the evolution of crystallographic texture during deformation processing has a significant influence on the formability of sheets. The evolution of various deformation components during recrystallization depends on the nucleation of these components and their subsequent growth. Physically based understanding of these phenomena cannot be obtained from these empirical models.

A major development over the above analytical models is the microstructural model for static recrystallization after hot deformation developed by Vatne et al. [68,69] that provides a physical basis for the nucleation and growth of various texture components during recrystallization. The basis for the model is the large amount of experimental work on the evolution of microstructure and texture in
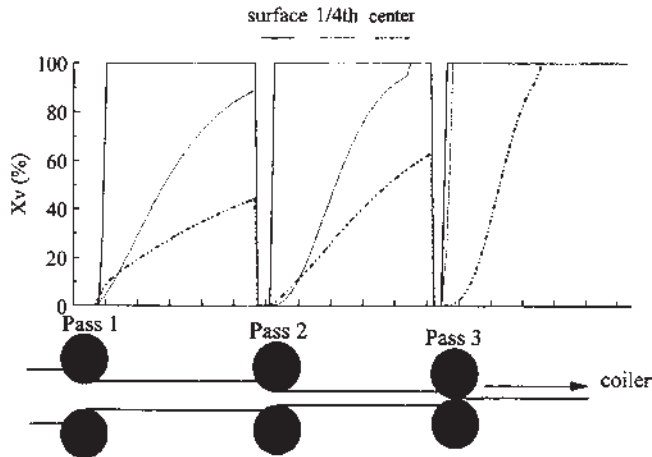
Figure 4 Through-thickness variation in recrystallization kinetics of AA5182 strip. Surface recrystallizes more quickly than interior locations due to variation in driving force for recrystallization. (From Ref. [67], reprinted with permission from TMS and ASM International.)
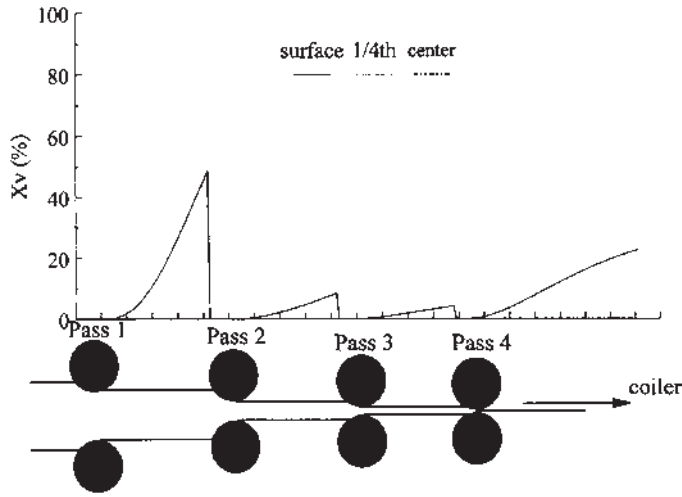
Figure 5 Inter-pass recrystallization of AA5052 during commercial hot-strip rolling. Recrystallization is slower than in AA5182 because of lower driving force for recrystallization due to decreased Mg content in alloy. (From Ref. [67], reprinted with permission from TMS and ASM International.)

commercial aluminum alloys during hot deformation [70–72]. The nucleation of three different texture components is considered in the model: (i) cube texture, (ii) texture due to growth of particle-stimulated nuclei, and (iii) grain boundary nucleated texture components. The nucleation of cube components is based on the following experimental observations.

  (i)   Initial cube grains survive the hot deformation process and are present as elongated bands in the deformed microstructure. These bands serve as potential cube nuclei during subsequent annealing
  (ii) The cubes situated next to the S deformation component have the potential to form the special high-mobility boundary which has a 40° rotation about the $\langle 1\,1\,1 \rangle$ axis.
  (iii) The deformation substructure within the cube is such that the cells have a growth advantage over other surrounding deformation components by having a long tail in the size distribution

Based on the above observations, the following equation was derived for the density of cube nucleation sites:

$$N_c = \frac{2C_c\delta_c R_c(1-R_c)R_S S_c^*}{\bar{D}}[\exp(\varepsilon) + \exp(-\varepsilon) + 1] \tag{21}$$

where ε is the plastic strain, $\bar{D}$ is the average initial size of the cube grains, R_c is the instantaneous volume fraction of cube in the as-deformed structure, R_S is the volume fraction of the S component after deformation given by the empirical equation

$$R_S = \begin{cases} 0.04 + 0.17\varepsilon & \varepsilon < 2 \\ 0.38 & \varepsilon > 2 \end{cases} \tag{22}$$

R_c is the instantaneous volume fraction of cube in the as-deformed structure given by the following empirical equation:

$$R_c = 2.5R_c^0(1 + 0.25\varepsilon)^{-1.8}(0.1\ln Z - 1.4)^{-1.8} \tag{23}$$

δ_c is the average cube subgrain size given by

$$\frac{1}{\delta_c} = \frac{RT}{A^*}\ln\left(\frac{Z\delta_c^2}{B^*}\right) \tag{24}$$

where A^* and B^* are empirical constants, S_c^* is the number of cube subgrains larger than the critical value δ^* for nucleation given by

$$S_c^* = \frac{1}{\delta_c^3}\exp(-5a)\left[1 + 5a + \frac{1}{2!}(5a)^2 + \frac{1}{3!}(5a)^3 + \frac{1}{4!}(5a)^4\right] \tag{25}$$

where $a = \delta^*/\delta_c$, and C_c is a modeling parameter of the order of unity.

The density of PSN nuclei is given by

$$N_{PSN} = C_{PSN}N_0\exp\left(\frac{4L\gamma_{GB}}{P_D - P_Z}\right) \tag{26}$$

where N_0 is a constant that depends on the size distribution of the coarse particles, γ_{GB} is the specific boundary energy between the nucleus and the deformation matrix, and $(P_D - P_Z)$ is the effective driving pressure for recrystallization, where P_D is the stored energy and P_Z is the Zener drag due to smaller particles.

The density of grain boundary nuclei that give rise to other random recrystallization texture components is given by

$$N_{GB} = C_{GB}\delta(1 - R_c S_{GB}^*)[\exp(\varepsilon) + \exp(-\varepsilon) + 1] \tag{27}$$

where δ is the average subgrain size, C_{GB} is a modeling constant, and S_{GB}^* is the density of overcritical nuclei given by an equation analogous to Eq. (25). The growth rate of each texture component is given by

$$G = M(P_D - P_Z) \tag{28}$$

where M is the temperature-dependent boundary mobility given by

$$M = M_0\exp\left(\frac{-U_{GB}}{RT}\right) \tag{29}$$

where U_{GB} is the activation energy for boundary migration. The driving pressure P_D is given by

$$P_D = \frac{Gb}{\delta}\left[\frac{\alpha\varphi}{4\pi(1-v)}\ln\left(\frac{e\varphi_c}{\phi}\right) + 0.5C_1^2\frac{b}{\delta}\right] \tag{30}$$

The recrystallized fraction is again determined using JMAK kinetics based on

$$X_t = 1 - \exp[-X_{ext}(t)] \tag{31}$$

where X_{ext} is the extended volume given by $4/3N_{tot}(Gt)^3$.

The above model by Vatne et al. is the first analytical model that combines the microstructural description of the dynamically recovered hot worked deformation substructure with the JMAK model to provide a microstructure-based kinetic model for recrystallization. In addition to predicting the kinetics of recrystallization, it also predicts the evolution of recrystallization texture as well as the recrystallized grain size. Therefore, it is a truly microstructural model that can be combined with numerical deformation and heat transfer models as before to predict the through-thickness variations in the recrystallized microstructure and texture in a hot-rolled sheet. Sellars and Zhu [73] extended the above recrystallization model to the deformation substructure obtained by transient deformation conditions of increasing or decreasing strain rate.

A main drawback in the Vatne model is that a number of alloy-specific empirical constants need to be experimentally determined before it can be applied. Also, there are other mechanistic assumptions that may not be valid. First, it is assumed that boundary mobility is independent of orientation. However, it is known from experiments and atomistic simulations that boundary mobility is not only a function of misorientation but also, for certain special misorientations, it is possible to have special boundaries that have high mobility and low energy [74]. In fact, the orientation dependence of boundary mobility gives rise to the so-called "orientation pinning" of texture components, a mechanism proposed by Juul Jensen [75]. As a result of this mechanism, a cube nucleus will enjoy unrestricted growth in the deformation structure, since the boundary between the cube and a non-cube will always remain a high-angle boundary during its growth. However, a grain boundary nucleus close to one of the deformation texture components will encounter regions of almost identical orientation, and therefore, the boundary mobility will become extremely low. The deformation substructure in many instances is characterized by long-range orientation gradients that result in the formation of nuclei by discontinuous subgrain growth. The above analytical model does not capture such mechanisms.

In polycrystalline materials subjected to deformation, the stored energy is a function of the grain orientation through the Taylor factor [76,77]. In addition to these grain-to-grain variations, the presence of constraints at grain boundaries and triple lines introduces intragranular variations in stored energy. Therefore, at a microstructural length scale, the stored energy is inhomogeneous. Inhomogeneous distribution of stored energy impacts not only the spatial distribution of the nuclei but also the growth rate by imparting a variable driving force [78]. Consequently, the assumptions of random distribution of nuclei and constant and isotropic growth rate that are inherent in the JMAK model may not be valid for real alloy systems. In fact, the incorporation of such effects into recrystallization simulations was the motivation for the development of various mesoscopic models of recrystallization, which will be discussed in a subsequent section. The prediction of recrystallization texture by the above model is limited in that it only gives the volume fractions of the cube and non-cube orientations, and it is not possible to predict the actual orientations of the non-cube components.

However, in spite of these drawbacks, the large amount of experimental work in $3xxx$ aluminum alloys has resulted in the accurate estimation of the empirical constants in the model, and in the successful application of the model to predict the evolution of cube texture in these alloys during multi-stage industrial hot working [79]. Figure 6 shows a comparison between model predictions and experimental
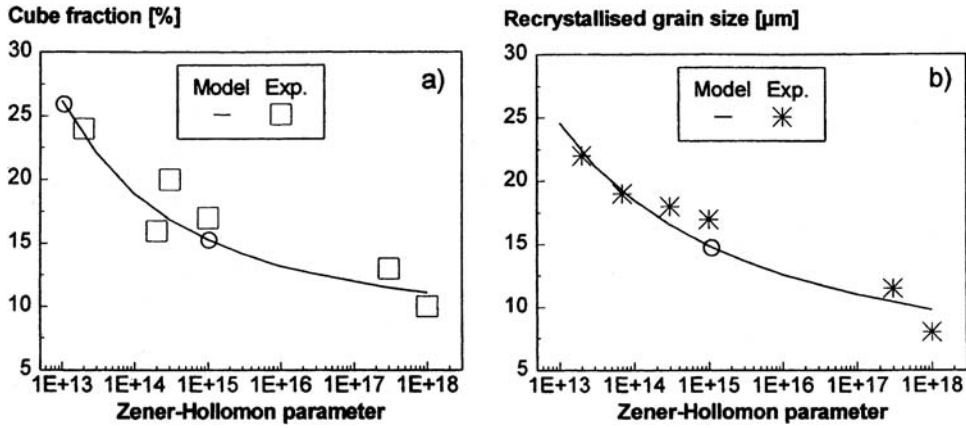
Figure 6 Model prediction and experimental data on (a) volume fraction of cube texture component and (b) recrystallized grain size in alloy AA3004 as function of Zener–Hollomon parameter. (From Ref. [69], reprinted with permission from Monterey Institute of Advanced Studies.)

results of cube fraction and recrystallized grain size after hot compression of AA3004 under plane strain conditions. At higher Z values (higher strain rates and lower temperatures), the cube volume fraction decreases because of the increased efficiency of PSN of recrystallization that gives rise to a random texture. The stability of the cube orientations during deformation also decreases with increasing Z because they tend to rotate to other orientations, probably because of the lack of slip in the {1 1 0} ⟨1 1 0⟩ slip systems. The model also predicts the experimental observation that the recrystallized cube texture increases with increasing cube in the initial micro-structure, as shown in Fig. 7. The recrystallized cube fraction increases with plastic
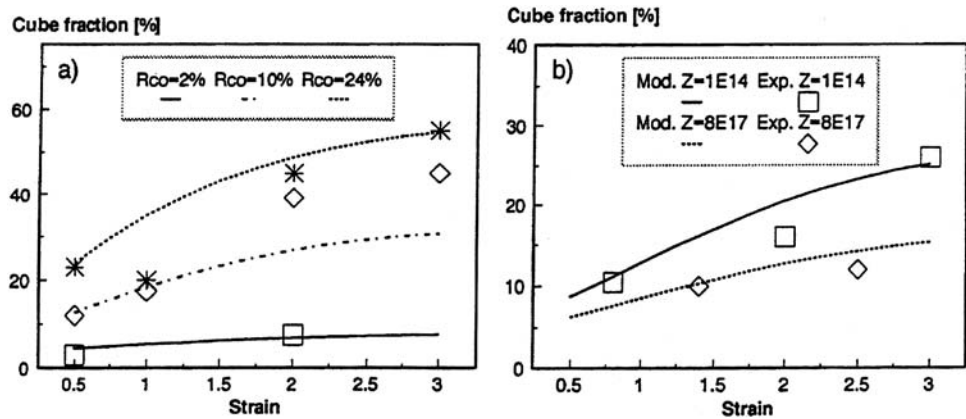
Figure 7 Comparison between model predictions and experimental data on volume fraction of cube texture component as function of plastic strain for (a) AA1050 alloy and (b) AA3004 alloy. (From Ref. [69], reprinted with permission from Monterey Institute of Advanced Studies.)

strain because of the increased surface area of the cube bands, and also due to the fact that the fraction of S deformation component increases with strain, thus increasing the probability of finding S—cube interfaces in the alloy.

2.1.1 Recrystallization Textures

One of the long-standing controversies in the prediction of recrystallization textures in aluminum alloys is whether the recrystallization texture is controlled by oriented growth (OG) or oriented nucleation (ON) [64]. In the ON concept, it is assumed that nuclei of the major recrystallization texture components form more frequently than the other minor components. This argument is supported by the Dillamore and Katoh model [80] based on cube nucleation in transition bands, and the Ridha and Hutchinson model [81] based on the preferential dynamic recovery of cube-oriented subgrains. Oriented nucleation was observed by Hjelen et al. [82] and Doherty and Panchanadeeswaran [83] for the formation of cube recrystallization texture in cold-rolled and hot-rolled aluminum, because the cube grains did not appear to have any specific growth advantage compared to other non-cube grains. Also, the fraction of cube-oriented grains after deformation was significantly higher than expected based on the assumption of random nucleation [83]. According to the OG concept, the growth rates of certain texture components present in the deformation texture are significantly higher than the growth rates of other components, so that these fast-growing orientations would consume the majority of the deformation structure during recrystallization. Experimental evidence for OG in which different growth rates were observed for nuclei with different orientation relationships with the matrix is quite large [84]. A particularly striking case of OG is the boundary between the cube grains and the S-oriented matrix, which is characterized by a $40°$ $\langle 1\ 1\ 1 \rangle$ orientation relationship. Although earlier work by Hjelen et al. [82] did not support this enhanced growth, there seems to be a quite convincing case of enhanced growth of the boundary between cube and S-grains. However, one of the manifestations of OG for cube grains could simply be the significantly less orientation pinning of the cube nuclei compared to nuclei of other deformation texture components, without an inherent enhanced boundary mobility. Recent molecular dynamics simulations of grain boundary mobility in fcc bicrystals also seem to support the fact that there is a peak in the mobility around $38°$ $\langle 1\ 1\ 1 \rangle$ boundary rotation, with an attendant reduction in the grain boundary energy, as shown in Fig. 8, taken from the work of Upmanyu [74]. More recent EBSP measurements taken in a variety of aluminum alloys show that the $40°$ $\langle 1\ 1\ 1 \rangle$ boundary does have a higher mobility than boundaries characterized by other rotations [85]. In fact, such observations led to the development of an analytical model for texture evolution, as described below.

The analytical model of Vatne has been used by Engler [86] to develop a model to predict the recrystallization textures in aluminum alloys from a knowledge of the deformation texture, based on the premise that the recrystallization texture can be described by a transformation of the deformation texture, which reflects the fact that the fastest-growing texture components during recrystallization are those that are related to the deformation texture by a $40°$ $\langle 1\ 1\ 1 \rangle$ rotation. According to this model, the probability of existence of a recrystallization texture, $f(g)^{\text{sim}}$, is given by

$$f(g)^{\text{sim}} = f(g)^{\text{nucl}} f(g)^{\text{grow}} \tag{32}$$
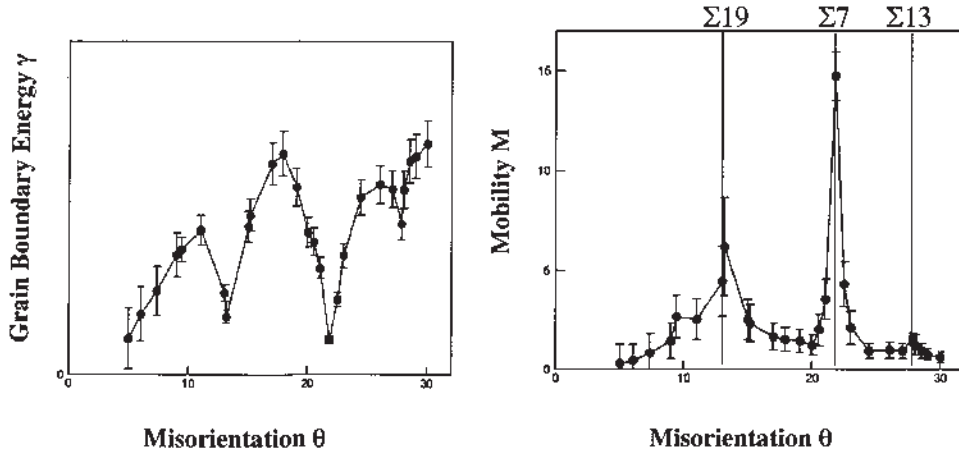
Figure 8 Variations of grain boundary energy and mobility with misorientation for {1 1 1} tilt boundaries showing minima in energy, and corresponding maxima in boundary mobility at coincidence boundaries (from Ref. [74]). For $\theta > 30°$, $\theta = 60° - \theta$.

where $f(g)^{nucl}$ is the probability of nucleation of a new grain and $f(g)^{grow}$ is the probability of growth of the nucleus. In polycrystalline aluminum alloys, nucleation can occur either in the cube bands, in grain boundaries between different deformation components, or in the vicinity of coarse hard particles. Therefore, the probability of nucleation, $f(g)^{nuc}$, is given by

$$f(g)^{nucl} = x_{Cube} f(g)^{nucl}_{Cube} + x_{GB} f(g)^{nucl}_{GB} + x_{PSN} f(g)^{nucl}_{PSN} \tag{33}$$

where x_i denotes the efficiency of nucleation at each site. The x_i's are calculated based on Vatne's model described above using Eqs. (21), (26), and (27). The nucleus orientation distribution functions for cube band, grain boundary, and PSN nucleus are obtained from experimental observations. The growth probability function, $f(g)^{grow}$, is based on the $40° \langle 1 1 1 \rangle$ rotation relationship between the deformed matrix and the orientation of the nucleus. The model has been successfully applied to predict the recrystallization textures of alloys AA1145 and AA3004 for a wide range of deformation conditions based on a knowledge of the deformation texture and the experimental determination of orientations near the grain boundaries, particle deformation zones, and cube bands [86]. Figure 9 shows the experimental and model predictions of recrystallization texture in AA3004 as a function of the Zener–Hollomon parameter, obtained using the above texture model.

Lee [87] proposed a theory for evolution of recrystallization textures according to which the orientations of the nuclei that preferentially form and grow during recrystallization are those for which the direction of minimum Young's modulus coincides with the direction of maximum internal stress. This is known as the strain energy release maximization (SERM) model. The model originates from the presumption that the stored energy due to dislocations is the major driving force for recrystallization. The absolute maximum internal stress direction may be obtained
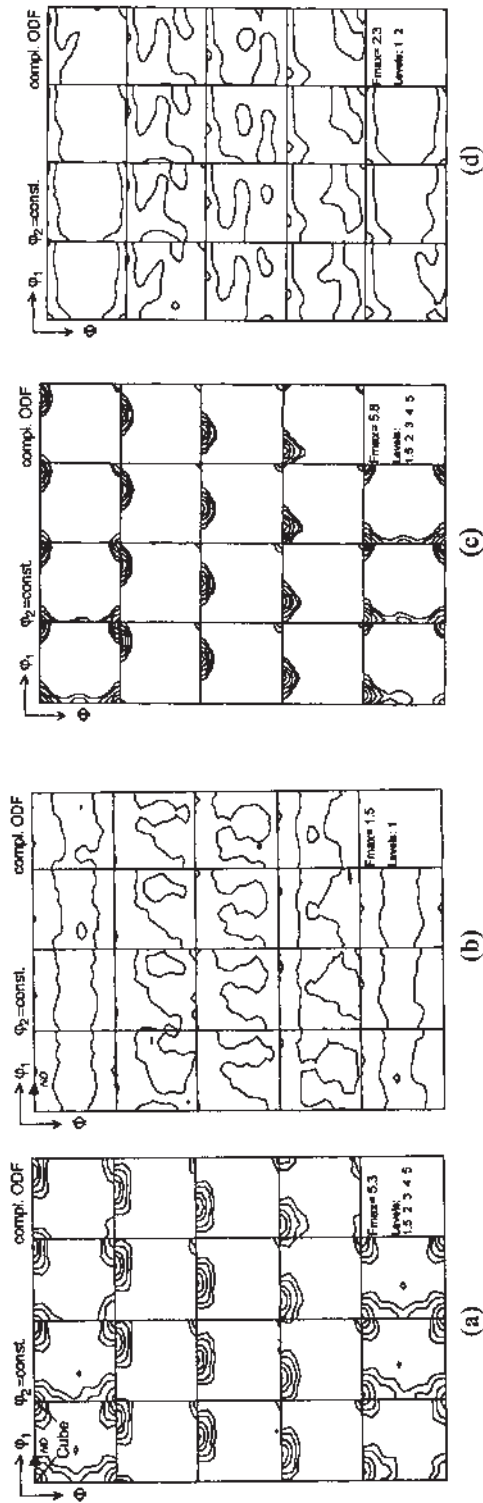
Figure 9 Experimental recrystallization texture (a and b) and model predictions (c and d) in AA3004 alloy. (a) and (c) $Z = 5.3 \times 10^{12}\,\text{sec}^{-1}$; (b) and (d) $Z = 3.9 \times 10^{16}\,\text{sec}^{-1}$. Recrystallization annealed at 430°C for 30 sec. (From Ref. [86], reprinted with permission from TMS.)

from the operating slip systems, which are related to the deformation mode and texture. The absolute maximum stress direction is shown to be parallel to the Burgers vector or the slip direction, and its magnitude is proportional to the total strain on a given slip plane. In the presence of multiple slip, the contribution from each slip plane must be calculated, and a resultant direction of maximum internal stress must be calculated. The theory has been successfully used to predict the evolution of recrystallization textures in axi-symmetrically deformed aluminum where the deformation texture based on $\langle 1\,1\,1 \rangle$ fiber changes to a $\langle 1\,0\,0 \rangle$ fiber texture upon recrystallization, the formation of rotated cube recrystallization texture from a Copper deformation texture in aluminum obtained by channel-die compression of an initially Goss oriented single crystal, and the formation of cube recrystallization texture from plane strain rolled polycrystalline aluminum. While OG and ON theories would require the presence of S–Cube interface to form the Cube texture during recrystallization, the SERM theory postulates that Cube will form at the expense of Copper. Experimental evidence does indicate that Copper is consumed during recrystallization of AA8011 aluminum alloy.

There is experimental evidence that the recrystallization texture in aluminum does not always have a $40°\ \langle 1\,1\,1 \rangle$ rotation with the deformation texture. Blicharski et al. [88] studied the deformation and recrystallization textures of aluminum bicrystals containing two variants of the S deformation component. The bicrystals were deformed in channel-die to 90% and annealed at $185°C$ for 5 min. Comparison of the deformation and recrystallization textures indicated that there was no such orientation relationship between the deformation and recrystallization textures. Therefore, the results could not be explained on the basis of Vatne and Engler's theories described above. However, the results could be explained on the basis of the SERM model. Rajmohan and Szupunar [89] also used the concept of maximum strain energy release as a criterion for nucleation in modeling the evolution of recrystallization texture in aluminum alloy AA3104. Both Blicharski et al. [88] and Rajmohan and Szupunar [89] concluded that either the ON or the OG acting alone was inadequate for explaining the evolution of recrystallization textures and believed that both OG and ON are equally important in texture evolution.

2.2 Mesoscale Simulations

By definition, these techniques apply to length scales comprising a collection of microstructural features, such as grain boundaries, coarse particles, and other second phases. The techniques that are widely used in recrystallization simulations include the ones based on Monte Carlo (MC) and cellular automaton (CA). There are several other mesoscale models, such as vertex simulations, boundary dynamics models, vertex dynamics models, and coupled vertex and boundary dynamics models. These models are still not well developed to handle complex three-dimensional deformation substructures, and therefore will not be described in any great detail. Simulations of isotropic grain growth using the phase field (PF) technique are now being extended to simulate anisotropic growth, and therefore may become increasingly important in simulating recrystallization. A brief description of these simulation techniques and their applications to recrystallization in aluminum will be presented.

2.2.1 MC Simulations

In this technique, the microstructure is discretized and mapped to a regular grid of points. Simulations have been carried out using square or hexagonal grids in two-dimensional as well as fcc, hcp, and simple cubic grids in three dimensions. Each lattice site is associated with a crystallographic orientation. The grain/subgrain boundaries are not modeled explicitly. However, a boundary is assumed to be present whenever the crystallographic orientations of two adjacent points are different. The simulations were initially developed for modeling microstructural evolution during isotropic [90] and anisotropic [91] grain growth in materials. The energy of a site and its neighborhood is given by

$$E = -J\left[\sum_{j=1}^{nn} \delta_{S_i S_j} - 1\right] \tag{34}$$

where J is the grain boundary energy, δ_{ij} is the Kronecker δ function, S_i is the orientation of the lattice site visited, and S_j is the orientation of the nearest-neighbor site. The summation is taken over all the nearest neighbors of the site. The change in the energy of the site before and after the attempted flip, ΔE, is calculated, and the flip is carried out with a probability p given by

$$p = \begin{cases} 1 & \Delta E \leq 0 \\ \exp\left(\frac{-\Delta E}{kT}\right) & \Delta E > 0 \end{cases} \tag{35}$$

The simulation proceeds by randomly visiting each lattice site and calculating the energy of the site and its neighborhood before and after an attempted flip of the orientation of the site to another orientation that is randomly chosen from the set of orientations initially present in the system. Recently, the above switching probabilities have been modified in order to take into account the effect of variable driving force and boundary mobility as follows [92]:

$$p = \begin{cases} \frac{J(S_i,S_j)}{J_{\max}} \frac{M(S_i,S_j)}{M_{\max}} & \Delta E \leq 0 \\ \frac{J(S_i,S_j)}{J_{\max}} \frac{M(S_i,S_j)}{M_{\max}} \exp\left(\frac{-\Delta E}{kT}\right) & \Delta E > 0 \end{cases} \tag{36}$$

The grain growth simulations were later extended to recrystallization simulations where, in addition to the grain boundary energy, the neighboring sites may also have a stored energy of deformation. The energy of the site and its neighborhood is defined as [93,94]

$$E = -J\left[\sum_{j=1}^{nn} \delta_{S_i S_j} - 1\right] + \sum_{j=1}^{nn} H_j \tag{37}$$

where H is the stored energy of deformation.

In recrystallization simulations, different types of nucleation scenarios, such as homogeneous nucleation [93] and heterogeneous nucleation [94] under site-saturated or constant nucleation rate conditions, were considered, and their influence on the recrystallization kinetics was examined. The simulation has also been used to account for the effect of inhomogeneous stored energy distribution on the recrystallization

kinetics [95]. The simulations were able to capture the influence of nucleation type on the JMAK exponent and also the deviation from ideal JMAK kinetics due to non-uniform stored energy of deformation.

A drawback of the MC technique is that both the simulation time and the physical size of the simulation domain are expressed as dimensionless quantities. It is not straightforward to relate these to the corresponding real quantities. Another parameter that is not physical is the lattice temperature, which is not the real temperature, but in mesoscale simulations is effectively a mathematical perturbation required to prevent premature locking of the microstructure depending upon the lattice type used. In systems with anisotropic grain boundary properties, it is necessary to locally scale the lattice temperature in order to avoid undesirable roughening of the grain boundaries that might introduce a mesoscale artifact in the boundary mobility [96].

2.2.2 Cellular Automaton Simulations

A cellular automaton is an array of contiguous cells, which may be arranged in one, two, or three dimensions. For the purposes of simulating microstructural evolution, two- or three-dimensional automata are used. The cells are characterized by certain attributes, such as temperature, material state (solid, liquid, deformed, recrystallized), orientation, etc. Rules are prescribed for changing the attributes of a cell based on the "transformation" to be modeled. In the case of recrystallization, rules are prescribed for changing the state from deformed to recrystallized. The neighborhood of a cell is also defined based on the first nearest, second nearest, or third nearest neighbors depending upon the problem to be simulated. In the case of recrystallization simulation, rules are also provided to initiate nucleation, i.e., convert one of the deformed cells into a nucleus. The switching rules are then applied to transform a deformed cell into a recrystallized cell. In the case of recrystallization, the interface velocity is directly proportional to the driving force:

$$V = M\Delta E \tag{38}$$

where ΔE is the driving force due to stored energy. The cellular automaton is evolved as a function of time based on Eq. (38). Davies [97] used different velocity formulations (a constant velocity formulation based on an analytical approach and a variable velocity formulation based on the Cahn–Hagel relationship), incorporated them in 3-D CA simulations of recrystallization, and compared the results with experimental recrystallization kinetics. It was found that the agreement between simulation and experiment was different for the two velocity formulations. The theoretical formulation based on constant velocity provided a better fit to the experimental recrystallization data at lower volume fractions, while the use of velocity derived using the Cahn–Hagel relationship provided a good fit at high volume fractions.

Recently, Davies et al. applied the CA technique to simulate static recrystallization in cold-rolled AA1050 [98]. Through EBSP measurements of texture components that evolve during recrystallization, the grains were classified into cube, deformation texture, and random classes, and the nucleation and growth parameters for each class were extracted from experimental data by fitting the data to the JMAK and the Cahn–Hagel relationships. The overall simulation kinetics was found to be

in good agreement with experimental data. However, at low times, the simulation kinetics showed significant deviation from the experimental data, which was attributed primarily to errors in determining the nucleation and growth parameters for each texture class, although the possibility of errors inherent to the simulation was not discounted. Similarly, for each texture class, the mean grain size and volume fraction determined from the simulation were in good agreement with the experimentally determined values. When the data from the three texture classes were combined, the mean grain size of the simulation was comparable to the experimentally determined mean grain size. The experimental and simulated grain size distributions for each texture class are shown in Fig. 10. The simulated rolling and random classes had a form similar to their experimentally determined counterparts, as did the distribution of all texture classes combined. Thus, the CA technique was shown to be a valid technique for simulating the evolution of microstructure. The size distribution of the cube class showed anomalies at low and high grain sizes. Both anomalies were connected to the fact that the cube class of grains was the fastest-growing class.

It should be noted that the above CA simulations require extensive experimental data, and in fact, the motivation for doing the simulations was to
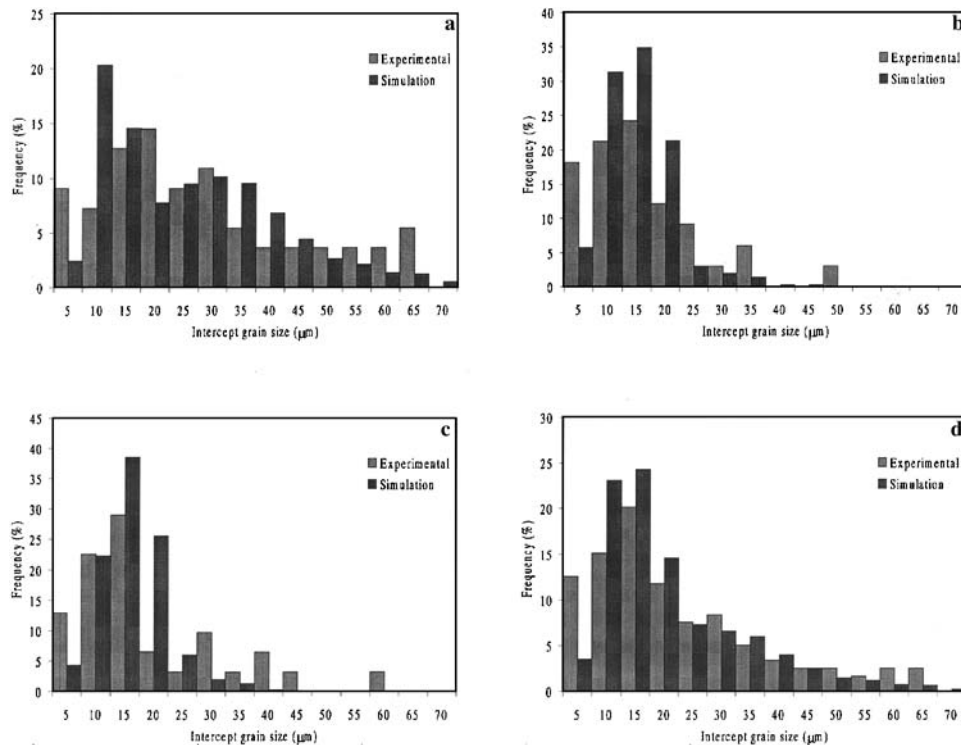
Figure 10 Comparison of simulated and experimental grain size distribution for each texture class arising from recrystallization in AA1050 (a) cube, (b) ''random'', (c) rolling, and (d) all three combined. (From Ref. [98], reprinted with permission from Elsevier Science.)

determine if CA could be used as a viable technique for simulating recrystallization. The simulations do not have a good model for nucleation. The nuclei were randomly distributed in the microstructure in the simulations, while in a real material the nucleus distribution could be non-random, which has a significant influence on the recrystallization kinetics [78].

In a complicated deformation substructure, it is possible that both ΔE and M vary with position because of the heterogeneous nature of the deformation substructure and variation of misorientation across the boundary. Recently, Raabe [99] developed a probabilistic CA model where the switching is calculated in a probabilistic manner by first computing a switching probability for each site as

$$p = \frac{M^{\text{local}} \Delta E^{\text{local}}}{M^{\text{max}} \Delta E^{\text{max}}} \tag{39}$$

and then effecting the switch with the probability, p.

While the MC technique can effectively capture the linear relationship between velocity and driving force when the driving force is curvature driven (grain growth), it cannot capture such a relationship when the driving force is stored energy. This is inherent in the switching probability shown in Eqs. (35) and (36). For example, when $\Delta E = 0$, the simulation would allow a finite probability for switching. This is critical, because kinks in the boundary are required to exist and execute random walks along the boundary to capture curvature-driven growth. However, when the boundary migration is due to differences in the stored energy of deformation, then the boundary velocity should be zero when $E = 0$, which the CA technique can capture correctly. However, the CA technique cannot be used to accurately capture the curvature-driven migration of boundaries. Rollett and Raabe [100] recently addressed this problem by developing a hybrid model that combines the site-switching probabilities of the Monte Carlo and cellular automaton techniques. In the hybrid model, either the MC switching rule or the CA switching rule is used based on a ratio, $\Gamma/(\Gamma+1)$, where Γ depends on the relative magnitude of the grain boundary energy to the stored energy.

2.2.3 Phase Field Simulations

In the phase field simulations, an arbitrary single-phase polycrystalline microstructure is described by many non-conserved order parameter fields $\eta_1, \eta_2, \ldots, \eta_p$, with each of them representing a specific crystallographic orientation. In contrast to the sharp interface approach, where the microstructure is described by mathematically sharp interfaces of zero thickness, the phase field methods characterize an arbitrary polycrystalline microstructure using the order parameters, which change continuously across grain boundaries. For this reason, the phase field models are often referred to as diffuse-interface models. The spatiotemporal evolution of the order parameters that describe the microstructural evolution during grain growth is given by Ginzburg–Landau-type kinetic equations [101]:

$$\frac{\partial \eta_i}{\partial t} = -L \frac{\delta F}{\delta \eta_i} \tag{40}$$

where L is the kinetic coefficient that characterizes the grain boundary mobility and F is a free energy functional that can be written in the following form:

$$F = F_0 + \int d^3r \left[f_0(\eta_1, \eta_2, \ldots, \eta_p) + \frac{k}{2} \sum_{i=1}^{p} |\nabla \eta_i|^2 \right] \tag{41}$$

where $f_0(\eta_i)$ is the free energy density and k is the gradient energy coefficient, which together determine the width and energy of the grain boundary regions.

The exact form of the free energy density function $f_0(\eta_i)$ is not important as long as it provides the correct topology for the free energy functional, which contains a large number of energetically degenerate potential wells at the points in the $\{\eta_i\}$ space located at $(\eta_1, \eta_2, \ldots, \eta_p) = (\pm 1, \ldots, 0), \ldots, (0, \ldots, \pm 1)$, where it is assumed that $\eta_i = 1$ inside the ith grain type (grain with a particular spatial orientation) and zero otherwise. A simple function that satisfies these requirements can be written in the form

$$f = \sum_{i=1}^{p} \left(-\frac{a_1}{2} \eta_i^2 + \frac{a_2}{2} \eta_i^4 + a_3 \sum_{j \neq i} \eta_i^2 \eta_j^2 \right) \tag{42}$$

where a_1, a_2, a_3 are positive constants. Combining Eqs. (40)–(42), we have

$$\frac{\partial \eta_i}{\partial t} = -L \left(-a_1 \eta_i + a_2 \eta_i^3 + 2a_3 \eta_i \sum_{j \neq i}^{p} \eta_j^2 - k_i \nabla^2 \eta_i \right) \tag{43}$$

The above equation is solved over the computational domain consisting of a square grid of points using a finite difference technique. For constant L and k, the above equations describe subgrain growth in isotropic systems. Anisotropic boundary energy and mobility as a function of boundary misorientation have recently been introduced in PF simulations by making L and k functions of the boundary misorientation [102], and the methodology has been used to simulate abnormal subgrain growth [103]. The simulations can ultimately be used to model recrystallization in high–stacking fault energy materials such as aluminum in which recrystallization occurs essentially through an abnormal subgrain growth process.

2.2.4 Network Models

In this class of models, an initial microstructure is constructed that has the positions of all the boundaries and the vertices. These are mainly two-dimensional models, although in some cases three-dimensional equivalents have been developed. At every time step, the equations of motion of the boundary points and/or the vertices are used to determine the new coordinates. Predetermined grain switching criteria are used when two vertices meet each other due to grain shrinking. In the boundary dynamics models, the displacements of the boundaries are calculated from their local curvatures and the positions of the vertices are adjusted to fulfill equilibrium requirements [104]. In the vertex dynamics models, the grain structure is essentially represented by the positions of the vertices or the triple points. The underlying assumption is that the boundary mobility is much higher than the triple point mobility, so that triple point motion essentially controls the structure evolution [105,106]. The model has been used to simulate the formation of recrystallized nuclei by the strain-induced boundary migration (SIBM) mechanism [107]. In the coupled boundary–vertex dynamics

models, the assumption of low vertex mobility is removed, and it is possible to consider circumstances where the vertex mobility is much higher than the boundary mobility. The simulation methodology involves a modified Kawasaki model in which the motion of the triple points along with ''virtual vertices'' situated on boundary segments is considered. Application of this technique for normal grain growth gives the striking result that the triple point mobility is not rate-controlling and that local equilibrium is indeed maintained at the triple junctions.

Although the simulation techniques described above can be used to capture the mesoscale effects on recrystallization, such as the effect of non-random nucleus distribution, anisotropic growth, variable stored energy, concurrent precipitation during recrystallization, concurrent recovery, etc., that cause significant deviations from recrystallization kinetics, they have been used mainly to illustrate such effects in hypothetical systems and not in any real aluminum alloy. These models could be used more effectively in developing a detailed microstructural understanding of recrystallization in commercial aluminum alloys if the deformation substructures could be properly input into the mesoscale simulations. With the advent of the EBSP, it is now routinely possible to take microtexture measurements and feed the data into the mesoscale models, but it is only recently that the appropriate data have started to become available in the form of, for example, orientation-dependent grain boundary mobilities [108–110]. However, the generation of 3-D microstructural data is more difficult, since it requires generation of microtexture and microstructure information from several sections and the utilization of computer codes that will use the serial-section data to generate the three-dimensional microstructure. Alternatively, 3-D x-ray microscopic techniques are being developed that use high-energy synchrotron radiation to penetrate large depths in aluminum and obtain 3-D orientation information with a lateral resolution of 0.1 μm. Early work [111], with a minimum observable grain size of around 1 μm, has shown that in situ synchrotron investigations can be conducted; however, data analysis is time consuming.

2.2.5 Coupled Deformation–Recrystallization Models

As described previously, one of the biggest hurdles in our understanding of microstructure and texture evolution during recrystallization in aluminum alloys is the lack of understanding of the deformation microstructure itself. While the analytical models of deformation do provide the mean values for the cell/subgrain size, internal dislocation density, and subgrain misorientation, they do not provide a quantitative estimate of the variations in these quantities at the mesoscopic length scale. Also, it is not possible to know the spatial distribution of the above quantities, which is important in determining the number, orientation, and spatial distribution of the nuclei that form during subsequent recrystallization. Recently, the microstructural deformation models based on crystal plasticity have been coupled with mesoscale simulation techniques, such as MC and CA, to understand the influence of deformation microstructure on the subsequent microstructural and microtextural evolution during recrystallization. Radhakrishnan et al. [78] developed a nucleation criterion based on the product of the stored energy of deformation at a given site and the mean misorientation of the site with its neighboring sites. This is based on the fact that a site with a large amount of stored energy would recover quickly by heterogeneous subgrain growth, which would be also aided by local
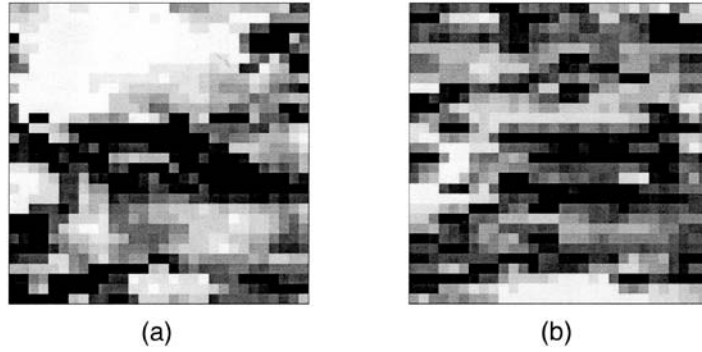
Figure 11 Distribution of nucleation parameter after (a) 50% deformation and (b) 68% deformation of polycrystalline aluminum. (From Ref. [78], reprinted with permission from Elsevier Science.)

orientation gradients to form supercritical cells bounded by high-angle grain boundaries that constitute the recrystallization nuclei. The MC recrystallization simulation took into account the variation of boundary mobility with misorientation. The simulations were carried out for plane strain cold deformation and subsequent static recrystallization of commercial purity aluminum. Figure 11 shows the non-random distribution of the nucleation parameter after 50% deformation that becomes more random after a reduction of 68%. As shown in Fig. 12, when the nuclei form in a non-random fashion, impingement of nuclei occurs fairly early during growth for a prior deformation of 50%, while a uniform three-dimensional growth occurs when the nuclei are more randomly distributed, as in the case of a 68% prior deformation.

The above analysis has been subsequently modified [112] based on the fact that recrystallization in high-stacking fault energy metals such as aluminum is essentially a discontinuous subgrain growth process driven by local heterogeneities in the stored energy of deformation. These heterogeneities result in local variation of the mean subgrain size or subgrain misorientation. Instead of a phenomenological description of the nucleation parameter described above, nucleation by heterogeneous subgrain growth could be made to occur naturally from the deformation structure if the deformation substructure consisting of subgrains could be extracted from the microstructural deformation model. In this case, recrystallization would be modeled as a one-step phenomenon of heterogeneous growth rather than the conventional models that treat recrystallization as a two-step process of nucleation and growth. The deformation substructure is extracted by assuming that the stored energy of deformation in the element per unit volume is equal to the surface energy per unit volume of the subgrain boundaries in the volume element based on the equation

$$H = \frac{\gamma_0}{2d}\left[\frac{\Delta\omega}{\omega^*}\left(1 - \ln\frac{\Delta\omega}{\omega^*}\right)\right] \tag{44}$$

where γ_0 is the specific energy of a high-angle boundary, $\Delta\omega$ is the mean misorientation among subgrains, ω^* is the misorientation limit for low-angle boundaries taken to be $15°$, and d is the mean subgrain size. The deformation

MCS = 6   MCS = 9   MCS = 7   MCS = 10

MCS = 19   MCS = 100   MCS = 13   MCS = 100

50% deformation   **68% deformation**
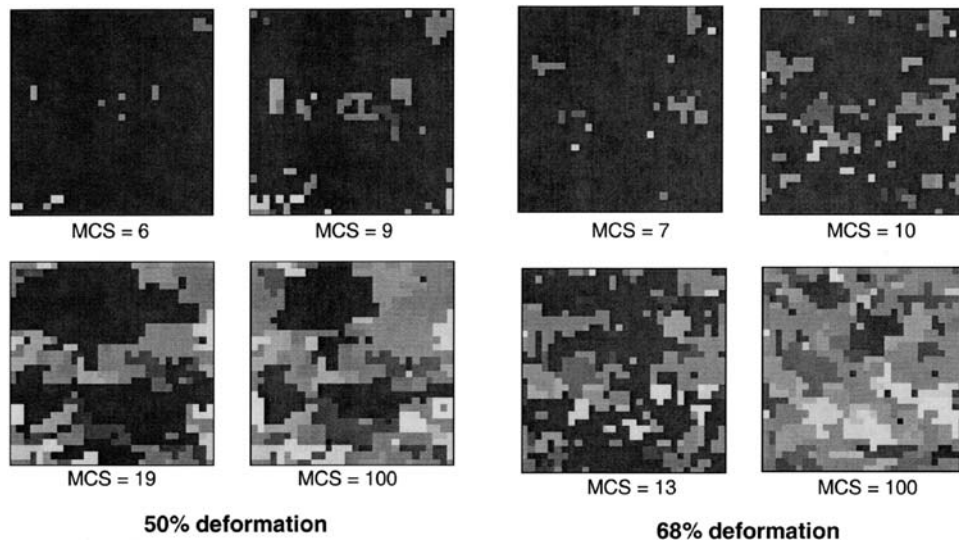
Figure 12 Temporal evolution of recrystallization in polycrystalline aluminum obtained using coupled finite element–Monte Carlo simulation approach. Non-random nucleation leads to early impingement of growing nuclei for prior deformation of 50% (left), while for prior deformation of 68%, the nuclei are more randomly distributed and impingement is significantly delayed during growth (right). (From Ref. [78], reprinted with permission from Elsevier Science.)

substructure in each volume element is assumed to consist of a set of subgrains, with a mean subgrain size of d and a mean misorientation of $\Delta\omega$ scattered about the mean orientation ω of the element. The substructure is generated by applying Eq. (44) to each volume element, using the stored energy and orientation information obtained from the deformation model, thus capturing the heterogeneities present in the deformation substructure. The evolution of the deformation substructure is simulated using an MC technique for subgrain growth, where the boundary energy and the boundary mobility are functions of the boundary misorientation. The variation of boundary energy with misorientation is given by the Read–Schockly relationship [64]:

$$
\gamma = \begin{cases}
0 & \omega = 0 \\
\gamma_0 \frac{\omega}{\omega^*}\left\{1 - \ln\left(\frac{\omega}{\omega^*}\right)\right\} & 0 < \omega \le \omega^* \\
\gamma_0 & \omega > \omega^*
\end{cases}
\tag{45}
$$

and the variation of boundary mobility with misorientation is given by the following sigmoidal relationship based on experimental observations:

$$
M = M_{\text{HAGB}}[1 - \exp(-\lambda\omega^n)]
\tag{46}
$$

where M_{HAGB} is the relative mobility of a high-angle boundary, and λ and n are constants.

The simulation captures the formation of nuclei at prior grain boundaries and triple lines by the SIBM mechanism and the discontinuous growth of the nuclei to

consume the rest of the deformation substructure. The advantage of the technique is that there is no need to make any phenomenological assumption about the formation of recrystallization nuclei. Also, since recrystallization is simulated as a one-step process of discontinuous subgrain growth, both discontinuous growth (discontinuous recrystallization) and normal growth occurring simultaneously can be modeled easily. Thus, there is no need to have a hybrid model like the one described previously for simultaneously evolving stored-energy-driven growth and curvature-driven growth. The temporal evolution of recrystallization using the above simulation technique for polycrystalline aluminum is shown in Fig. 13.

Raabe and Becker [113] used a similar coupled approach for modeling static recrystallization by combining the crystal plasticity microstructural deformation model with a probabilistic cellular automaton technique for simulating static recrystallization. The stored energy and the orientation information from the finite element mesh were mapped to a finer CA grid. Nucleation was handled using two types of phenomenological models. In the first, the misorientation between a cell and its neighbor is checked. If the misorientation is greater than $15°$ and if the stored energy in the cells is greater than a prescribed value, the two cells were assumed to be
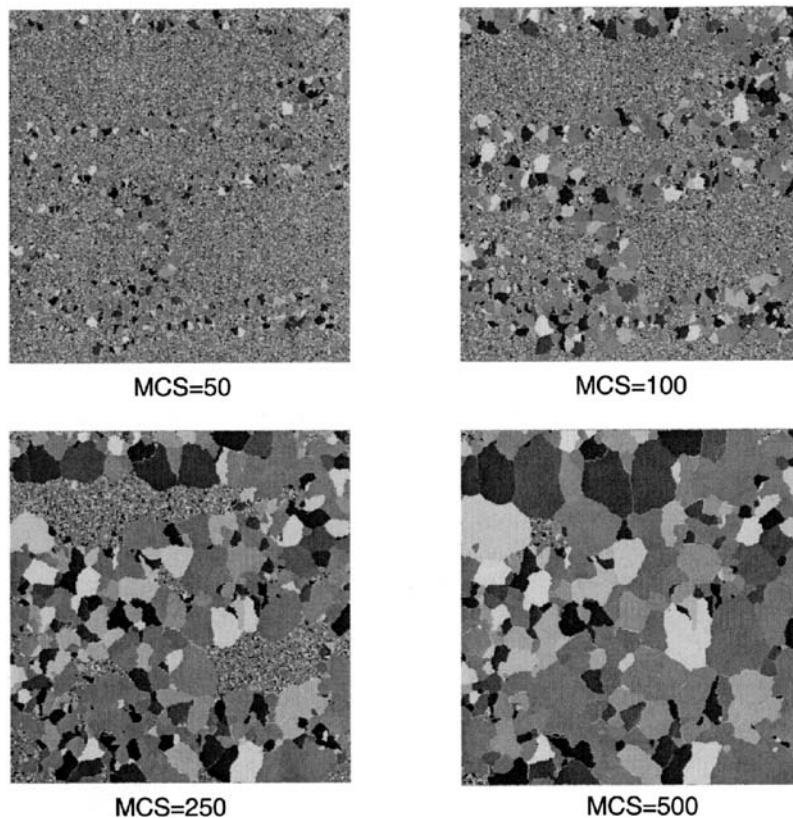
Figure 13 Monte Carlo simulation of temporal evolution of deformation substructure extracted from crystal plasticity finite element modeling of microstructural deformation in polycrystalline aluminum.

nuclei whose orientation is midway between the orientations of the two cells. In the second approach, a single cell with high stored energy and high misorientation with the surroundings was selected as the nucleus without any orientation change. Growth was accomplished using the CA technique, where the cells were switched from a deformed state to a recrystallized state using a scaled and normalized switching probability given by Eq. (39). The simulations, like the previously described ones, were also able to capture the inhomogeneous distribution of nuclei and the deviation of recrystallization kinetics from ideal JMAK behavior. The temporal evolution of the recrystallized microstructure using the above simulation technique is shown in Fig. 14. The upper figures in Fig. 14 show the orientation image, i.e., the microtexture of the material. The lower figures show the dislocation density. The recrystallized regions correspond to a dislocation density of zero and appear white. Preferential nucleation at certain grain boundary sites and growth of these nuclei to consume the rest of the deformed microstructure are clearly seen.

While the coupled deformation–recrystallization models described above are quite powerful in that they are able to explicitly model the effect of initial microstructure and the deformation conditions on subsequent microstructure evolution during recrystallization, they can only be applied to small volume elements. This is because the computations are time-consuming, and they cannot be carried out rigorously for the length scale of the entire workpiece. However, the analysis could be carried out in a decoupled way to model the influence of deformation path on local texture evolution. For example, the deformation of volume elements close to the surface of the sheet being rolled would have a larger shear component than those in the mid-thickness region. A larger length scale FE model based on microstructure-free constitutive behavior can be used to model the deformation history of these volume elements, which can then be used in the microstructural deformation model.

3 PRECIPITATION

Although the precipitation response of aluminum alloys has been exploited successfully for many years without recourse to modeling, there are several reasons why one may want to predict the precipitate evolution in an aluminum alloy exposed to a thermal cycle. Principal among these is the desire to optimize either the alloy composition or its processing in order to achieve improvements to some mechanical property (usually strength, sometimes toughness, and often a combination of the two). A secondary, and related, consideration is the need to be able to predict the precipitation response during processes that are not as "well behaved" as isothermal aging, such as welding or thermomechanical processing, where dissolution may also play a role. Because of this, both formation and dissolution are dealt with in this section, and even though they are not necessarily complementary processes, from the modeling point of view they are treated as such.

Probably because strengthening is the motivation for the artificial aging of aluminum alloys, models can be classified as those which model the strengthening response only (i.e., without considering the underlying microstructural evolution), and those which model the microstructural evolution and then consider the contribution to strengthening. This section groups models by length scale as analytical models, mesoscale models, and atomistic models. Independent of length scale,
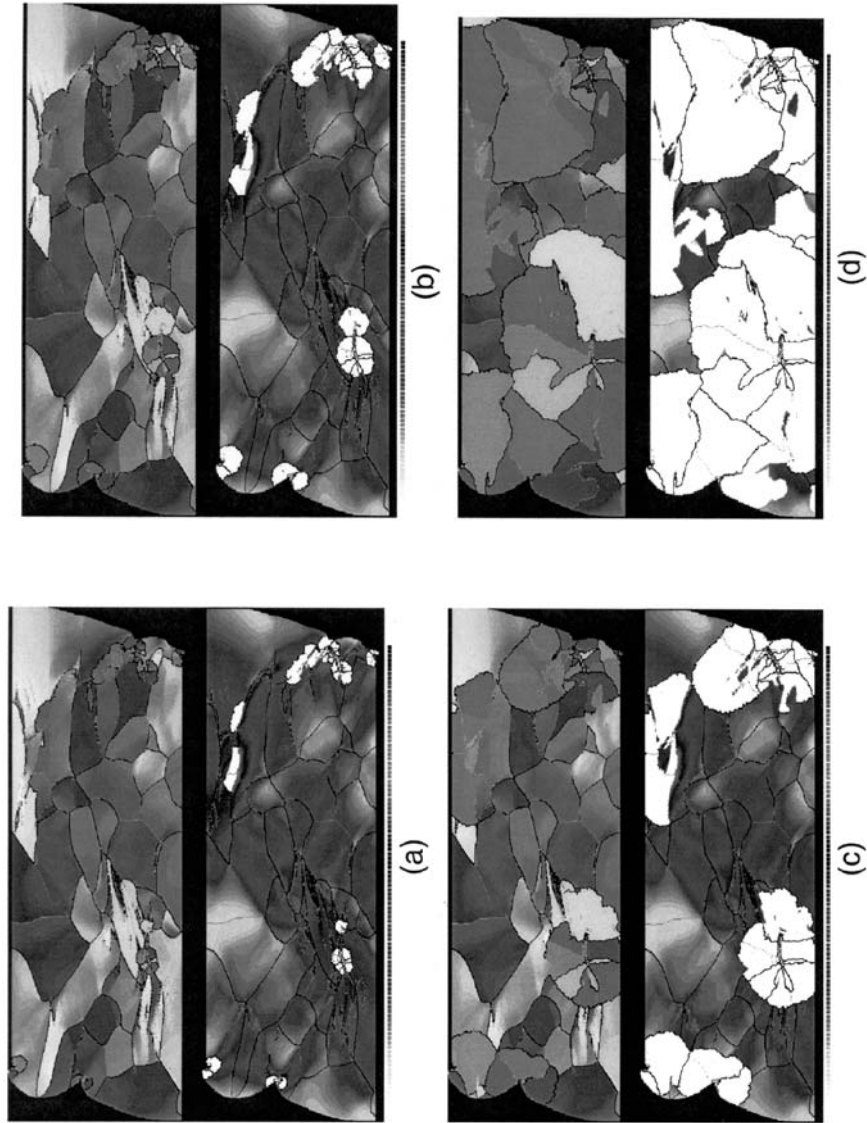
Figure 14 Deformation structure obtained by crystal plasticity—based finite element model (upper figures) and temporal evolution of recrystallization (lower figures) obtained using cellular automaton technique: (a) 3%; (b) 13.6%; (c) 28.8%; (d) 68.7%. (From Ref. [113], courtesy Raabe.)

"complete" models must address the nucleation, growth, and stability of the precipitating phase. Conventional precipitation hardening heat treatment is a relatively simple process in that it is carefully controlled and is conducted under near-isothermal conditions, and by far, the bulk of the models deal with this situation. However, non-isothermal conditions are routinely encountered during and after thermomechanical processing, including during welding, and analytical process models have been developed to deal with these situations. Two-step aging treatments have also been addressed by similar methods. Finally, the influence of strain is important when aging treatments are conducted after forming operations. Some models have been developed to address these more complex circumstances.

Complications arise because (a) not all precipitates are spherical; (b) in general, the precipitation reaction proceeds by the nucleation and growth of a series of metastable phases that replace one another, although not sequentially, until the equilibrium phase is achieved. Commonly, however, the reaction is not allowed to proceed to completion. The 2xxx (Al-Cu), 6xxx (Al-Mg-Si), 7xxx (Al-Zn-Mg), and 8xxx (Al-Li) series alloys are all susceptible to precipitation hardening; however, modeling has focused on the 6xxx and 7xxx series alloys.

3.1 Analytical Models

Models of precipitation that deal with the bulk behavior of the metal are the most advanced in terms of their ability to predict the outcome of processing. Many analytical models adopt an internal state variable (ISV) approach, in which the evolution of microstructure is defined by a series of differential equations:

$$\frac{dS_1}{dt} = f_1(S_1, S_2, \ldots, T)$$
$$\frac{dS_2}{dt} = f_2(S_2, S_1, \ldots, T)$$

(47)

where the value of any state variable, S_i, at a time increment dt from the present depends only on the current state defined by the values of all variables S_i and the temperature. Precipitation models adopt as state variables the precipitate volume fraction and the average particle radius, and occasionally, some measure of the size distribution. Crucially, the state variable should not depend on the material history, and this means that we can define a "template" for the reaction for which the kinetics of the reaction collapse onto a single curve or line. This is known as the isokinetic approach, and a discussion of its applicability has been detailed by Grong and Shercliff [114]. Under some circumstances, this criterion can be disregarded with no adverse effect on the outcome of the model, but there also are instances when the ISV approach breaks down. One important example of this is the case of continuous nucleation of precipitate particles, and this will be illustrated later in the section.

Classically, the kinetics of the reaction are modeled by the JMAK relation, with the volume transformed depending on the number of precipitate particles, the availability of solute, and the distribution of precipitate sites. In its general form, the JMAK relation is

$$X = 1 - \exp(-X_{ext})$$

(48)

where X_{ext} is the volume fraction of precipitate particles transformed in the absence of any impingement, and is known as the extended volume fraction (cf. Eqs. (17) and (31) earlier). Impingement is taken to mean either "hard" impingement − physical contact between particles − or "soft" impingement − the overlapping of solute-depleted regions − as particles grow. At low volume fractions, V is equal to the observed volume fraction, but in a random distribution of precipitates, the observed and calculated volume fractions quickly diverge. It is rarely recognized that the JMAK relation is strictly a limit at large numbers of precipitate particles (N) of the more general relation for all N [115]:

$$X = 1 - \left[1 - \frac{X_{ext}}{N}\right]^N \tag{49}$$

Starink and coworkers [116−119] implement an alternative to the JMAK approach, based on a more general form of the governing differential equation [120,121]:

$$\frac{dX}{dX_{ext}} = (1 - X)^\lambda \tag{50}$$

A value of $=1$ yields the JMAK relation; $=2$ gives the Austin−Rickett formulation. The rationale for this alternative approach is the apparent failure of the JMAK relation for diffusion-controlled growth [122]. Starink conducts a simple thought experiment to show that the JMAK assumption of random nucleation cannot be maintained once solute depletion commences, because nucleation is dependent on solute concentration and thus there will exist some regions in which nucleation is not possible. He further shows that JMAK underestimates the impingement effect for a non-regular array. However, no similar reasoning is presented to support the alternative, and an independent evaluation would seem necessary to resolve this point of difference.

Upon integration, Eq. (50) yields

$$X = 1 - \left(\frac{X_{ext}}{\eta} + 1\right)^{-\eta} \tag{51}$$

where $\eta = 1/(\lambda - 1)$. For the condition $\lambda 1$ (i.e., large values of η), or for small values of the extended volume fraction, Eq. (51) tends to Eq. (48). As η in Eq. (51) and N in Eq. (49) increase, the two equations converge to the same limit.

The extended volume around a single particle is given by [122]

$$V_p = A[G(t - z)]^n \tag{52}$$

where A is a constant, G is the average growth rate, z is the time at which the nucleus formed, and n is the growth rate exponent. The volume of all particles transformed in a time interval is

$$dV_{ext} = V_0 I(T)A[G(t - z)]^n dz \tag{53}$$

where $I(T)$ is the nucleation rate per unit volume, and V_0 is the total sample volume. Replacing the term V_{ext}/V_0 by the extended volume fraction allows us to

model the transformation using Eq. (53). In practice, the extended volume fraction is modeled as

$$X_{\text{ext}} = [k(T)t]^n \tag{54}$$

where $k(T)$ is a factor that captures both the nucleation and growth parameters:

$$k(T) = k_0 \exp\left(-\frac{E_{\text{iso}}}{k_{\text{B}}T}\right) \tag{55}$$

E_{iso} is an effective activation energy for isothermal conditions, and k_0 is a constant. These parameters are calibrated using differential scanning calorimetry.

The model was successfully used to predict the calorimetry traces in an Al–6% Si alloy and β and β precipitation in Al–Mg alloys [117,123]. It is here that the internal state variable approach is shown to break down, at least for β precipitation. Starink and Zahra [117] show that the equation for the fraction transformed represented in ISV form as

$$\frac{\text{d}\alpha}{\text{d}t} = g(T)f(\alpha) \tag{56}$$

cannot be valid, because $f(\alpha)$ depends on temperature, violating the definition of an ISV differential equation. Consequently, the activation energy for the process cannot be evaluated in the conventional manner. In practice, though, the extended volume fraction can be considered to be a state variable under certain restrictive conditions, and Borrego and González-Doncel [124,125] model the precipitation of β in AA6061 using a JMAK formulation on the assumption that the nucleation is completed rapidly, i.e., that site-saturated conditions can be assumed (Fig. 15). In the revision of this model [125], the JMAK exponent, n, is allowed to vary, with a consequent improvement in the fit of model to experimental data (Figs. 16 and 17). Whereas it is unusual to vary the JMAK exponent, Cahn [126] shows that a decreasing n value can result from a non-random precipitate distribution.

More recently, Starink et al. [118,119] have coupled the transformation model to a coarsening model and applied it to the isothermal aging of the aluminum-lithium alloy 8090 and metal matrix composites based on 8090 (Fig. 18). The coupled model predicts the evolution of volume fraction and size of GPB zones, δ', and S' precipitate phases with good accuracy, and is further incorporated into a model for yield stress using the conventional expression

$$\sigma_y = M\Delta\tau_{\text{tot}} \tag{57}$$

where M is the Taylor factor, and $\Delta\tau_{\text{tot}}$ is the contribution of all strengthening mechanisms, which, in the 8090 alloy, are order hardening due to the δ precipitates, modulus hardening due to the GPB zones, and dispersion hardening due to S. For the underaged alloy, the order hardening contribution due to the δ phase is expressed as

$$\Delta\tau_{\text{ord,ua}} = \frac{\gamma_{\text{apb}}}{2b}\sqrt{\frac{3\pi^2\gamma_{\text{apb}}f\bar{r}}{32\Gamma}} \tag{58}$$
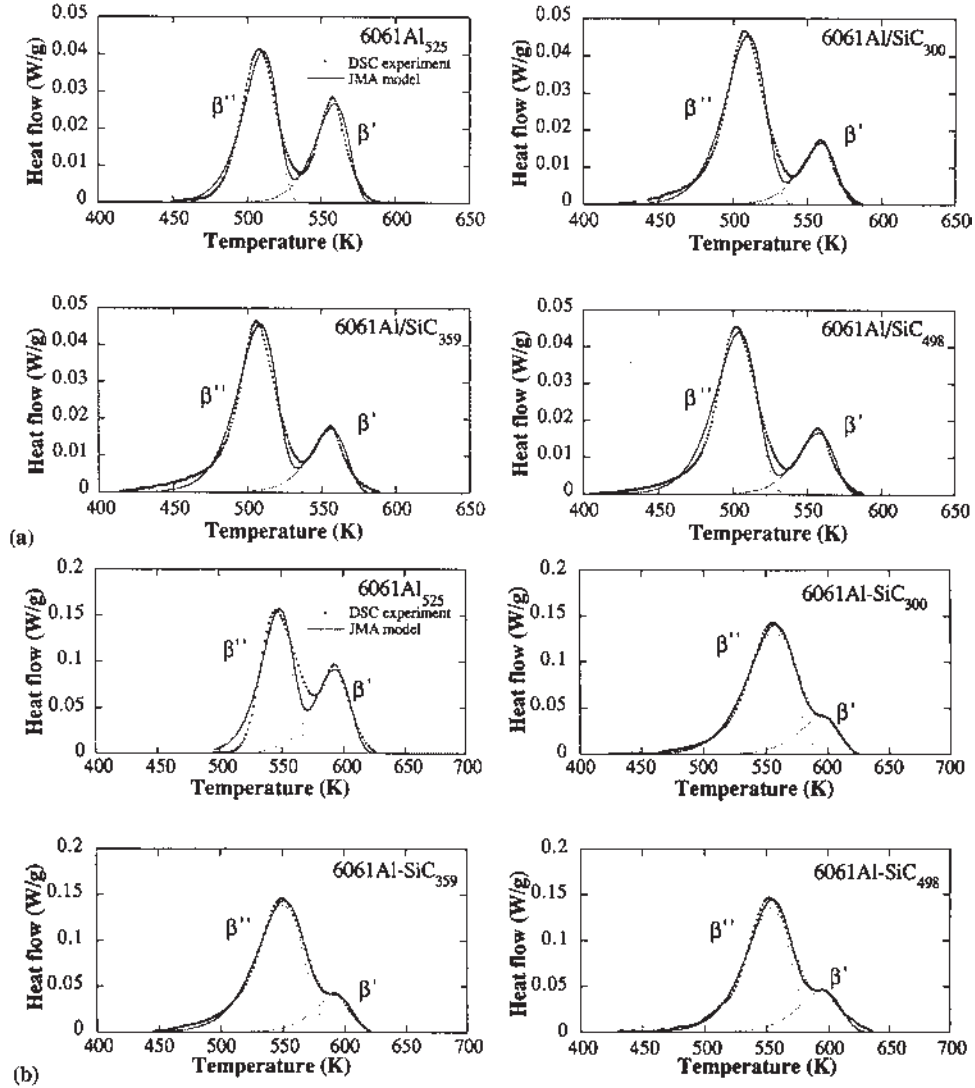
Figure 15 JMAK fits to DSC curves at (a) $10\,\text{K}\,\text{min}^{-1}$ and (b) $40\,\text{K}\,\text{min}^{-1}$ using a constant JMAK exponent. Dotted lines are approximations for β' and β'' precipitation, and subscripts indicate extrusion temperatures. (Reprinted from Ref. [125], copyright 1998, with permission from Elsevier Science.)

where γ_{apb} is the antiphase boundary energy on $\{1\,1\,1\}$ planes, f is the volume fraction of precipitates, $\bar{r}$ is the mean particle size, and Γ is the line tension. For a peak aged structure, the contribution is

$$\Delta\tau_{\text{ord,pa}} = 0.81\frac{\gamma_{\text{apb}}}{2b}\sqrt{\frac{3\pi f}{8}} \qquad (59)$$
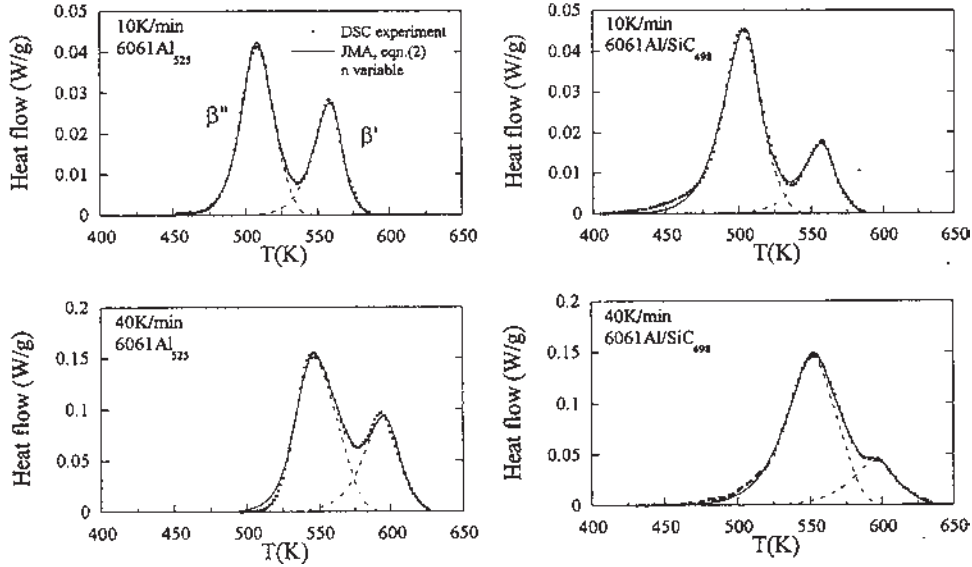
Figure 16 JMAK fits to DSC curves at 10 and $40\,\mathrm{K\,min}^{-1}$, using a variable JMAK exponent. (Reprinted from Ref. [125], copyright 1998, with permission from Elsevier Science.)
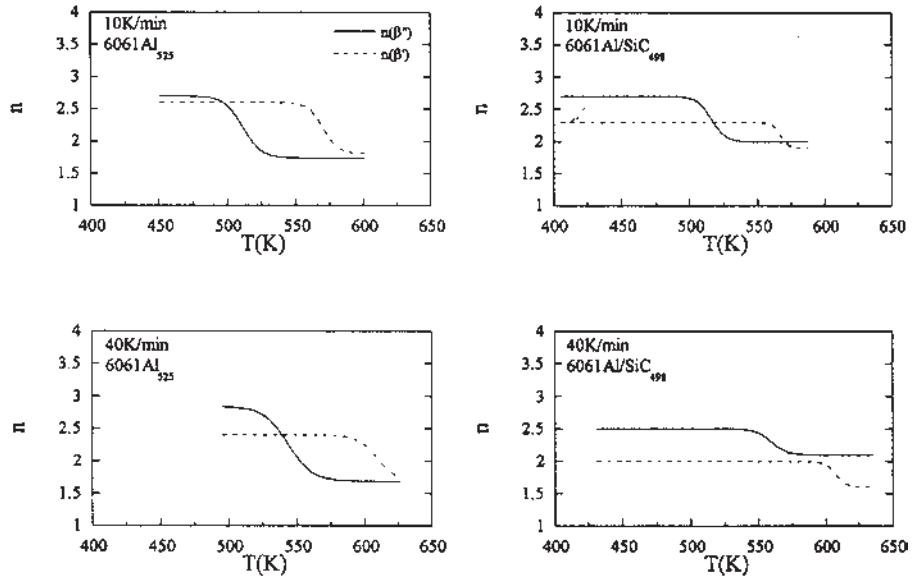
Figure 17 Variation of JMAK exponent, *n*, with temperature for different precipitate types under conditions shown. (Reprinted from Ref. [125], copyright 1998, with permission from Elsevier Science.)
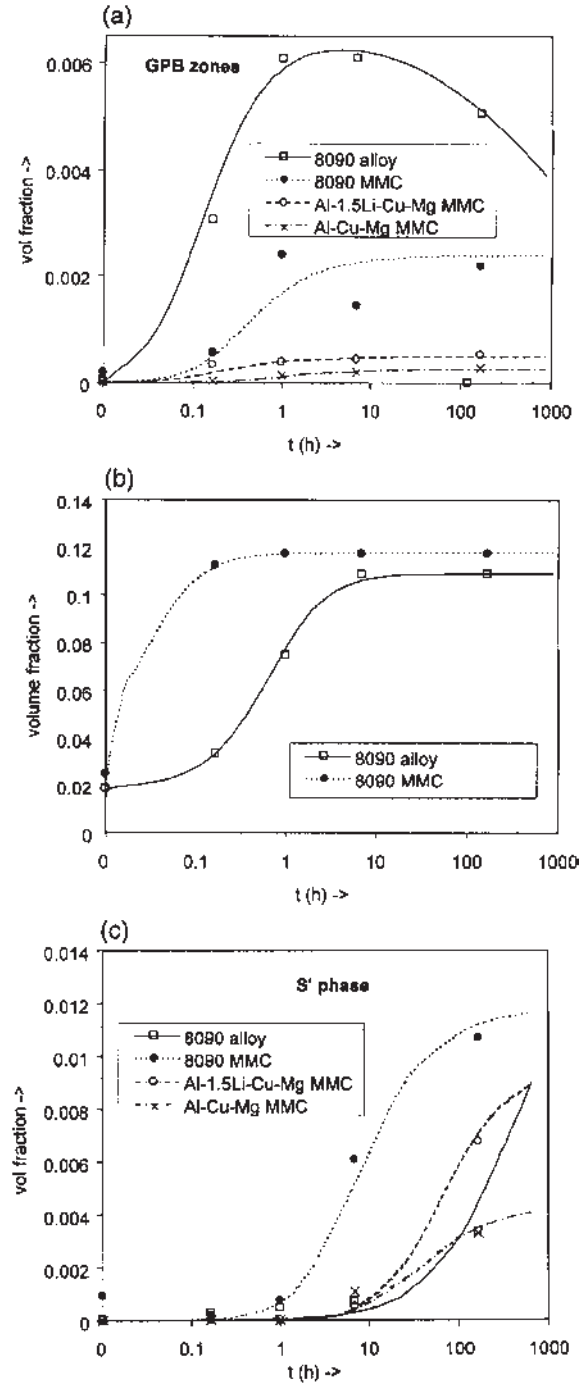
Figure 18 Predicted evolution of precipitating species in AA8090 and its composites: (a) GP zones, (b) δ, and (c) S. Lines indicate model, and points are data from DSC experiments. (Reprinted from Ref. [118], copyright 1999, with permission from Elsevier Science.

and the actual contribution to strengthening is determined from

$$\Delta\tau_{\mathrm{ord}} = \mathrm{Min}\{\Delta\tau_{\mathrm{ord,ua}}, \Delta\tau_{\mathrm{ord,pa}}\} \tag{60}$$

Modulus hardening in the GPB zones is modeled as

$$\Delta\tau_{\mathrm{mod}} = \frac{\Delta G}{4\pi\sqrt{2}}\sqrt{f} \tag{61}$$

where ΔG is the difference in shear modulus between the GPB zones and matrix, and the GPB zone modulus is estimated as a simple average of the shear moduli of the constituent magnesium and copper. Dispersion strengthening due to the S' phase is represented by

$$\Delta\tau_{\mathrm{dis}} = 0.81\frac{Gb}{2\pi(1-v)^{1/2}}\left(\frac{\ln(d/b)}{0.651d(2\pi/3f)^{1/2}-d}\right) \tag{62}$$

where d is the diameter of the rod-like S' precipitates. In the complete strengthening model, solute strengthening is also considered, and for the MMCs, additional strengthening due to misfit dislocations and weakening due to precipitate-free zones at ceramic particles are also treated.

Both the strength and weakness of this approach lie in the need to calibrate the model using experimental data that are then themselves "modeled". A partial solution to this problem is to use a set of reference data to which other, modeled, systems are compared.

Bratland et al. [127] model the kinetics for systems in which the nucleation and growth rates do not change during the transformation as

$$X = 1 - [1 - X_{\mathrm{c}}]^{(t/t^*)^n} \tag{63}$$

where t^* is the time taken to precipitate a certain fraction, X_{c}, of the second phase, and n is a "JMAK exponent". The value of t^* is determined from classical nucleation theory, and is found in practice by calibration to experimental data for an alloy of the same family at a set reference temperature:

$$t^* = t_{\mathrm{r}}^*\left(\frac{N_{\mathrm{v}}^*}{N_{\mathrm{v}}}\right)\exp\left[\frac{1}{R}\left(\frac{A_0(T_{\mathrm{eq}})^2}{T(T_{\mathrm{eq}}-T)^2} - \frac{A_0^*(T_{\mathrm{eq}}^*)^2}{T_{\mathrm{r}}(T_{\mathrm{eq}}^*-T_{\mathrm{r}})^2}\right) + \frac{Q_{\mathrm{d}}}{R}\left(\frac{1}{T}-\frac{1}{T_{\mathrm{r}}}\right)\right] \tag{64}$$

The subscript "r" refers to data at the reference temperature, and the * on the right-hand side of the equation distinguishes the known parameters for a reference *alloy*. A thermodynamic model is used to predict the solvus boundaries and the equilibrium dissolution temperatures (T_{eq}). The model is used to predict the kinetics of β Mg$_2$Si evolution for three Al–Mg–Si alloys based on one reference alloy (Fig. 19). However, the strength of this approach does not lie in the isothermal models, because, importantly, A_0 is a tuning parameter that is adjusted for each alloy.
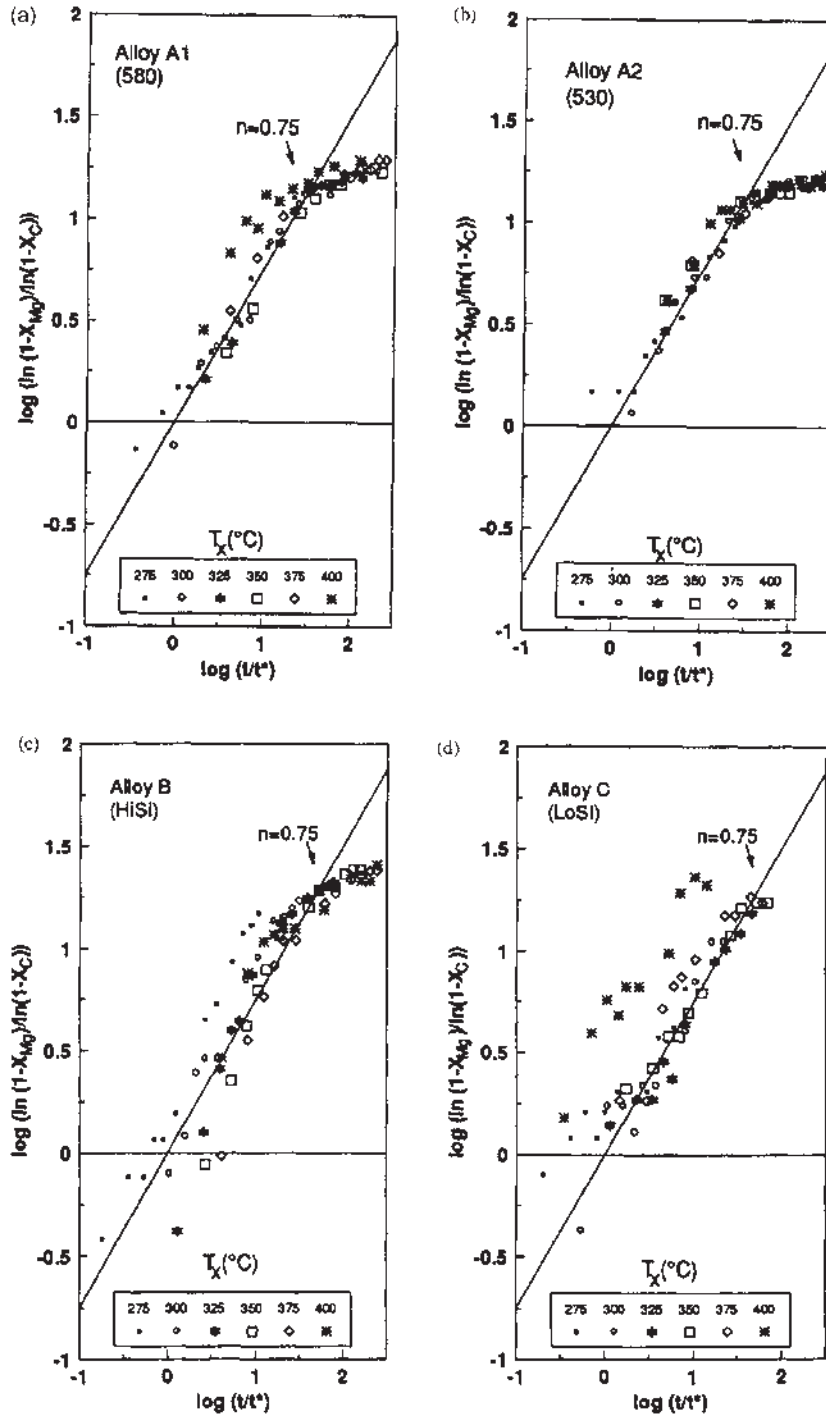
Figure 19 Kinetics of β precipitation for four different alloys. Note how curves deviate from constant *n* at long times. (Reprinted from Ref. [127], copyright 1997, with permission from Elsevier Science.)

Equation (63) is readily adapted to non-isothermal conditions based on the principles of additivity, whereupon the equation is rendered as

$$X = 1 - [1 - X_c]^{m} \tag{65}$$

and

$$I = \int_0^t \frac{\mathrm{d}t}{t^*} \tag{66}$$

This formulation is applied to predict behavior on continuous cooling [127] and consequently develop process diagrams for the extrusion of alloy AA6082 (Fig. 20). These diagrams predict the yield strength after extrusion and allow the precipitation model to be used as the basis not only for alloy design but also for the optimization of cooling schedules after extrusion − what the authors call "smart cooling systems". Essentially the same model is exploited by Bjorneklett et al. [128] for the *dissolution* of particles during continuous heating and cooling cycles, and welding.

Whereas mathematically correct analytical models can act as excellent signposts to guide industrial processing, we should not expect to accurately model an industrial process in a single differential equation. Even a small variation in temperature can lead to a large difference in the outcome of the model. In order to model more complex situations, analytical models are coupled with numerical methods that discretize the spatiotemporal phase field.

Myhr and Grong [129,130] and Myhr et al. [131] have extended the above approach to explicitly recognize coupled nucleation, growth, and coarsening, and particle size evolution is modeled by a finite difference discretization of the size distribution. Several others use a similar formulation to model precipitation in various Al–Zn–Mg alloys [132,133], Al–Zr alloys [134], and Al–Li alloys [135], each to good effect. All of these models owe something to the work of Wagner and Kampmann [136]; we shall, however, focus on the model developed by Myhr and Grong as an exemplar of the approach. The nucleation rate in the absence of any incubation period is defined by the standard equation

$$j = j_0 \exp\left(-\frac{\Delta G_{\text{het}}^*}{RT}\right) \exp\left(-\frac{Q_\text{d}}{RT}\right) \tag{67}$$

and

$$\Delta G_{\text{het}}^* = \frac{A_0^3}{(RT)^2 [\ln(C_\text{m}/C_\text{e})]^2} \tag{68}$$

which ignores the effect of elastic coherency strains. The rate of growth or dissolution is

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{C_\text{m} - C_\text{i}}{C_\text{p} - C_\text{i}} \frac{D}{r} \tag{69}$$

(a)

(b)
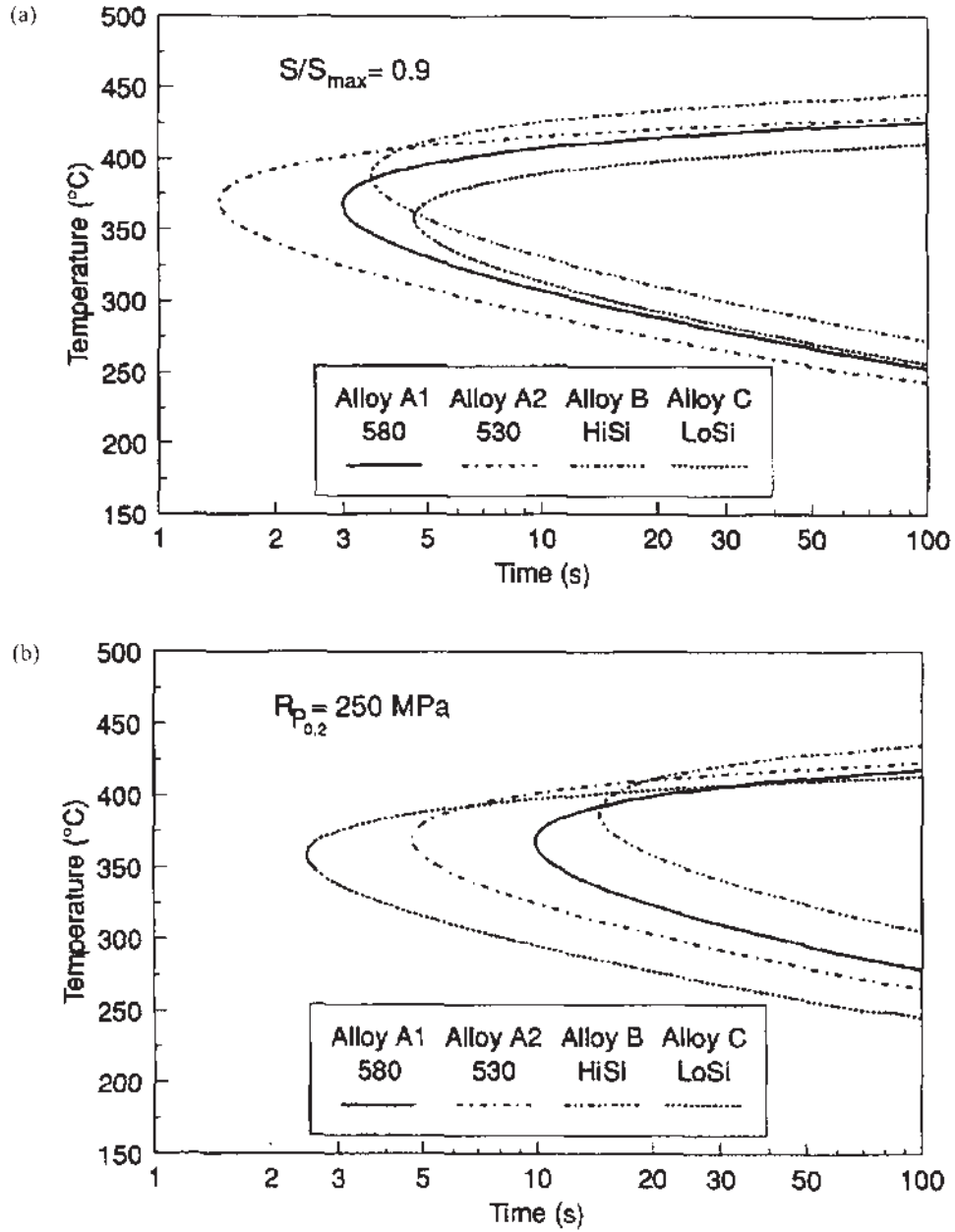
Figure 20 Theoretical temperature–time–property diagrams for (a) 90% of maximum yield strength and (b) constant yield strength, for AA6082. (Reprinted from Ref. [127], copyright 1997, with permission from Elsevier Science.)

where C_p is the particle solute content, and C_i is the interface concentration, which is related to the equilibrium concentration, C_e, by the Gibbs–Thomson relation. From this the size of particle that delineates the boundary between growth and dissolution is determined:

$$r_{crit} = \frac{2\sigma V_m}{RT} \left[\ln\left(\frac{C_m}{C_e}\right) \right]^{-1} \tag{70}$$

where σ is the interface energy, and V_m is the molar volume of the particle. Under non-isothermal conditions, r^* will vary with temperature, and must be computed at each step in a continuous cooling or heating cycle.

The particle size distribution is introduced by dividing the distribution into a series of elements of size Δr through which the flux of particles is calculated. The mean solute concentration is governed by a mass balance continuity equation:

$$C_m = C_0 - (C_p - C_m) \sum_i \frac{4}{3}\pi r_i^3 N_i \tag{71}$$

where C_0 is the solute content of the alloy, i is the index of the size class, and $N_i = \varphi_i \Delta r_i$ and φ_i is the size distribution function. All precipitates are assumed to be spherical. Myhr et al. show that this model can be calibrated to successfully predict the microstructural and property evolution of several Al–Mg–Si alloys under isothermal and non-isothermal conditions (Figs. 21 and 22). The authors do, however, comment that "reliable input data appears to be a limiting factor for its application to real (industrial) problems involving non-isothermal transformations," and this point may be generalized for all process models that deal with complex behavior. The tractability of the problem scales inversely with the number of dimensions introduced, and this can only be managed by calibration of the model.

A second class of analytical models uses yield strength as a proxy for microstructural evolution, and several authors have used conceptually simple models to good effect. The yield strength is defined as a function of the state variables of interest:

$$Y = f(S_1, S_2, S_3, \ldots) \tag{72}$$

Poole and colleagues [137,138] define precipitate spacing, mean size, and fraction as the state variables in a model of the precipitation sequence in AA7475, but further reduce these to a consideration solely of precipitate spacing by modeling in the coarsening regime (i.e., constant volume fraction), recognizing that particle size and spacing are coupled. The yield stress is modeled as

$$\sigma_{ys} = \sigma_0 + (\Delta\sigma_{ppt}^2 + \Delta\sigma_{WH}^2)^{1/2} \tag{73}$$

where σ_0 is the intrinsic Peierls strength, and $\Delta\sigma_{WH}$ is the strain-dependent contribution from work hardening. The contribution from precipitates is given by

$$\Delta\sigma_{ppt} = 2S_0 \frac{(\lambda^*)^{1/2}}{1 + (\lambda^*)^{3/2}} \tag{74}$$
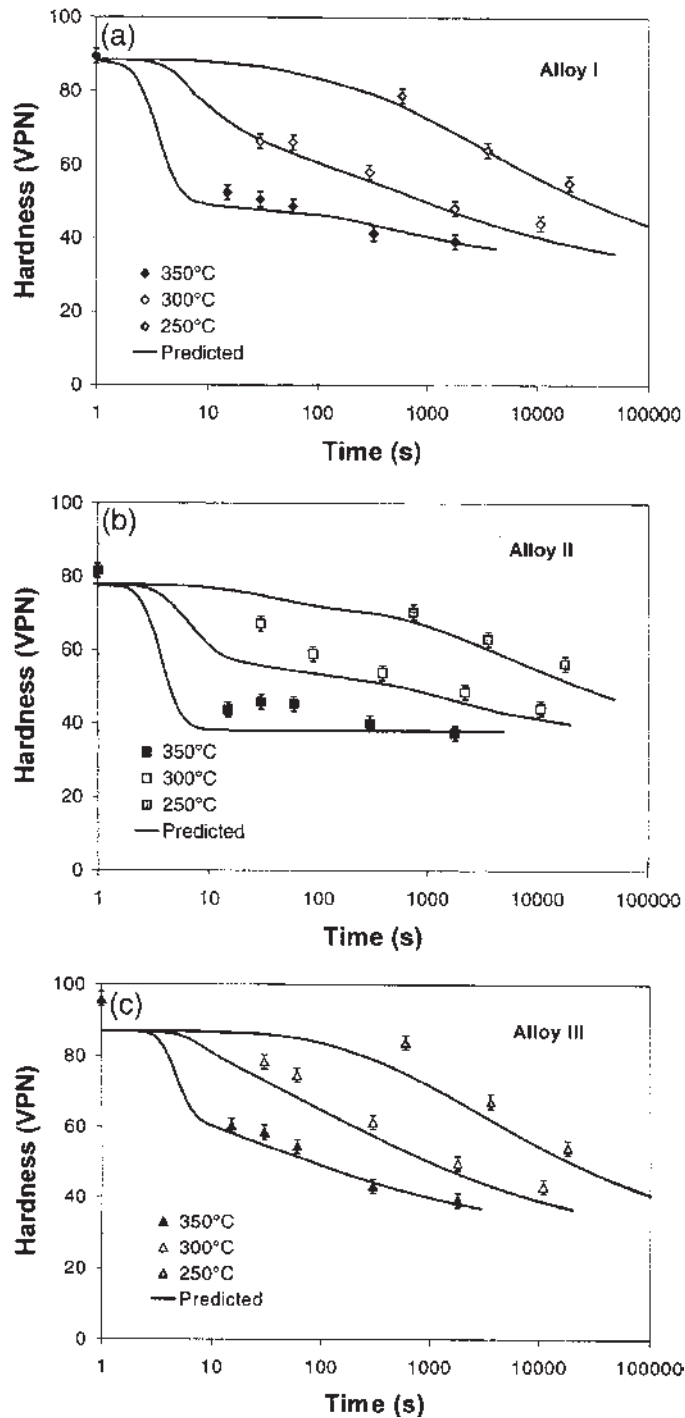
Figure 21 Predictions of reheat response of peak aged alloy for three alloys of different Mg/Si ratios: (a) 1:1; (b) 1.6:1; (c) 0.8:1. (Reprinted from Ref. [131], copyright 2001, with permission from Elsevier Science.)
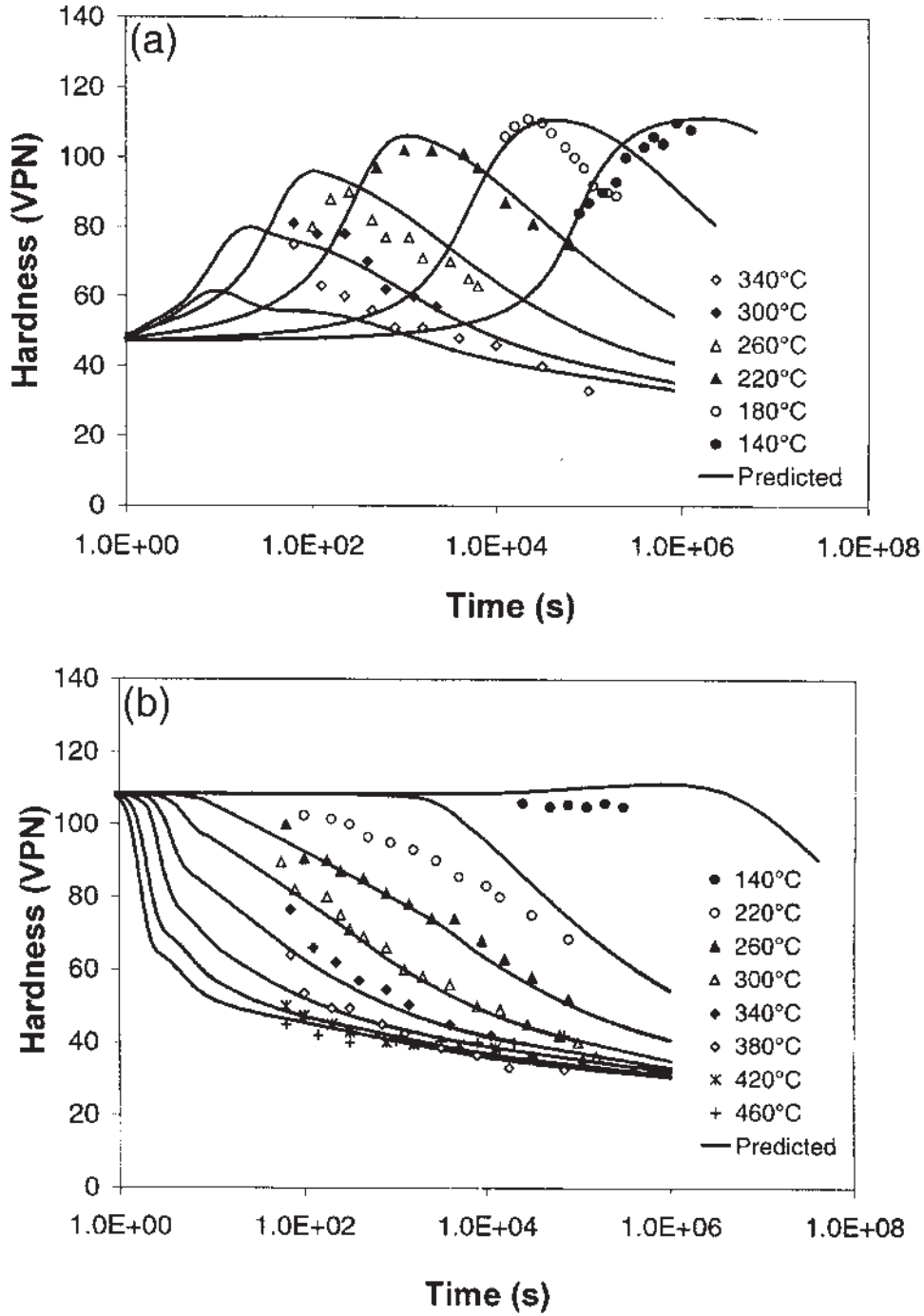
Figure 22 Prediction of (a) age hardening response and (b) reheat response of peak aged AA6082. (Reprinted from Ref. [131], copyright 2001, with permission from Elsevier Science.)

where S_0 is the peak precipitation strength of the alloy (i.e., excluding work hardening), and λ^* is the mean particle spacing normalized with respect to the spacing at peak strength. Evolution of particle spacing is calculated from conventional coarsening theory, except that ''kinetic strength'' substitutes for time:

$$\lambda^* = (P^*)^{1/3} \tag{75}$$

where the kinetic strength is given by

$$P^* = \frac{P}{P_p} \tag{76}$$

$$P = \sum \Delta t \frac{D_T}{D_{110°C}} \tag{77}$$

where P_p is the peak strength, D_T is the diffusion coefficient at the temperature of interest, and $D_{110°C}$ is the coefficient at a reference temperature. Under isothermal conditions, the kinetic strength is simply time at temperature. However, for non-isothermal conditions, the time is adjusted by the diffusion coefficient at the new temperature to give an equivalent time at the reference temperature, rather in the manner of the Zener–Hollomon parameter used to compare deformation (see Eq. (1), and cf. Eq. (19)). Although this model requires considerable calibration, it does allow two-step aging to be modeled once the calibration has been done. The disadvantage compared to the previously described models would appear to be its stiffness.

The evolution of precipitate shape coupled to crystallographic orientation has received little attention. Nie and coworkers [139,140] have applied simple analytical models for microstructural evolution to determine the effect of differing precipitate shape and orientations on strength, and have outlined how such models can be used to guide microstructural design. Separate expressions are derived for plates and rods on different crystallographic planes, and the transition from shear strengthening to Orowan strengthening is also modeled. The models result in good agreement with experimental data, and are remarkable for the prediction that ''further improvements in the strength of existing Al alloys might be achieved by increasing the number density and/or aspect ratio of rationally oriented precipitate plates'' [139]. Similar to the remarks of Bratland et al. [127] noted earlier, thus we see models being used to inform alloy design.

3.2   Mesoscale Simulations

As defined earlier, mesoscale models apply to length scales at the microstructural level — typically in the range 10^{-8} to 10^{-5} m — with the features explicitly modeled by numerical techniques. Perhaps because of the abundance of analytical models, and the size range that a model must span, very little has been accomplished to date using typical mesoscale techniques, such as MC or CA. Phase field simulations have been applied to the coarsening of δ in Al–Li alloys [141], in both two and three dimensions, and to the nucleation and growth of θ in Al–Cu alloys [142]. The form of the governing equations is similar to Eqs. (40)–(43) presented earlier, with, in the case of δ precipitation, three order parameters to describe the structural phase field,

and the addition of a composition component such that the free energy density function is written as

$$f(c, \eta) = \frac{A_1}{2}(c - C_1)^2 + \frac{A_2}{6}(C_2 - c)\sum_{i=1}^{3}\eta_i^2 - \frac{A_3}{3}\prod_{i=1}^{3}\eta_i + \frac{A_4}{24}\sum_{i=1}^{3}\eta_i^4 + \frac{A_5}{24}\sum_{i\neq j}\eta_i^2\eta_j^2$$

(78)

where C_1 and C_2 are constants close to the equilibrium compositions of the matrix and δ' phases, and A_i are constants that depend on temperature and composition. The free energy function is

$$F = \int\left(f(c, \eta) + \sum_{i=1}^{3}\frac{\beta_i}{2}(\nabla\eta_i)^2 + \frac{\alpha}{2}(\nabla c)^2\right)\mathrm{d}V$$

(79)

where α and β_i are gradient energy coefficients (cf. Eq. (41)). In addition to the equation representing the evolution of the structural variables (Eq. (40)), we also solve for the compositional evolution:

$$\frac{\partial c}{\partial t} = M\nabla^2\frac{\delta F}{\delta c}$$

(80)

where M is the mobility.

3.3 Atomic-Scale Simulations

As is often the case, modeling and experimental developments go hand-in-hand. In concert with the maturing of experimental techniques, such as 3-D position-sensitive atom probe, modelers have focused their attention on early-stage precipitation events. There are essentially two approaches to the simulation of precipitate evolution at the atomic length scale. Ab initio simulations apply statistical mechanics to resolve the pairwise interactions of atoms and calculate the optimum atomic configuration for a given set of conditions [143–146]. Ideally, the simulations are "parameter free" [143], in that no experimental data are required as input, but in practice, experimental data are often used to augment the calculations. In contrast, CALPHAD (CALculation of PHAse Diagrams) techniques require experimental thermodynamic data as input [147]. In either case, the models are combined with other techniques in order to predict the evolution of precipitate size and/or morphology with time and/or temperature.

   Ab initio or "first-principles" calculations have been successful in describing other aspects of phase transformations (e.g., magnetic phase transitions), but until recently have been ignored in the context of the age hardening of aluminum alloys. This is partly because the computational power required to run even small simulations has become available only relatively recently but also because the methods have been largely overlooked by physical metallurgists. It is only with the advent of multi-disciplinary groups, often inspired by an industrial need, that statistical mechanics has been applied to age hardening problems. The approach relies on the assumption that calculations over small volumes can be applied to all length scales, and thus a computationally manageable simulation may be conducted using a small number of

atoms and then scaled up to precipitate volumes. Texts by Stauffer [148] and Yeomans [149] provide an introduction to the theory underlying ab initio models.

Wolverton and coworkers [143–145] calculate the equilibrium formation enthalpy of any potential A_pB_q compound relative to the bulk constituents at their equilibrium lattice constants:

$$\Delta H_f^{eq}(A_pB_q) = E(A_pB_q) - xE_A^{eq}(A) - (1-x)E_B^{eq}(B) \tag{81}$$

where $E(A_pB_q)$, E_A^{eq}, and E_B^{eq} are the energies per atom of the compound A_pB_q, and constituents A and B; $x = p/(p+q)$. This method is also applied to ternary systems by evaluating the formation enthalpy of binary compounds in relation to the pure constituents in the fcc structure and interpolating between the data points for the ternary compound. For example, for the compound A_2BC, the formation enthalpy is given by [143]

$$\Delta H^{fcc}(A_2BC) = \frac{1}{2}\Delta H^{fcc}(AB) + \frac{1}{2}\Delta H^{fcc}(AC) + \frac{1}{4}\Delta H^{fcc}(BC) \tag{82}$$

The model is applied to Al–Cu–Mg–(Si) and Al–Mg–Zn alloys to evaluate proposed structural models based on experimental observations [143]. Combined with a Monte Carlo simulation, the precipitate shape may be predicted [143,145], and it is here that the engineering interest lies. An alternative ab initio method is the cluster variation method originated by Kikuchi [150], which computes configurational entropy as well as enthalpy. Asta and Hoyt [146] couple their calculations with a Monte Carlo simulation to predict GP zone shapes in Al–Ag alloys.

Although not based on first-principles calculations, but rather reliant on extensive experimental data, the CALPHAD technique [147] is more accessible by virtue of its embodiment in the Thermocalc† and, more recently, MTDATA‡ software. Löchte et al. [147] couple calculations of the free energies of metastable phases with an eulerian finite difference formulation of the Cahn–Hilliard equation to determine spatial and temporal evolution of precipitates in Al–Cu alloys. Given the widespread use of this technique for simulating precipitation in other systems (most notably steels), it is surprising that its application to aluminum alloys is so limited. This is possibly due to the misapprehension that the energetics of metastable phases are not able to be computed, but Löchte et al. [147] show that this is indeed possible. Research of this type is necessarily supported by large data sets, and recently, the European COST-507 initiative has led to the construction of a light alloy thermodynamic database [151–153], which may be used in conjunction with Thermocalc or MTDATA. With CALPHAD techniques being coupled to mesoscale simulations in other alloy systems [154], it is certain that this technique will become more common for aluminum-based alloys.

4   SUMMARY AND OUTLOOK

This chapter has described the models used to describe the secondary processing of aluminum alloys, but it has also illustrated the evolution of modeling since the advent of computers. From empirical models to simulations, the science of modeling

---

†Thermocalc is a product of Thermocalc Software, Stockholm, Sweden.
‡MTDATA is a product of the National Physical Laboratory, Teddington, UK.

has tracked the development of techniques available to solve the problems: first graph paper, followed by linear programming, superseded by parallel programming and/or computation that consumes as much computational power as is available. Yet, there is still a need for better input data and smarter modeling.

From a scientific point of view, an adequate metallurgical description of the deformed state and the mobility of grain boundaries is lacking, although advances in the EBSD technique will surely resolve this problem in the near future. Nucleation is rather poorly handled in most simulations, both of recrystallization and precipitation. From the engineering perspective, models that operate in isolation — simulating recrystallization but not precipitation, grain size but not texture, or deformation but not recrystallization, for example — are inadequate, and it is only recently that researchers have addressed the idea of "through-process" models that attempt to capture the physics of a process at a level necessary to provide a reliable prediction.

In the next five years we can expect increased hybridization of modeling techniques, as has been illustrated here with finite element/cellular automaton simulations, or CALPHAD/phase field models. This will bring with it its own set of problems — most notably, the coupling of the techniques in a computationally efficient manner.

ACKNOWLEDGMENTS

REFERENCES

1. C. M. Sellars and W. J. M. Tegart, " Hot Workability," Met. Rev. 1972, 17, p. 1.
2. D. Kuhlmann-Wilsdorf, "Theory of Work Hardening 1934–1984," Metall. Trans. 1985, 16A, pp. 2091–2108.
3. H. Mecking and U. F. Kocks, "Kinetics of Flow and Strain Hardening," Acta Metall. 1981, 29, pp. 1865–1875.
4. F.' R. N. Nabarro, "Work Hardening and Dynamical Recovery of fcc Metals in Multiple Glide," Acta Metall. 1989, 37, pp. 1521–1546.
5. H. Mughrabi, "A 2-Parameter Description of Heterogeneous Dislocation Distributions in Deformed Metal Crystals," Mater. Sci. Eng. 1987, 85, pp. 15–31.
6. F. R. Castro-Fernandez, C. M. Sellars, and J. A. Whiteman, "Changes of Flow Stress and Microstructure During Hot Deformation of Al–1Mg–1Mn," Mater. Sci. Technol. 1990, 6, pp. 453–460.
7. G. Sevillano, P. Van Houtte, and P. Aernoudt, "Large Strain Work Hardening and Textures," Prog. Mater. Sci. 1980, 25, pp. 69–412.
8. E. Nes, "Constitutive Laws for Steady State Deformation of Metals, a Microstructural Model," Scr. Metall. Mater. 1995, 33, pp. 225–231.
9. E. Nes and T. Furu, "Application of Microstructurally Based Constitutive Laws to Hot Deformation of Aluminium Alloys," Scr. Metall. Mater. 1995, 33, pp. 87–92.
10. E. Nes, "Modelling of Work Hardening and Stress Saturation in fcc Metals," Prog. Mater. Sci. 1998, 41, pp. 129–193.
11. E. Nes and K. Marthinsen, "Modelling Plastic Deformation in Aluminium Alloys by a Microstructural Approach," in: *Hot Deformation of Aluminum Alloys II* (T. R. Bieler, L. A. Lalli, and S. R. MacEwen, eds.), TMS, Warrendale, 1998, pp. 171–184.

12. K. Marthinsen and E. Nes, "The ALFLOW Model − a Microstructural Approach to Constitutive Plasticity-Modelling of Aluminium Alloys," Mater. Sci. Forum 2000, 331−337, pp. 1231−1242.

13. E. Schmid, "Zn − Normal Stress Law," in: Proceedings of the 1st International Congress on Applied Mechanics, Delft, 1924, p. 342.

14. G. Sachs, "Zur Ableitung einer Fliessbedingung," Z. Verein. Dtsch. Ing. 1928, 72, pp. 734−736.

15. J. Hirsch, K. Lucke, and H. Mecking, "Comparison of Experimental and Theoretical Rolling Textures of fcc Metals," in: *ICOTOM-7*, Seventh International Conference of Textures of Materials, (C. M. Brakman, P. Jongenburger, E. J. Mittemeijere, eds.), Netherlands Society for Materials Science, 1984, pp. 83−88.

16. M. Holscher, D. Raabe, and K. Lucke, "Relationship Between Rolling Textures and Shear Textures in fcc and bcc Metals," Acta Metall. Mater. 1994, 42, pp. 879−886.

17. A. Fieldly and H. J. Rowan, "Comparison of Anisotropy and Flow Behavior in two Al−Zn−Mg Extrusions − Recrystallized Versus Nonrecrystallized," Mater. Sci. Forum 1997, 242, pp. 135−140.

18. G. Winther, D. J. Jensen, and N. Hansen, "Modelling Flow Stress Anisotropy Caused by Deformation Induced Dislocation Boundaries," Acta Mater. 1997, 45, pp. 2455−2465.

19. W. M. Mao, "Modeling of Rolling Textures in Aluminum," Mater. Sci. Eng. A 1998, 257(1), pp. 171−177.

20. R. von Mises, "Mechanik der Plastischen Formänderung von Kristallen," Z. Angew. Math. Mech. 1928, 8, pp. 161−185.

21. G. I. Taylor, "Plastic Strain in Metals," J. Inst. Met. 1938, 62, pp. 307−324.

22. J. F. W. Bishop and R. Hill, "A Theory of the Plastic Distortion of a Polycrystalline Aggregate Under Combined Stresses," Philos. Mag. 1951, 42, pp. 414−427.

23. J. F. W. Bishop and R. Hill, "A Theoretical Derivation of the Plastic Properties of a Polycrystalline Face-Centered Metal," Philos. Mag. 1951, 42, pp. 1298−1307.

24. R. J. Asaro and A. Needleman, "Texture Development and Strain Hardening in Rate Dependent Polycrystals," Acta Metall. 1985, 33, pp. 923−953.

25. H. Honeff and H. Mecking, "A Method for the Determination of the Active Slip Systems and Orientation Changes During Single Crystal Deformation," in: *ICOTOM-V*, Fifth International Conference on Textures of Materials, (G. Gottstein and K. Lücke, eds.), Springer-Verlag, Berlin, 1978, pp. 265−275.

26. U. F. Kocks and G. R. Canova, "How Many Slip Systems and Which?" in: *Deformation of Polycrystals: Mechanisms and Microstructures* (N. Hansen, et al., eds.), Risø National Laboratory, Roskilde, Denmark, 1981, pp. 35−44.

27. G. R. Canova, U. F. Kocks, and J. J. Jonas, "Theory of Torsion Texture Development," Acta Metall. 1984, 32, pp. 211−226.

28. C. Tome, G. R. Canova, U. F. Kocks, N. Christodoulou, and J. J. Jonas, "The Relation Between Macroscopic and Microscopic Strain-Hardening in fcc Polycrystals," Acta Metall. 1984, 32, pp. 1637−1653.

29. E. Kroner, "Zur Plastischen Verformung des Vielkristalls," Acta Metall. 1961, 9, p. 155.

30. B. Budiansky and T. T. Wu, "Theoretical Prediction of Plastic Strains of Polycrystals", in: Proceedings of the Fourth U.S. National Congress on Applied Mechanics, American Society of Mechanical Engineers, New York, 1962, p. 1175.

31. R. Hill, "Continuum Micro-Mechanics of Elastoplastic Polycrystals," J. Mech. Phys. Solids, 1965, 13, pp. 89−101.

32. J. J. Gracio, A. B. Lopes, and E. F. Rauch, "Analysis of Plastic Instability in Commercially Pure Al Alloys," J. Mater. Process. Technol. 2000, 103, pp. 160−164.

33. S.-H. Choi, J. C. Brem, F. Barlat, and K. H. Oh, "Microscopic Anisotropy in AA5019A Sheets," Acta Mater. 2000, 48, pp. 1853−1863.

34. G. B. Sarma and P. R. Dawson, "Texture Predictions Using a Polycrystal Plasticity Model Incorporating Neighbor Interactions," Int. J. Plast. 1996, 12, pp. 1023−1054.

35. K. K. Mathur and P. R. Dawson, "On Modeling the Development of Crystallographic Texture in Bulk Forming Processes," Int. J. Plast. 1989, 5, pp. 67–94.

36. K. K. Mathur, P. R. Dawson, and U. F. Kocks, "On Modeling Anisotropy in Deformation Processes Involving Textured Polycrystals with Distorted Grain Shape," Mech. Mater. 1990, 10, pp. 183–202.

37. H. Aretz, R. Luce, M. Wolske, R. Kopp, M. Goerdeler, V. Marx, G. Pomana, and G. Gottstein, "Integration of Physically Based Models into FEM and Application in Simulation of Metal Forming Processes," Model. Simul. Mater. Sci. Eng. 2000, 8, pp. 881–891.

38. A. J. Beaudoin, Jr., P. R. Dawson, and K. K. Mathur, "Analysis of Anisotropy in Sheet Forming Using Polycrystal Plasticity," in: Proceedings of the Tenth International Conference on Engineering Mechanics, Vol. 2, (S. Stein, ed.), American Society of Civil Engineers, New York, 1995, pp. 1018–1021.

39. S. Balasubramanian and L. Anand, "Single Crystal and Polycrystal Elasto-Viscoplasticity: Application to Earing in Cup Drawing of fcc Materials," Comput. Mech. 1996, 17, pp. 209–225.

40. J. Hu, J. J. Jonas, and T. Ishikawa, "FEM Simulation of the Forming of Textured Aluminum Sheets," Mater. Sci. Eng. A 1998, 256(1–2), pp. 51–59.

41. K. Inal, P. D. Wu, and K. W. Neale, "Simulation of Earing in Textured Aluminum Sheets," Int. J. Plast. 2000, 16, pp. 635–648.

42. Y. P. Chen, X. H. Dong, C. L. Xie, Z. G. Li, and E. Nakamachi, "Prediction of Earing in Textured Aluminum Sheets Using Crystal Plasticity," Acta Mech. Sol. Sin. 2001, 14, pp. 95–103.

43. A. J. Beaudoin, P. R. Dawson, K. K. Mathur, U. F. Kocks, and D. A. Korzekwa, "Application of Polycrystal Plasticity to Sheet Forming," Comput. Meth. Appl. Mech. Eng. 1994, 117, pp. 49–70.

44. J. Hu, T. Ishikawa, and J. J. Jonas, "Finite Element Analysis of Damage Evolution and the Prediction of the Limiting Draw Ratio in Textured Aluminum Sheets," J. Mater. Process. Technol. 2000, 103, pp. 374–382.

45. S. E. Schoenfeld and R. J. Asaro, "Through Thickness Texture Gradients in Rolled Polycrystalline Alloys," Int. J. Mech. Sci. 1996, 38, pp. 661–683.

46. T. Aukrust, S. Tjotta, H. E. Vatne, and P. Van Houtte, "Coupled FEM and Texture Modeling of Plane Strain Extrusion of Aluminum Alloy," Int. J. Plast. 1997, 13, pp. 111–125.

47. F. Perocheau and J. Driver, "Texture Gradient Simulations for Extrusion and Reversible Rolling of fcc Metals," Int. J. Plast. 2000, 16, pp. 73–89.

48. R. Becker, "Analysis of Texture Evolution in Channel Die Compression. 1. Effects of Grain Interaction," Acta Metall. Mater. 1991, 39, pp. 1211–1230.

49. C. A. Bronkhorst, S. R. Kalidindi, and L. Anand, "Polycrystalline Plasticity and the Evolution of Crystallographic Texture in fcc Metals," Philos. Trans. Roy. Soc. London A 1992, 341, pp. 443–477.

50. L. Anand and S. R. Kalidindi, "The Process of Shear-Band Formation in Plane-Strain Compression of fcc Metals – Effects of Crystallographic Texture," Mech. Mater. 1994, 17, pp. 223–243.

51. R. Becker and S. Panchanadeeswaran, "Effects of Grain Interactions on Deformation and Local Texture in Polycrystals," Acta Metall. Mater. 1995, 43, pp. 2701–2719.

52. P. Bate, "Modelling Deformation Microstructure with the Crystal Plasticity Finite-Element Method," Philos. Trans. Roy. Soc. London A 1999, 357, pp. 1589–1601.

53. A. J. Beaudoin, P. R. Dawson, K. K. Mathur, and U. F. Kocks, "A Hybrid Finite-Element Formulation for Polycrystal Plasticity with Consideration of Macrostructural and Microstructural Linking," Int. J. Plast. 1995, 11, pp. 501–521.

54. A. J. Beaudoin, H. Mecking, and U. F. Kocks, "Development of Localized Orientation Gradients in fcc Polycrystals," Philos. Mag. A 1996, 73, pp. 1503–1517.

55. G. B. Sarma, B. Radhakrishnan, and T. Zacharia, "Finite Element Simulations of Cold Deformation at the Mesoscale," Comput. Mater. Sci. 1998, 12, pp. 105–123.

56. D. P. Mika and P. R. Dawson, "Effects of Grain Interaction on Deformation in Polycrystals," Mater. Sci. Eng. A 1998, 257(1), pp. 62–76.

57. P. R. Dawson and E. B. Marin, "Computational Mechanics for Metal Deformation Processes Using Polycrystal Plasticity," Adv. Appl. Mech. 1998, 34, pp. 77–169.

58. B. Radhakrishnan, G. Sarma, H. Weiland, and P. Baggethun, "Simulations of Deformation and Recrystallization of Single Crystals of Aluminium Containing Hard Particles," Model. Simul. Mater. Sci. Eng. 2000, 8, pp. 737–750.

59. H. Yamagata, "Microstructural Evolution of Single Crystalline Aluminum During Multiple Peak Stress Oscillation at 623 K," Scr. Metall. Mater. 1992, 27, pp. 1157–1160.

60. F. R. Castro-Fernandez and C. M. Sellars, "Static Recrystallization and Recrystallization During Hot Working of Al–1Mg–1Mn Alloy," Mater. Sci. Technol. 1988, 4, pp. 621–627.

61. S. Gourdet and F. Montheillet, "An Experimental Study of the Recrystallization Mechanism During Hot Deformation of Aluminium," Mater. Sci. Eng. A 2000, 283, pp. 274–288.

62. H. J. McQueen and W. Blum, "Dynamic Recovery: Sufficient Mechanism in the Hot Deformation of $Al < 99.99$," Mater. Sci. Eng. A 2000, 290, pp. 95–107.

63. W. Blum, Q. Zhu, R. Merkel, and H. J. McQueen, "Geometric Dynamic Recrystallization in Hot Torsion of Al–5Mg–0.6Mn AA5083," Mater. Sci. Eng. A 1996, 205, pp. 23–30.

64. F. J. Humphreys and M. Hatherly, *Recrystallization and Related Annealing Phenomena*, Elsevier Science Ltd., Oxford, 1995.

65. M. S. Mirza, C. M. Sellars, K. Karhausen, and P. Evans, "Multipass Rolling of Aluminum Alloys: Finite Element Simulations and Microstructural Evolution," Mater. Sci. Technol. 2001, 17, pp. 874–879.

66. B. K. Chen, P. F. Thompson, and S. K. Choi, "Computer Modelling of Microstructure During Hot Flat Rolling of Aluminum," Mater. Sci. Technol. 1992, 8, pp. 72–77.

67. M. A. Wells, D. J. Lloyd, I. V. Samarasekera, J. K. Brimacombe, and E. B. Hawbolt, "Modeling the Microstructural Changes During Hot Tandem Rolling of Aluminum Alloys: Part III. Overall Model Development, Validation," Metall. Mater. Trans. B 1998, 29, pp. 709–719.

68. H. E. Vatne, T. Furu, R. Ørsund, and E. Nes, "Modelling Recrystallization After Hot Deformation of Aluminium," Acta Mater. 1996, 44, pp. 4463–4473.

69. H. E. Vatne, T. Furu, and E. Nes, "Microstructural Modelling of Recrystallization Applied to Industrial Processing of Metals," in: Proceedings of ReX'96, The Third International Conference on Recrystallization and Related Phenomena (T. R. McNelley, ed.), MIAS, Monterey, 1997, pp. 95–108.

70. H. E. Vatne, T. Furu, and E. Nes, "Nucleation of Recrystallized Grains from Cube Bands in Hot Deformed Commercial Purity Aluminium," Mater. Sci. Technol. 1996, 12, pp. 201–210.

71. O. Daaland and E. Nes, "Origin of Cube Texture During Hot Rolling of Commercial Al–Mn–Mg Alloys," Acta Mater. 1996, 44, pp. 1389–1411.

72. O. Daaland and E. Nes, "Recrystallization Texture Development in Al–Mn–Mg Alloys," Acta Mater. 1996, 44, pp. 1413–1435.

73. C. M. Sellars and Q. Zhu, "Microstructural Modelling of Aluminium Alloys During Thermomechanical Processing," Mater. Sci. Eng. A 2000, 280, pp. 1–7.

74. M. Upmanyu, Private Communication, Oak Ridge National Laboratory, 2002.

75. D. Juul Jensen, "Microstructural and Crystallographic Aspects of Recrystallization," in: Proceedings of the 16th Risø International Symposium (N. Hansen, et al., eds.), Risø, Roskilde, 1995, p. 119.

76. T. Furu, R. Ørsund, and E. Nes, "Substructure Evolution During Different Hot Deformation Processes of Commercial Non-Heat Treatable Aluminium Alloys," Mater. Sci. Eng. A 1996, 214, pp. 122–132.

77. Q. Liu, D. J. Jensen, and N. Hansen, "Effect of Grain Orientation on Deformation Structure in Cold-Rolled Polycrystalline Aluminium," Acta Mater. 1998, 46, pp. 5819–5838.

78. B. Radhakrishnan, G. Sarma, and T. Zacharia, "Modeling the Kinetics and Microstructural Evolution During Static Recrystallization − Monte Carlo Simulation of Recrystallization," Acta Mater. 1998, 46, pp. 4415–4433.

79. H. E. Vatne, K. Marthinsen, R. Orsund, and E. Nes, "Modeling Recrystallization Kinetics, Grain Sizes and Textures During Multipass Hot Rolling," Metall. Mater. Trans. A 1996, 27, pp. 4133–4144.

80. I. L. Dillamore and H. Katoh, "Mechanisms of Recrystallization in Cubic Metals with Particular Reference to their Orientation Dependence," Metal Sci. 1974, 8, pp. 73–83.

81. A. A. Ridha and W. B. Hutchinson, "Recrystallization Mechanisms and the Origin of Cube Texture in Copper," Acta Metall. Mater. 1982, 30, pp. 1929–1939.

82. J. Hjelen, R. Ørsund, and E. Nes, "On the Origin of Recrystallization Textures in Aluminium," Acta Metall. Mater. 1991, 39, pp. 1377–1404.

83. R. D. Doherty and S. Panchanadeeswaran, "Direct Observation of Recrystallization Texture in Commercial Purity Aluminum," Acta Metall. Mater. 1993, 41, pp. 3029–3053.

84. M. G. Ardakani and F. J. Humphreys, "The Annealing Behavior of Deformed Particle-Containing Aluminium Alloys," Acta Metall. Mater. 1994, 42, pp. 763–780.

85. O. Engler, "On the Influence of Orientation Pinning on Growth Selection of Recrystallization," Acta Mater. 1998, 46, pp. 1555–1568.

86. O. Engler and H. E. Vatne, "Modeling the Recrystallization Textures of Aluminum Alloys After Hot Deformation," JOM-J. Min. Met. Mat. S. 1998, June, pp. 23–27.

87. D. N. Lee, "Strain Energy Release Maximization Model for Evolution of Recrystallization Textures," Int. J. Mech. Sci. 2000, 42, pp. 1645–1678.

88. M. Blicharski, J. Liu, and H. Hu, "Annealing Behavior of Aluminum Bicrystals with S-Orientations Deformed by Channel Die Compression," Acta Metall. Mater. 1995, 43, pp. 3125–3138.

89. N. Rajmohan and J. A. Szupunar, "A New Model for the Recrystallization of Heavily Cold-Rolled Aluminum Using Orientation-Dependent Stored Energy," Acta Mater. 2000, 48, pp. 3327–3340.

90. D. J. Srolovitz, M. Anderson, and G. S. Grest, "Grain Growth in 2 Dimensions," Scr. Metall. Mater. 1983, 17, pp. 241–246.

91. G. S. Grest, D. J. Srolovitz, and M. P. Anderson, "Computer Simulation of Grain Growth. 4. Anisotropic Grain Boundary Energies," Acta Metall. Mater. 1985, 33, pp. 509–520.

92. A. D. Rollett and E. A. Holm, "Abnormal Grain Growth − the Origin of Recrystallization Nuclei?" in: Proceedings of ReX'96, The Third International Conference on Recrystallization and Related Phenomena (T. R. McNelley, eds.), MIAS, Monterey, 1997, pp. 31–41.

93. D. J. Srolovitz, G. S. Grest, and M. P. Anderson, "Computer Simulation of Recrystallization. 1. Homogeneous Nucleation and Growth," Acta Metall. 1986, 34, pp. 1833–1845.

94. D. J. Srolovitz, G. S. Grest, and M. P. Anderson, "Computer Simulation of Recrystallization. 2. Heterogeneous Nucleation and Growth," Acta Metall. 1988, 36, pp. 2115–2128.

95. A. D. Rollett, et al., "Computer Simulation of Recrystallization in Non-Uniformly Deformed Metals," Acta Metall. 1989, 37, pp. 627–639.

96. B. Radhakrishnan and T. Zacharia, "The Effect of Lattice Temperature on Abnormal

Subgrain Growth Simulations Using a Monte Carlo Technique," Interf. Sci., 2002, 10, pp. 171–180.

97. C. H. J. Davies, "Growth of Nuclei in a Cellular Automaton Simulation of Recrystallization," Scr. Mater. 1997, 36, pp. 35–40.

98. C. H. J. Davies and L. Hong, "The Cellular Automaton Simulation of Static Recrystallization in Cold-Rolled AA1050," Scr. Mater. 1999, 40, pp. 1145–1150.

99. D. Raabe, "Introduction of a Scalable Three-Dimensional Cellular Automaton with a Probabilistic Switching Rule for the Discrete Mesoscale Simulation of Recrystallization Phenomena," Philos. Mag. 1999, 79, pp. 2339–2358.

100. A. D. Rollett and D. Raabe, "A Hybrid Model for Mesoscopic Simulation of Recrystallization," Comput. Mater. Sci. 2001, 21, pp. 69–78.

101. L.-Q. Chen, "A Novel Computer Simulation Technique for Modeling Grain Growth," Scr. Metall. Mater. 1995, 32, pp. 115–120.

102. A. Kazaryan, Y. Wang, S. A. Dregia, and B. R. Patton, "Generalized Phase-Field Model for Computer Simulation of Grain Growth in Anisotropic Systems," Phys. Rev. 2000, 61, pp. 14275–14278.

103. B. Radhakrishnan, G. Sarma, T. Zacharia, A. Kazaryan, and Y. Wang, "Mesoscale Modeling of Abnormal Subgrain Growth," in: Proceedings of the Symposium on Microstructure Modeling and Prediction during Thermomechanical Processing (R. Srinivasan, et al., eds.), TMS, Warrendale, 2001, pp. 47–57.

104. H. J. Frost, C. V. Thompson, C. L. Howe, and J. H. Whang, "A 2-Dimensional Computer Simulation of Capillarity-Driven Grain Growth − Preliminary Results," Scr. Metall. Mater. 1988, 22, pp. 65–70.

105. A. Soares, A. C. Ferro, and M. A. Fortes, "Computer Simulation of Grain Growth in a Bidimensional Polycrystal," Scr. Metall. Mater. 1985, 19, pp. 1491–1496.

106. K. Kawasaki, T. Nagai, and K. Nakashima, "Vertex Models for Two-Dimensional Grain Growth," Philos. Mag. 1989, 60, pp. 399–421.

107. F. J. Humphreys, "A Network Model for Recovery and Recrystallization," Scr. Metall. Mater. 1992, 27, pp. 1557–1562.

108. Y. Huang and F. J. Humphreys, "Measurement of Grain Boundary Mobility During Recrystallization of a Single Phase Aluminium Alloy," Acta Mater. 1999, 47, pp. 2259–2268.

109. Y. Huang, F. J. Humphreys, and M. Ferry, "The Annealing Behaviour of Deformed Cube-Oriented Aluminium Single Crystals," Acta Mater. 2000, 48, pp. 2543–2556.

110. Y. Huang and F. J. Humphreys, "Subgrain Growth and Low Angle Boundary Mobility in Aluminium Crystals of Orientation {110} ⟨0 0 1⟩," Acta Mater. 2000, 48, pp. 2017–2030.

111. E. M. Lauridsen, D. Juul Jensen, and H. F. Poulsen, "Kinetics of Individual Grains During Recrystallization," Scr. Mater. 2000, 43, pp. 561–566.

112. B. Radhakrishnan, G. Sarma, and T. Zacharia, "Monte Carlo Simulation of Deformation Substructure Evolution During Recrystallization," Scr. Mater. 1998, 39, pp. 1639–1645.

113. D. Raabe and R. C. Becker, "Coupling of a Crystal Plasticity Finite Element Model with a Probabilistic Cellular Automaton for Simulating Primary Static Recrystallization in Aluminium," Model. Simul. Mater. Sci. Eng. 2000, 8, pp. 445–462.

114. Ø. Grong and H. R. Shercliff, "Microstructural Modelling in Metals Processing," Prog. Mater. Sci. 2002, 47, pp. 163–282.

115. S. Fletcher, M. Thomson, and T. Tran, "Numerical Analysis of 2D Nucleation/Growth/Collision Processes: Part 1, Test of the Kolmogoroff–Avrami Theorem," J. Electroanal. Chem. 1986, 199, pp. 241–247.

116. M. J. Starink, C. Y. Zahra, and A.-M. Zahra, "Analysis of Precipitation in Al-Based

Alloys Using a Novel Model for Nucleation and Growth Reactions," J. Thermal Anal. Calorim. 1998, 51, pp. 933–942.

117. M. J. Starink and A.-M. Zahra, "Kinetics of Isothermal β Precipitation in Al–Mg Alloys," J. Mater. Sci. 1999, 34, pp. 1117–1127.

118. M. J. Starink, P. Wang, I. Sinclair, and P. J. Gregson, "Microstructure and Strengthening of Al–Li–Cu–Mg Alloys and MMCS: I. Analysis and Modelling of Microstructural Changes," Acta Mater. 1999, 47, pp. 3841–3853.

119. M. J. Starink, P. Wang, I. Sinclair, and P. J. Gregson, "Microstructure and Strengthening of Al–Li–Cu–Mg Alloys and MMCS: II. Modelling of Yield Strength," Acta Mater. 1999, 47, pp. 3855–3868.

120. E.-S. Kim and Y. G. Kim, "A Transformation Kinetic Model and its Application to Cu–Zn–Al Shape Memory Alloys − I Isothermal Conditions," Acta Metall. Mater. 1990, 38, pp. 1669–1676.

121. E.-S. Kim and Y. G. Kim, "A Transformation Kinetic Model and its Application to Cu–Zn–Al Shape Memory Alloys − II Non-Isothermal Conditions," Acta Metall. Mater. 1990, 38, pp. 1677–1686.

122. M. J. Starink, "Kinetic Equations for Diffusion Controlled Precipitation Reactions," J. Mater. Sci. 1997, 32, pp. 4061–4070.

123. M. J. Starink and A.-M. Zahra, "β and β Precipitation in an Al–Mg Alloy Studied by DSC and TEM," Acta Mater. 1998, 46, pp. 3381–3397.

124. A. Borrego and G. González-Doncel, "Calorimetric Study of 6061-Al–15 vol% SiC$_w$ PM Composites Extruded at Different Temperatures," Mater. Sci. Eng. A 1998, 245, pp. 10–18.

125. A. Borrego and G. González-Doncel, "Analysis of the Precipitation Behavior in Aluminum Matrix Composites from a Johnson–Mehl–Avrami Kinetic Model," Mater. Sci. Eng. A 1998, 252, pp. 149–152.

126. J. W. Cahn, "The Kinetics of Grain Boundary Nucleated Reactions," Acta Metall. 1956, 4, pp. 449–459.

127. D. H. Bratland, Ø. Grong, H. Shercliff, O. R. Myhr, and S. Tjotta, "Overview No. 124: Modelling of Precipitation Reactions in Industrial Processing," Acta Mater. 1997, 45, pp. 1–22.

128. B. I. Bjorneklett, Ø. Grong, O. R. Myhr, and A. O. Kluken, "Additivity and Isokinetic Behaviour in Relation to Particle Dissolution," Acta Mater. 1998, 46, pp. 6257–6288.

129. Ø. Grong and O. R. Myhr, "Additivity and Isokinetic Behaviour in Relation to Diffusion Controlled Growth," Acta Mater. 2000, 48, pp. 445–452.

130. O. R. Myhr and Ø. Grong, "Modelling of Non-Isothermal Transformations in Alloys Containing a Particle Distribution," Acta Mater. 2000, 48, pp. 1605–1615.

131. O. R. Myhr, Ø. Grong, and S. J. Andersen, "Modelling of the Age Hardening Behaviour of Al–Mg–Si Alloys," Acta Mater. 2001, 49, pp. 65–75.

132. A. Deschamps and Y. Bréchet, "Influence of Predeformation and Ageing of an Al–Zn–Mg Alloy II Modelling of Precipitation Kinetics and Yield Stress," Acta Mater. 1999, 47, pp. 293–305.

133. J. C. Werenskiold, A. Deschamps, and Y. Bréchet, "Characterization and Modeling of Precipitation Kinetics in an Al–Zn–Mg Alloy," Mater. Sci. Eng. A 2000, 293, pp. 267–274.

134. J. D. Robson and P. B. Prangnell, "Dispersoid Precipitation and Process Modelling in Zirconium Containing Commercial Aluminum Alloys," Acta Mater. 2001, 49, pp. 599–613.

135. A. W. Zhu, "Evolution of Size Distribution of Shearable Ordered Precipitates Under Homogeneous Deformation: Application to an Al–Li Alloy," Acta Mater. 1997, 45, pp. 4213–4223.

136. R. Wagner and R. Kampmann, "Homogeneous Second Phase Precipitation," in: *Phase Transformations* (R. W. Cahn, P. Haasen, and E. J. Kramer, eds.), Materials Science and Technology, a Comprehensive Treatise, Vol. 5, VCH, Weinheim, 1991.

137. W. J. Poole, H. R. Shercliff, and T. Castillo, "Process Model for Two Step Age Hardening of 7475 Aluminium Alloy," Mater. Sci. Technol. 1997, 13, pp. 897–904.

138. W. J. Poole, J. A. Sæter, S. Skjervold, and G. Waterloo, "A Model for Predicting the Effect of Deformation After Solution Treatment on the Subsequent Artificial Aging Behavior of AA7030 and AA7108 Alloys," Metall. Mater. Trans. A 2000, 31, pp. 2327–2338.

139. J. F. Nie and B. C. Muddle, "Microstructural Design of High Strength Aluminum Alloys," J. Phase Equilibria 1998, 19, pp. 543–551.

140. J. F. Nie, B. C. Muddle, and I. J. Polmear, "The Effect of Precipitate Shape and Orientation on Dispersion Strengthening in High Strength Aluminium Alloys," Mater. Sci. Forum 1996, 217–222, pp. 1257–1262.

141. V. Vaithyanathan and L. Q. Chen, "Coarsening Kinetics of δ-Al$_3$Li Precipitates: Phase Field Simulation in 2D and 3D," Scr. Mater. 2000, 42, pp. 967–973.

142. D. Y. Li and L. Q. Chen, "Computer Simulation of Stress-Oriented Nucleation and Growth of Precipitates in Al–Cu Alloys," Acta Mater. 1998, 46, pp. 2573–2585.

143. C. Wolverton, "Crystal Structure and Stability of Complex Precipitate Phases in Al–Cu–Mg–Si and Al–Zn–Mg Alloys," Acta Mater. 2001, 49, pp. 3129–3142.

144. C. Wolverton, "First-Principles Theory of 250,000-Atom Coherent Alloy Microstructure," Model. Simul. Mater. Sci. Eng. 2000, 8, pp. 323–333.

145. S. Muller, C. Wolverton, L.-W. Wang, and A. Zunger, "Predicting the Size- and Temperature-Dependent Shapes of Precipitates in Al–Zn Alloys," Acta Mater. 2000, 48, pp. 4007–4020.

146. M. Asta and J. J. Hoyt, "Thermodynamic Properties of Coherent Interfaces in fcc-Based Ag–Al Alloys: A First Principles Study," Acta Mater. 2000, 48, pp. 1089–1096.

147. L. Löchte, A. Gitt, G. Gottstein, and I. Hurtado, "Simulation of the Evolution of GP Zones in Al–Cu Alloys: An Extended Cahn–Hilliard Approach," Acta Mater. 2000, 48, pp. 2969–2984.

148. D. Stauffer, In*troduction to Percolation Theory*, Taylor and Francis, London, 1985.

149. J. M. Yeomans, *Statistical Mechanics of Phase Transitions*, Oxford University Press, Oxford, 1992.

150. R. A. Kikuchi, "Theory of Cooperative Phenomena," Phys. Rev. 1951, 81, pp. 988–1003.

151. S. G. Fries and T. Jantzen, "Compilation of CALPHAD Formation Enthalpy Data − Binary Intermetallics in the COST507 Gibbsian Database," Thermochim. Acta 1998, 314, pp. 23–33.

152. T. Buhler, S. G. Fries, P. J. Spencer, and H. L. Lukas, "A Thermodynamic Assessment of the Al–Cu–Mg Ternary System," J. Phase Equilibria 1998, 19, pp. 317–331.

153. H. J. Siefert, P. Liang, H. L. Lukas, F. Aldinger, S. G. Fries, M. G. Harmelin, F. Faudot, and T. Jantzen, "Computational Phase Studies on Commercial Aluminium and Magnesium Alloys," Mater. Sci. Technol. 2000, 16, pp. 1429–1433.

154. U. Grafe, B. Böttger, J. Tiaden, and S. G. Fries, "Coupling of Multicomponent Thermodynamic Databases to a Phase Field Model: Application to Solidification and Solid State Transformations of Superalloys," Scr. Mater. 2000, 42, pp. 1179–1186.

8

Texture–Property Relationships in Aluminum Alloys: Simulations and Experiments

DIERK RAABE

*Max Planck Institute for Iron Research
Düsseldorf, Germany*

1 INTRODUCTION

Modern well-tailored and optimized aluminum alloys provide a huge and steadily growing application spectrum to customers of formed and cast products. A good example is the 6xxx series aluminum alloys which gradually gain momentum in the automotive industry. The high demands with respect to mechanical properties and surface appearance faced by these materials increasingly require adequate quantitative characterization measures, simulation tools, and basic understanding to build a bridge between materials scientists, materials producers, and final product designers.

Current approaches for conducting simulations, for instance, of plastic deformation of aluminum and its alloys are usually based on solving large sets of differential equations associated with a well-posed forming problem by using finite element methods. Primary objectives of such simulations are the prediction of the material shape after forming, in particular, the thickness distribution, the minimization of material failure, the optimization of the material flow during forming, the simulation of the springback effect, and the calculation of the final mechanical properties of the formed sample. Related essential applications are in the fields of optimizing tool designs, predicting pressing forces, and simulating the final surface appearance of the part. The latter aspect involves both macroscopical (e.g. wrinkling) as well as microstructural (e.g. ridging, strain localization, orange peel) mechanisms of surface changes during forming.

Rendering such continuum-type metal-forming simulations scientifically sound, predictive at the microstructure scale, in good accord with experiment, and at the same time economically rewarding requires that the involved materials are properly specified in terms of their respective constitutive behavior. For this purpose, finite element approaches for advanced metal-forming simulations typically employ three sets of material input data covering strain hardening, failure behavior and forming limits, as well as texture and anisotropy (Fig. 1). The current chapter deals, in particular, with the latter aspect and consists of three parts. The first part presents an introduction to some of the employed experimental and theoretical methods. The second part reviews some microscopic aspects of texture development focusing on plastic deformation, primary static recrystallization, and the relationship between texture and surface properties. The third part reviews concepts for the integration of crystallographic texture and anisotropy into large scale metal-forming simulations.

2 EXPERIMENTAL AND THEORETICAL METHODS

2.1 Experimental Determination of Surface Strains

Surface strains on plastically deformed samples can be experimentally determined using photogrametry. This is a digital image analysis method which quantifies changes in surface patterns after straining [1,2]. Both the natural characteristics of an unprepared sample surface or a stochastic color spray applied to a polished surface can provide an exploitable input pattern (Fig. 2). In order to measure the
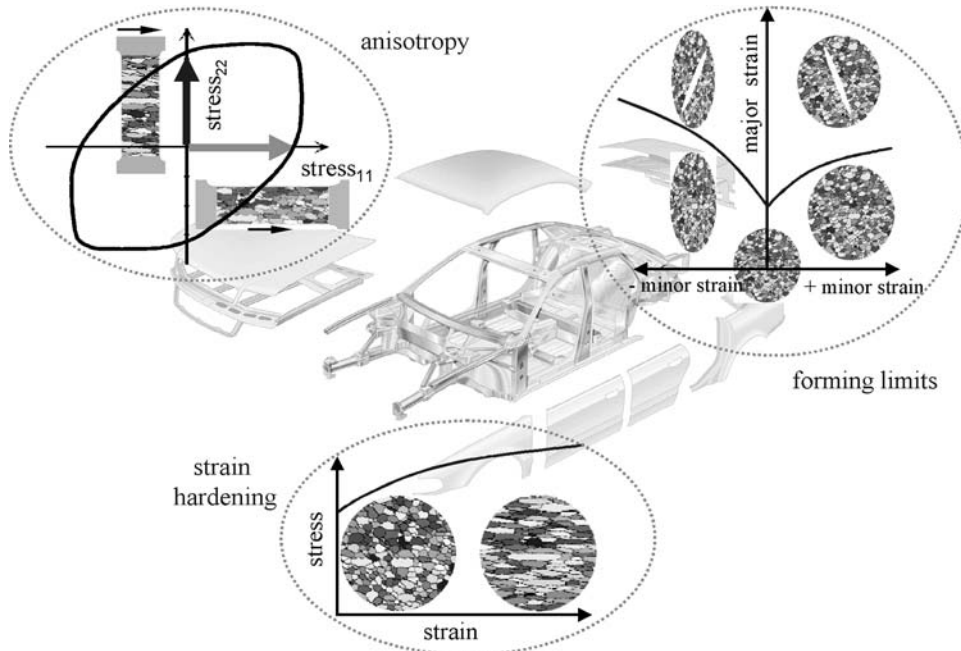
Figure 1   Modern finite element approaches for realistic metal-forming simulation typically require input data about strain hardening, forming limits, and anisotropy.
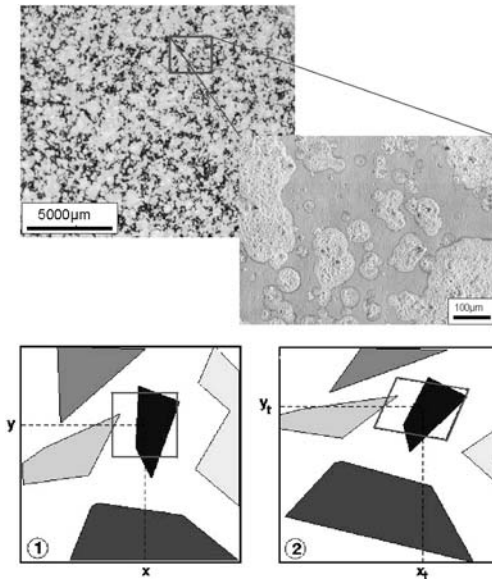
Figure 2 Plastic strain fields can be measured by photogrametry. This is a digital image analysis method which is based on the recognition of changes in the gray scale distribution of natural or artificial surface patterns after a straining step. (From Refs. 1 and 2.)

three-dimensional displacements at the surface, digital stereo pair images of the samples are acquired using two high-resolution CCD cameras. Pattern recognition is carried out by mapping a rectangular grid onto the image. The grid points are defined by three-dimensional coordinates and by the gray scale distribution in their proximity. After plastic straining, the distorted surface pattern is again determined. From the change in border coordinates containing the initial gray scale distribution around each grid point, the three-dimensional displacement gradient tensor field is determined with a lateral resolution given by the cameras and the grid point spacing. This pattern correlation method is based on the assumption that the gray scale distribution changes its spatial arrangement around a lattice point but remains otherwise constant during straining. The displacement gradient data are used to calculate the surface components of the strain tensor at each grid point in the form of the first-order approximation of the standard polar decomposition. The strain resolution is below 1% since the method uses the match of the complete gray scale distribution before and after loading as a measure to determine the exact shift in border coordinates. This procedure provides a larger precision than the determination of the new border coordinates from discrete pixels. For aluminum, a fine white color spray usually provides sufficient contrast for obtaining good gray scale pattern on a polished sample surface [1,2].

2.2 Experimental Determination of Nanotextures and Microtextures

Nano- and microtexture evolution in aluminum is discussed in terms of crystal orientation maps of deformed and recrystallized samples. Orientation mapping is

a technique for analyzing the two-dimensional topology of texture and grain boundaries in crystalline material. Local lattice orientations are measured on a regular grid by automated acquisition and processing of electron backscatter diffraction patterns (Kikuchi patterns) in a high-resolution scanning electron microscope (SEM). The microstructure can subsequently be reconstructed by coloring similar orientations on the measured grid with similar colors. In addition, the pattern quality of the Kikuchi patterns which is a measure for the local perfection of the crystal lattice is determined for each point [3–7]. In the present investigations, the lateral resolution of the microtexture measurement lies between 20 nm and 15 μm.

2.3 Constitutive Laws for Crystal Plasticity Finite Element Simulations

In the large-strain constitutive crystal plasticity model [8] modified for the present study [9,10], one assumes the stress response at each macroscopic continuum material point to be potentially given by one crystal or by a volume-averaged response of a set of grains comprising the respective material point. The latter assumption can be referred to as a local Taylor-type or local strain-rate homogenization assumption. In the case of a multi-grain description, the volume-averaged stress amounts to

$$\langle \mathbf{T} \rangle = \sum_{k=1}^{N} (\omega_k \mathbf{T}_k) \tag{1}$$

where N is the total number of individual orientations mapped onto an integration point using the Taylor assumption, ω_k the volume fraction of each single orientation, $\mathbf{T}_k$ the Cauchy stress produced by the kth individual orientation, and $\langle \mathbf{T} \rangle$ the volume-average stress produced by all orientations mapped at an integration point. The constitutive equation for the stress in each grain is then expressed in terms of

$$\mathbf{T}^* = \mathbf{C}\mathbf{E}^* \tag{2}$$

where $\mathbf{C}$ is the fourth-order elastic tensor, and $\mathbf{E}^*$ is an elastic strain measure obtained by polar decomposition

$$\mathbf{E}^* = \frac{1}{2}(\mathbf{F}^{*\mathbf{T}}\mathbf{F}^* - \mathbf{1}) \tag{3}$$

which leads to a stress measure which is the elastic work conjugate to the strain measure $\mathbf{E}^*$,

$$\mathbf{T}^* = \mathbf{F}^{*-1}(\det(\mathbf{F}^*)\mathbf{T})\mathbf{F}^{*-\mathbf{T}} \tag{4}$$

where $\mathbf{T}$ is the symmetric Cauchy stress tensor in the grain, and $\mathbf{F}^*$ is a local elastic deformation gradient defined in terms of the local *total* deformation gradient $\mathbf{F}$ and the local *plastic* deformation gradient $\mathbf{F}^\mathrm{p}$. The relation between the elastic and the plastic portion of $\mathbf{F}$ amounts to

$$\mathbf{F}^* = \mathbf{F}\mathbf{F}^{\mathrm{p}^{-1}}, \quad \det(\mathbf{F}^*) > 0, \quad \det(\mathbf{F}^{\mathrm{p}}) = 1 \tag{5}$$

The plastic deformation gradient is given by the flow rule

$$\dot{\mathbf{F}}^{p} = \mathbf{L}^{p}\mathbf{F}^{p} \tag{6}$$

with its crystalline portion

$$\mathbf{L}^{p} = \sum_{k=1}^{N} \dot{\gamma}_{k}m_{k}, \quad m_{k} = \hat{b}_{k} \otimes \hat{n}_{k} \tag{7}$$

where m_k are the kth dyadic slip products of unit vectors $\hat{b}_k$ in the slip direction and $\hat{n}_k$ normal to the slip plane, and $\dot{\gamma}_k$ are the shear rates on these systems. The specific constitutive functions for the plastic shearing rates $\dot{\gamma}_k$ on the slip systems are taken as

$$\dot{\gamma}_{k} = \dot{\gamma}_{0} \left| \frac{\tau_{k}}{\tau_{k,\mathrm{crit}}} \right|^{1/m} \mathrm{sign}(\tau_{k}) \tag{8}$$

where τ_k is the resolved shear stress for the slip system k, and $\tau_{k,\mathrm{crit}}$ is the actual critical shear stress on the kth slip system. $\dot{\gamma}_0$ and m are material parameters representing shearing rate and the rate sensitivity of slip. $\tau_{k,\mathrm{crit}}$ can be calculated by accounting for latent hardening through the use of a hardening matrix,

$$\dot{\tau}_{k,\mathrm{crit}} = \sum_{i} h^{ki} |\dot{\gamma}^{i}|, \quad h^{ki} = q^{ki} h^{(i)} \tag{9}$$

where h^{ki} is the rate of strain hardening on kth slip system due to a shearing on ith slip system, q^{ki} the hardening matrix describing the latent hardening behavior of a crystallite, and $h^{(i)}$ is the hardening rate of a single-slip system i. In all simulations presented in this chapter 12 slip systems with crystallographic $\langle 1\,1\,0 \rangle$ slip directions and $\{1\,1\,1\}$ slip planes are used for the plastic deformation of aluminum. The matrix h^{ki} can be taken as

$$h^{ki} = \begin{bmatrix} A & qA & qA & qA \\ qA & A & qA & qA \\ qA & qA & A & qA \\ qA & qA & qA & A \end{bmatrix} \tag{10}$$

where q is the ratio of the latent hardening rate to the self-hardening rate, and A is a 3×3 matrix populated by ones. Using this constitutive description renders the finite element method an elegant tool for detailed joint simulation studies of texture and strain evolution under realistic boundary conditions.

2.4 Constitutive Laws of Cellular Automata for Recrystallization Simulations

The recrystallization simulations discussed in this chapter are based on a cellular automaton formulation with a probabilistic switching rule [11–15]. Independent variables are time t and space $\mathbf{x} = (x_1, x_2, x_3)$. Space is discretized into an array of equally shaped quadratic cells. Each cell is characterized in terms of the dependent variables. These are scalar (mechanical, electromagnetic) and con-figurational (interfacial) contributions to the driving force and the crystal orientation

$\mathbf{g} = \mathbf{g}(\varphi_1, \phi, \varphi_2)$, where $\mathbf{g}$ is the rotation matrix, and φ_1, ϕ, φ_2 are the Euler angles. The driving force is the negative change in Gibbs enthalpy G_t per transformed cell. The starting data, i.e. the crystal orientation map and the spatial distribution of the driving force, can be provided by orientation imaging microscopy or by a crystal plasticity finite element simulation. Grains or subgrains are mapped as regions of identical crystal orientation, but the driving force may vary inside these areas.

The kinetics of cellular automata result from changes in the state of the cells (cell switches). They occur in accord with a transformation rule which determines the individual switching probability of each cell as a function of its previous state and the state of its neighbor cells [16–19]. The switching rule used in the simulations discussed in this chapter is designed for the simulation of primary static recrystallization. It reflects that the state of a non-recrystallized cell belonging to a deformed grain may change due to the expansion of a recrystallizing neighbor grain which grows according to the local driving force and boundary mobility. If such an expanding grain sweeps a non-recrystallized cell, the stored dislocation energy of that cell drops to 0 and a new orientation is assigned to it, namely that of the growing neighbor grain. To put this formally, the switching rule is cast in the form of a probabilistic analog of the linearized symmetric Turnbull rate equation which describes grain boundary motion in terms of isotropic single-atom diffusion processes perpendicular through a homogeneous planar grain boundary segment under the influence of a drop in Gibbs energy [20–22],

$$\dot{\mathbf{x}} = \mathbf{n}\nu_D \lambda_{gb} c \left\{ \exp\left(-\frac{\Delta G + \Delta G_t/2}{k_B T}\right) - \exp\left(-\frac{\Delta G - \Delta G_t/2}{k_B T}\right) \right\} \tag{11}$$

where $\dot{\mathbf{x}}$ is the grain boundary velocity, ν_D the Debye frequency, λ_{gb} the jump width through the boundary, c the intrinsic concentration of grain boundary vacancies or shuffle sources, $\mathbf{n}$ the normal of the grain boundary segment, ΔG the Gibbs enthalpy of motion through in the interface, ΔG_t the Gibbs enthalpy associated with the transformation, k_B the Boltzmann constant, and T is the absolute temperature. Replacing the jump width by the burgers vector and the Gibbs enthalpy terms by the total entropy, ΔS, and total enthalpy, ΔH, leads to the linearized form

$$\dot{\mathbf{x}} \approx \mathbf{n}\nu_D b \exp\left(-\frac{\Delta S}{k_B}\right)\left(\frac{pV}{k_B T}\right) \exp\left(-\frac{\Delta H}{k_B T}\right) \tag{12}$$

where p is the driving force, and V the atomic volume which is of the order of b^3, where b is the magnitude of the Burgers vector. Summarizing these terms reproduces Turnbull's expression

$$\dot{\mathbf{x}} = \mathbf{n}mp = \mathbf{n}m_0 \exp\left(-\frac{Q_{gb}}{k_B T}\right) p \tag{13}$$

where m is the mobility. The equations provide a well-known kinetic picture of grain boundary segment motion, where the atomistic processes are statistically described in terms of the pre-exponential factor of the mobility $m_0 = m_0(\Delta\mathbf{g},\mathbf{n})$ and the activation energy of grain boundary mobility $Q_{gb} = Q_{gb}(\Delta\mathbf{g},\mathbf{n})$. For dealing with competing switches affecting the same cell, the deterministic rate equation, Eq. (13), can be replaced by a probabilistic analog which allows one to calculate switching

probabilities. First, it is separated into a deterministic part, $\dot{x}_0$, which depends weakly on temperature, and a probabilistic part, w, which depends strongly on temperature:

$$\dot{\mathbf{x}} = \dot{\mathbf{x}}_0 w = \mathbf{n} \frac{k_B T m_0}{V} \frac{pV}{k_B T} \exp\left(-\frac{Q_{gb}}{k_B T}\right)$$

$$\text{with } \dot{\mathbf{x}}_0 = \mathbf{n} \frac{k_B T m_0}{V}, \quad w = \frac{pV}{k_B T} \exp\left(-\frac{Q_{gb}}{k_B T}\right) \tag{14}$$

The probability factor w represents the product of the linearized part $pV/(k_B T)$ and the non-linearized part $\exp(-Q_{gb}/(k_B T))$ of the original Boltzmann terms. According to Eq. (14), non-vanishing switching probabilities occur for cells which reveal neighbors with different orientation and a driving force which points in their direction. The automaton considers the first, second (2D), and third (3D) neighbor shells for the calculation of the total driving force acting on a cell. The local value of the switching probability depends on the crystallographic character of the boundary segment between such unlike cells.

The cellular automaton is usually applied to starting data which have a spatial resolution far above the atomic scale. This means that the automaton grid has some mesh size $\lambda_m \gg b$. If a moving boundary segment sweeps a cell, the grain, thus, grows (or shrinks) by λ_m^3 rather than b^3. Since the net velocity of a boundary segment must be independent of the imposed value of λ_m, an increase of the jump width must lead to a corresponding decrease of the grid attack frequency, i.e. to an increase of the characteristic time step, and vice versa. For obtaining a scale-independent grain boundary velocity, the grid frequency must be chosen in a way to ensure that the attempted switch of a cell of length λ_m occurs with a frequency much below the atomic attack frequency which attempts to switch a cell of length b. Mapping Eq. (14) on a grid which has a scaling length λ_m leads to

$$\dot{\mathbf{x}} = \dot{\mathbf{x}}_0 w = \mathbf{n}(\lambda_m v) w \quad \text{with } v = \frac{k_B T m_0}{V \lambda_m} \tag{15}$$

where v is the eigenfrequency of the chosen mesh characterized by the scaling length λ_m. The eigenfrequency given by this equation represents the attack frequency for *one* particular grain boundary with constant mobility. In order to use a whole *spectrum* of mobilities and driving forces in one simulation, it is necessary to normalize the equation by a common grid attack frequency v_0 rendering it into

$$\dot{\mathbf{x}} = \dot{\mathbf{x}}_0 w = \mathbf{n} \lambda_m v_0 \left(\frac{v}{v_0}\right) w = \hat{\dot{\mathbf{x}}}_0 \left(\frac{v}{v_0}\right) w = \hat{\dot{\mathbf{x}}}_0 \hat{w} \tag{16}$$

where the normalized switching probability amounts to

$$\hat{w} = \left(\frac{v}{v_0}\right) \frac{pV}{k_B T} \exp\left(-\frac{Q_{gb}}{k_B T}\right) = \frac{m_0 p}{\lambda_m v_0} \exp\left(-\frac{Q_{gb}}{k_B T}\right) \tag{17}$$

The value of the normalization (grid attack) frequency v_0 can be identified by using the assumption that the maximum occurring switching probability cannot be larger than 1

$$\hat{w}^{max} = \frac{m_0^{max} p^{max}}{\lambda_m v_0^{min}} \exp\left(-\frac{Q_{gb}^{min}}{k_B T}\right) \overset{!}{\leq} 1 \tag{18}$$

where m_0^{max} is the maximum occurring pre-exponential factor of the mobility, p^{max} the maximum possible driving force, v_0^{min} the minimum allowed grid attack frequency, and $Q_{\mathrm{gb}}^{\mathrm{min}}$ the minimum occurring activation energy. Assuming $\hat{w}^{\mathrm{max}} = 1$, one obtains the normalization frequency as a function of the upper bound input data.

$$v_0^{\mathrm{min}} = \frac{m_0^{\mathrm{max}} p^{\mathrm{max}}}{\lambda_{\mathrm{m}}} \; \exp\left(-\frac{Q_{\mathrm{gb}}^{\mathrm{min}}}{k_{\mathrm{B}} T}\right) \tag{19}$$

This frequency and the local values of the mobility and the driving force change the switching equation into

$$\begin{aligned}
\hat{w}^{\mathrm{local}} &= \frac{m_0^{\mathrm{local}} p^{\mathrm{local}}}{\lambda_{\mathrm{m}} v_0^{\mathrm{min}}} \exp\left(-\frac{Q_{\mathrm{gb}}^{\mathrm{local}}}{k_{\mathrm{B}} T}\right) \\
&= \left(\frac{m_0^{\mathrm{local}}}{m_0^{\mathrm{max}}}\right)\left(\frac{p^{\mathrm{local}}}{p^{\mathrm{max}}}\right) \exp\left(-\frac{Q_{\mathrm{gb}}^{\mathrm{local}} - Q_{\mathrm{gb}}^{\mathrm{min}}}{k_{\mathrm{B}} T}\right) = \left(\frac{m^{\mathrm{local}} p^{\mathrm{local}}}{m^{\mathrm{max}} p^{\mathrm{max}}}\right)
\end{aligned} \tag{20}$$

This expression is the central switching equation of the cellular automaton algorithm. It shows that the local switching probability can be quantified by the ratio of the local and the maximum mobility $m^{\mathrm{local}}/m^{\mathrm{max}}$, which is a function of the grain boundary character and by the ratio of the local and the maximum driving pressure $p^{\mathrm{local}}/p^{\mathrm{max}}$. The probability of the fastest occurring boundary segment (characterized by $m_0^{\mathrm{local}} = m_0^{\mathrm{max}}$, $p^{\mathrm{local}} = p^{\mathrm{max}}$, and $Q_{\mathrm{gb}}^{\mathrm{local}} = Q_{\mathrm{gb}}^{\mathrm{min}}$) to realize a cell switch is equal to 1. The equation shows that the mesh size does not influence the switching probability but only the time step elapsing during an attempted switch. The characteristic time constant of the simulation Δt is $1/v_0^{\mathrm{min}}$. The switching probability can also be formulated in terms of the local time $t = \lambda_{\mathrm{m}}/\dot{x}$ required by a grain boundary with velocity $\dot{x}$ to cross the automaton cell of size λ_{m}

$$\hat{w}^{\mathrm{local}} = \left(\frac{m^{\mathrm{local}} p^{\mathrm{local}}}{m^{\mathrm{max}} p^{\mathrm{max}}}\right) = \left(\frac{\dot{x}^{\mathrm{local}}}{\dot{x}^{\mathrm{max}}}\right) = \left(\frac{t^{\mathrm{max}}}{t^{\mathrm{local}}}\right) \tag{21}$$

Therefore, the local switching probability can also be regarded as the ratio of the distances that were swept by the local grain boundary and the grain boundary with maximum velocity, or as the number of time steps the local grain boundary needs to wait before crossing the encountered cell. This reformulates the same underlying problem, namely, that boundaries with different mobilities and driving forces cannot equally switch the state of the automaton within one common time step.

The above equations allow one to calculate the switching *probability* of a cell as a function of its previous state and the state of the neighbor cells. The actual *decision* about a switching event is made by a Monte Carlo step. The use of random sampling ensures that all cells are switched according to their statistical weight, i.e. according to the local driving force and grain boundary mobility between abutting cells. The simulation proceeds by first calculating the individual local switching probabilities $\hat{w}^{\mathrm{local}}$ and then sampling them using a Monte Carlo algorithm. This means that for each cell the calculated switching probability is compared to a randomly generated number r which lies between 0 and 1. The switch is accepted if the random number is

equal or smaller than the calculated switching probability. Otherwise, the switch is rejected

$$\text{random number } r \text{ between 0 and 1} \begin{cases} \text{accept switch if } r \leq \left(\frac{m^{\text{local}}p^{\text{local}}}{m^{\text{max}}p^{\text{max}}}\right) \\ \text{reject switch if } r > \left(\frac{m^{\text{local}}p^{\text{local}}}{m^{\text{max}}p^{\text{max}}}\right) \end{cases} \quad (22)$$

Except for the probabilistic evaluation of the analytically calculated transformation probabilities, the approach is entirely deterministic. The use of realistic or experimental input data for the grain boundaries enables one to make predictions on a real time and space scale. The switching rule is scalable to any mesh and to any spectrum of boundary mobility and driving force data. The state update of all cells is made in synchrony. Related cellular automaton models which are suited to tackle recrystallization of aluminum alloys were also suggested by other authors [23–28].

3 MICROSCOPIC ASPECTS OF TEXTURE EVOLUTION DURING PLASTIC DEFORMATION AND RECRYSTALLIZATION

3.1 Bicrystal Deformation

Figure 3 shows a bicrystal which was plane strain deformed (30% engineering thickness reduction) in a channel die set-up using Teflon foil as a lubricant. The tilt angle grain boundary separating the two crystals amounted to 7° before deformation and to 11° after deformation. The figure shows the experimentally determined distribution of the von Mises strain in the two abutting crystals together with a corresponding crystal plasticity finite element simulation. Although the simulation does not exactly match the experiments, both data sets clearly show that the two crystals do not deform as a micromechanical entity – as anticipated due to the small angle of the grain boundary between them – but deform as two mechanically individual grains. An additional experiment (Fig. 4) conducted on a second bicrystal with a large angle grain boundary (34.8° tilt boundary) between the two grains indeed shows a very similar behavior as the deformation experiment conducted on the bicrystal with the small angle grain boundary. The surprising similarity between the micromechanical influence exerted by a small angle grain boundary and that exerted by a large angle grain boundary shows that – although similar slip systems are active on either side of the small angle boundary – the grain boundary clearly impedes slip continuity and thereby generates discontinuities in the accumulated plastic strain. Also, the strong localization effects occurring at the open surfaces along the extension direction of the sample (Fig. 3) indicate that the main influence of the small angle grain boundary could lie in acting as a source for generating instability rather than acting as a rigid border to the dislocation flux. It is well known from studies on in-grain deformation banding that tiny orientation fluctuations within one grain can create neighboring areas with different slip system selection (the so-called differently deforming regions). This principle might help to explain the strain heterogeneity observed in Fig. 3 across the grain boundary. It is conceivable that it is energetically more favorable for a bicrystal to deform by the activation of
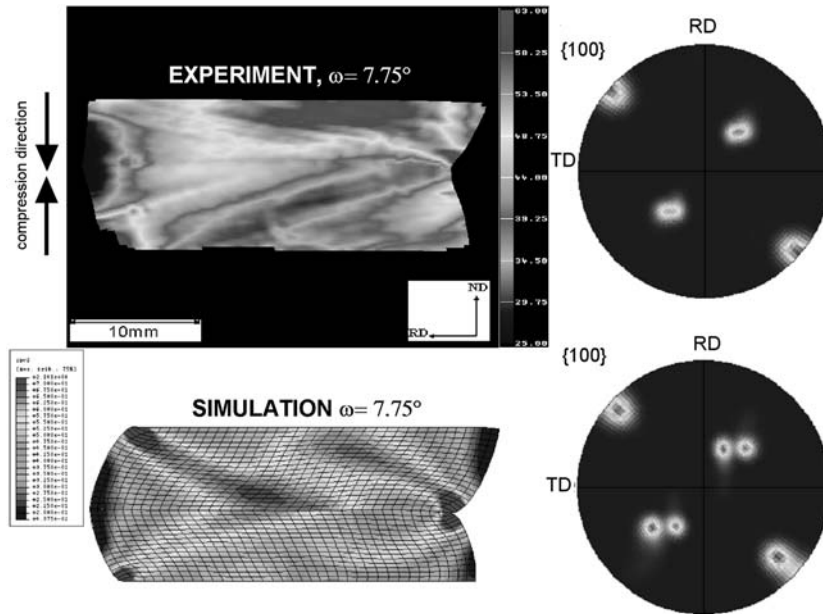
Figure 3 Measured and simulated spatial distribution of the von Mises strain in an aluminum bicrystal (small angle grain boundary, 7.75° tilt angle) which was plane strain deformed (30% engineering thickness reduction) in a channel die set-up. The right-hand side shows the measured and simulated {1 0 0} pole figures.

different single-slip systems in each crystal in order to follow the exerted loading rather than to symmetrically co-deform both crystals.

Figures 5 and 6 show a more detailed high-resolution analysis of the orientation changes on either side of the small angle grain boundary shown in Fig. 3 and of the large angle grain boundary shown in Fig. 4. In both the cases, it can be seen that the orientation change in front of the interface, expressing the stored content of geometrically necessary dislocations, is not symmetric. In other words, one grain reveals an increasing net orientation change when approaching the boundary and the other one does not.

3.2 Grain-Scale Heterogeneity of Polycrystal Deformation

In order to investigate strain heterogeneity during polycrystal deformation, an aluminum sample with a single layer of quasi-2D coarse grains was plastically deformed in a channel die plane strain set-up at ambient temperature and low strain rate. Lubrication was obtained by using a Teflon foil. The microtexture was determined by orientation mapping in the SEM. The distribution of the plastic microstrain at the sample surface was determined using photogrametry. The microtexture was mapped onto a finite element mesh. J2 based* and crystal plasticity

---

*J2 indicates the 2nd invariant of the stress deviator.
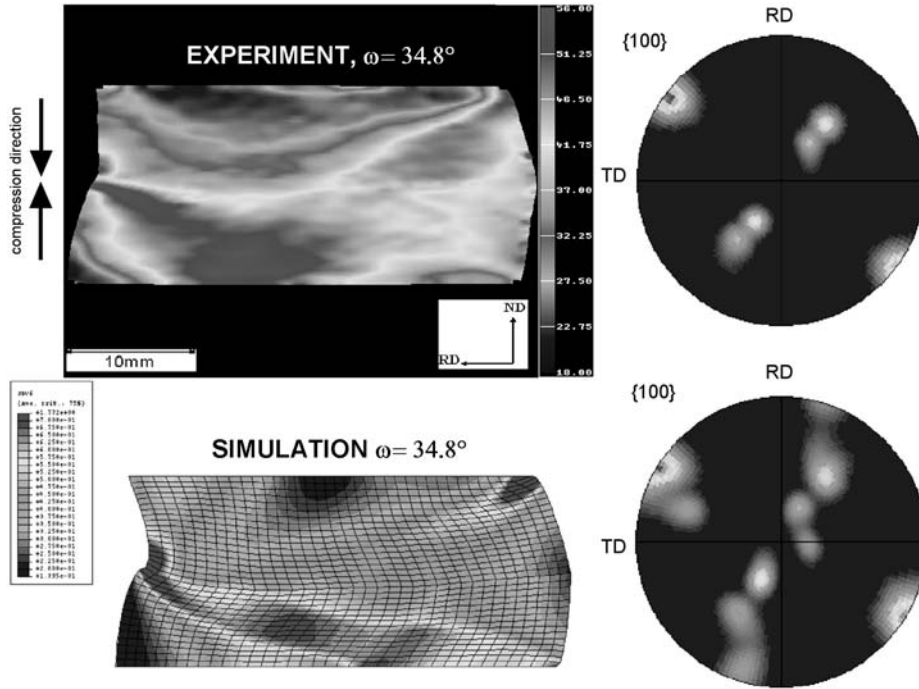
Figure 4 Measured and simulated spatial distribution of the von Mises strain in an aluminum bicrystal (high-angle grain boundary, 34.8° tilt angle) which was plane strain deformed (30% engineering thickness reduction) in a channel die set-up. The right-hand side shows the measured and simulated {1 0 0} pole figures.
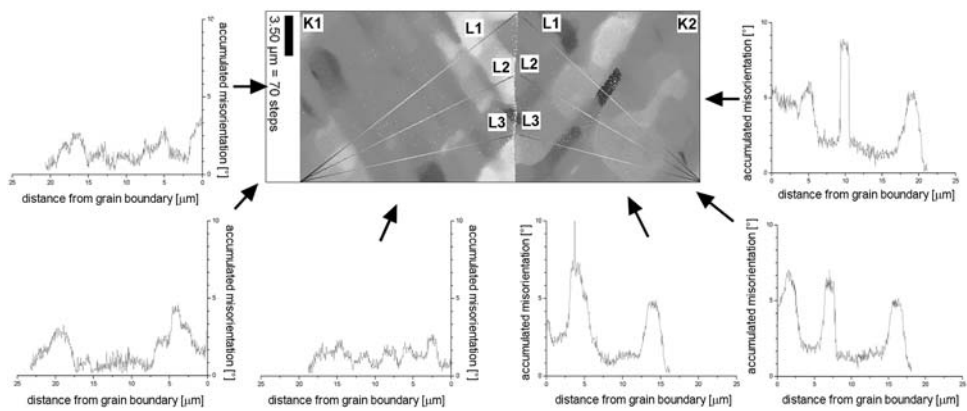
Figure 5 Measured orientation gradients on both sides of the small angle grain boundary (7.75° tilt angle) of the sample shown in Fig. 3. Note the orientation change in the crystal on the right-hand side indicating a substantial increase in the density of geometrically necessary dislocations.
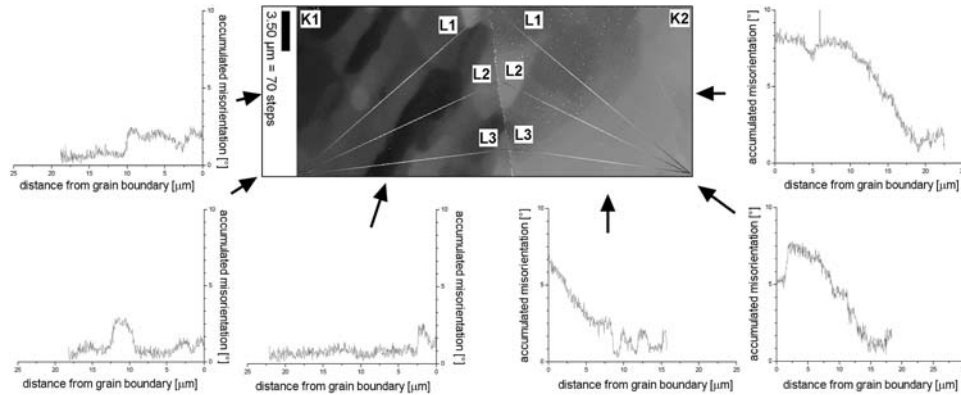
Figure 6 Measured orientation gradients on both sides of the high angle grain boundary (34.8° tilt angle) of the sample shown in Fig. 4. Note the orientation change in the crystal on the right-hand side indicating a substantial increase in the density of geometrically necessary dislocations.

finite element simulations were conducted using boundary conditions which approximated those of the channel die experiments [1,2].

Figure 7 shows the in-plane distribution of the accumulated plastic von Mises strain in the specimen after 3%, 8%, and 13% sample thickness reduction (given by $\Delta d/d \cdot 100\%$, where d is the sample thickness in compression direction). The grain boundaries indicated by lines were taken from microtexture data. The measurements reveal that some grains and grain portions show much higher accumulated plastic strain than others. For instance, in the 8% compressed sample, some of the grains have accumulated less than 1% von Mises strain while others show maximum strains beyond 15%, particularly close to the grain boundaries. Related to the total compressive strain of 8%, this establishes an enormous strain heterogeneity with deviations of nearly $\pm 90\%$ from the average value [1,2]. The phenomenon can be qualitatively interpreted in terms of the different geometrical hardness (orientation factor) of the grains. However, the observed lateral strain variation strongly exceeds the spread expected from the orientation factors. At larger strains, it was found that grain-scale strain hardening gradually equilibrates initial hardness differences among the grains, i.e. grain-to-grain strain heterogeneity is less pronounced at larger loadings. These observations have important consequences for the interpretation of data obtained by constitutive modeling, recrystallization, and electron microscopy. The strong heterogeneity observed suggests that local phenomena must be interpreted in terms of the *local* rather than the *global* strain state. The experimental data shown in Fig. 7 were analyzed using continuum simulations (J2, crystal plasticity; Fig. 8). The calculations reveal two important points. First, the macroscopic plastic strain path is not completely altered by the crystallographic texture, but *modulated* following soft crystals and avoiding hard crystals. Second, grain-scale mechanisms are strongly superimposed by effects arising from the macroscopic profile of strain [1,2]. Figure 9 shows some strain and texture details at a triple point between two soft and a hard grain. The nanotexture maps show the orientation
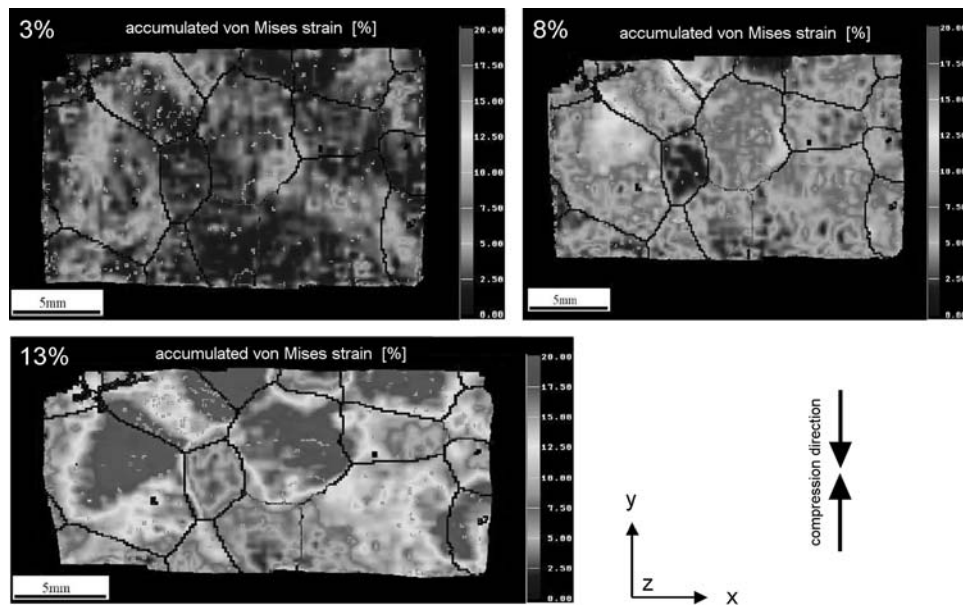
Figure 7   Distribution of the accumulated von Mises strain in a quasi-2D polycrystalline aluminum specimen after 3%, 8%, and 13% sample thickness reduction ($\Delta d/d$, where d is the sample extension along compression direction). The strain was determined using photogrametry. The grain boundaries indicated by black lines were taken from microtexture measurements (EBSD in an SEM).

changes between the grain interior and the triple point in the form of gray scales. It can be seen that the hard grain does not reveal substantial in-grain orientation changes between its interior and the triple point. The two soft grains show strong orientation changes when approaching the triple point. This detail suggests that triple points can be of importance at the early stages of strain localization and orientational in-grain subdivision.
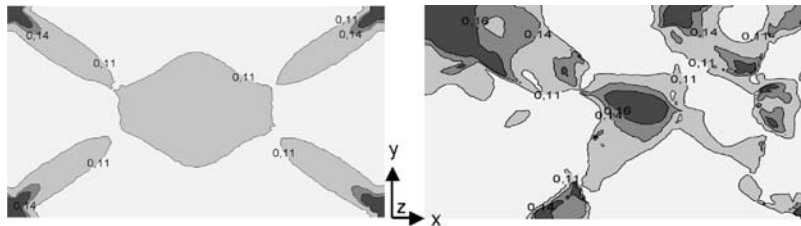
Figure 8   Simulated strain distribution in the 8% deformed polycrystalline aluminum sample shown in Fig. 7; left: J2-continuum plasticity; right: crystal continuum plasticity. The finite element calculations were conducted using plane strain boundary conditions and a friction coefficient of $\mu = 0.2$. The constitutive crystal plasticity law was implemented using 12 $\{1\,\underline{1}\,1\}\langle1\,1\,0\rangle$ slip systems and viscoplastic hardening. The details are given in [1,2].
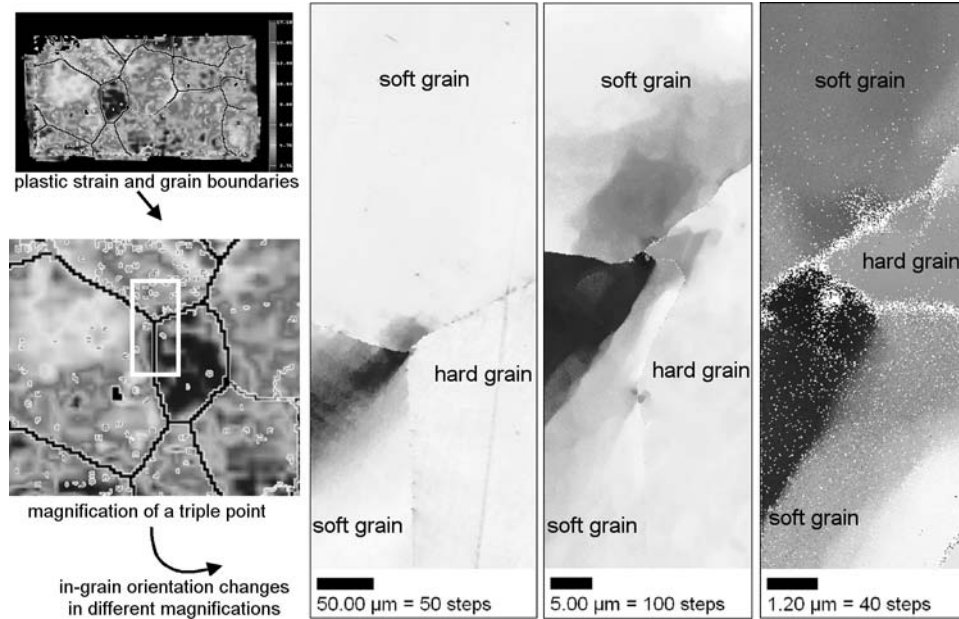
Figure 9 Details of the strain (left) and texture (right) distribution at a triple point between two soft and a hard grain. The three nanotexture maps (right) show the orientation changes between the grain interior far away from the triple point and the triple point in the form of gray scales. Black color indicates orientation changes of 15° between grain interior and the triple point. It can be seen that the hard grain does not reveal substantial in-grain orientation changes between its interior and the triple point. The two soft grains show strong orientation changes when approaching the triple point. This detail suggests that triple points can be of importance at the early stages of strain localization and orientational in-grain subdivision.

3.3   Plastic Grain Interaction and In-Grain Subdivision During Large-Strain Polycrystal Deformation

This section is about experiments and crystal plasticity simulations (plane strain) which address the dependence of in-grain orientation subdivision and deformation textures in aluminum polycrystals on grain interaction [1,2,29–35]. The finite element simulations are conducted by statistically varying the spatial arrangement of the grains in a polycrystal so that they have different neighbor grains in each simulation (Fig. 10). Each grain contains eight integration points. The orientation changes of the eight integration points in each grain are sampled for the different polycrystal arrangements. The influence of grain neighborhood on in-grain subdivision and texture is quantified by the use of a mean orientation concept which allows one to calculate the orientation spread among the eight originally identical in-grain orientation points after plastic straining for each crystal (Fig. 11).

An aluminum sample with commercial purity (99.9 mass% Al) was chosen for comparing the finite element predictions with experimental results. For preparing the sample, an aluminum block was forged in three mutually perpendicular directions several times with gradually decreasing thickness reductions. This procedure was conducted in order to produce fine grains and a random texture. After forging, the
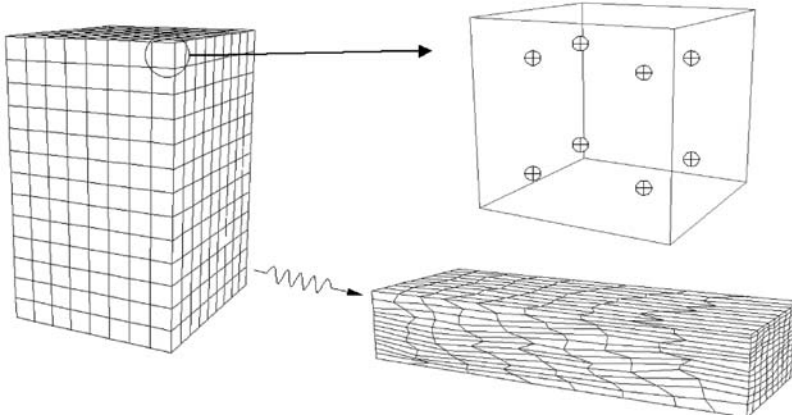
Figure 10   Finite element model set-up for evaluating the influence of neighboring grains on the deformation texture.

sample was cold rolled to 95% thickness reduction (engineering strain $\Delta d/d$, where d is the sheet thickness before rolling). The textures were examined by measuring four incomplete pole figures in Bragg back-reflection mode. From the experimental pole figures, the orientation distribution function was derived by the use of the series expansion method and subsequently ghost corrected using spherical Gauss functions (Fig. 11).

The study reveals five important points about the relationship between texture and grain interaction. First, the grain neighborhood has a significant influence on the
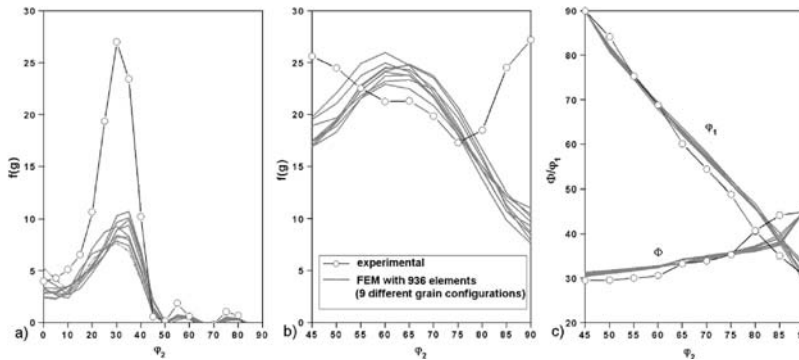
Figure 11   Texture results obtained from nine crystal plasticity finite element simulations conducted by statistically varying the arrangement of the grains in a polycrystal. Each grain is represented by eight integration points and has different neighbor grains in each of the nine simulations. There are only small differences among the nine calculations which show that details of the grain neighborhood can be regarded as a scalable statistical factor for texture evolution: (a) α-fiber including major components $\{0\,1\,1\}\langle 1\,0\,0\rangle$ (Goss-component, $\varphi_1 = 0°$, $\phi = 45°$, $\varphi_2 = 0°$), $\{0\,\underline{1}\,1\}\langle 2\,1\,1\rangle$ (brass-component, $\varphi_1 = 35°$, $\phi = 45°$, $\varphi_2 = 0°$), $\{0\,\underline{1}\,1\}\langle 1\,1\,1\rangle$, and $\{0\,\underline{1}\,1\}\langle 0\,1\,1\rangle$ (90° about the normal rotated Goss-component, $\varphi_1 = 90°$, $\phi = 45°$, $\varphi_2 = 0°$); (b) β-skeleton line including major components $\{\underline{2}\,1\,1\}\langle 1\,1\,1\rangle$ (copper-component, $\varphi_1 = 90°$, $\phi = 35°$, $\varphi_2 = 45°$), $\sim\{1\,2\,\underline{3}\}\langle 6\,3\,4\rangle$ (S-component, $\varphi_1 = 60°$, $\phi = 32°$, $\varphi_2 = 65°$), and the brass-component $\{0\,1\,1\}\langle 2\,\underline{1}\,1\rangle$; and (c) coordinates of β-skeleton line.

reorientation of a grain (up to 20% in terms of its end orientation and its orientation density), but its own initial orientation is more important for its reorientation behavior than the neighborhood. Second, the sharpness of the deformation texture is affected by grain interaction leading to an overall weaker texture when compared to predictions obtained without interaction. Third, the in-grain subdivision of formerly homogeneous grains occurring during straining is strongly dependent on their initial orientation (Fig. 12). For instance, some crystals build up in-grain orientation changes of more than 20° after 95% straining while others do practically not subdivide (Fig. 13). Fourth, the dependence of in-grain subdivision on the neighbor grains is different for crystals with different initial orientation (cube or rotated Goss grains reveal strong subdivision). Fifth, the upper bound for the variation of texture due to changes in grain neighborhood amounts at most to 5% in terms of the positions of the main texture components. In terms of the overall orientation density, all predictions (using different neighborhood configurations) remain within a narrow tube with an orientation scatter of 10% (β-fiber) to 20% (brass component, α-fiber) when the neighborhood changes [1,2,29,35,36].

3.4 Simulation of Primary Static Recrystallization and Comparison to Avrami Kinetics

Figure 14 shows the simulated kinetics and microstructures of a recrystallizing aluminum single crystal. The initial deformed crystal had a uniform Goss orientation (0 1 1)[1 0 0] and a dislocation density of 10^{15} m^{-2}. The driving force was due to the
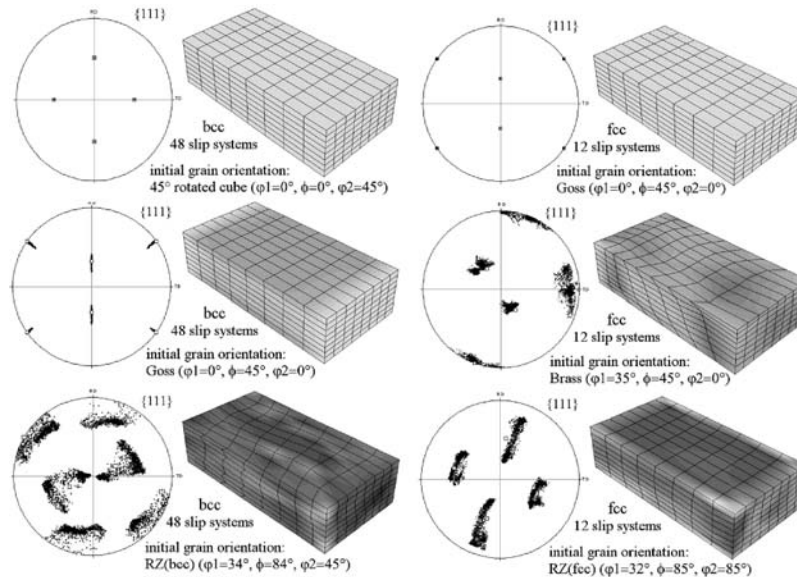
Figure 12   Accumulated in-grain crystalline misorientations in gray scale coding; 50% plane strain thickness reduction of initially uniformly oriented single grains; {111} pole figures; open squares show initial orientations; black dots show orientations of all integration points after deformation; left: body-centered crystal structure using 48 slip systems; right: face-centered crystal structure using 12 slip systems.
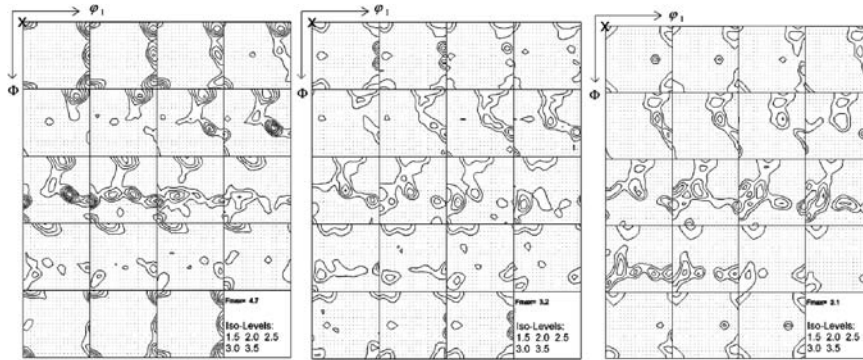
Figure 13 In-grain misorientation functions between a mean orientation and the orientations of eight integration points initially pertaining to the same grain (same finite element) for three crystal plasticity finite element simulations with different grain neighborhood. By using a Gauss approach, all misorientation results for the 936 different elements were mapped into Euler space according to their *initial* orientation position. The misorientation functions are normalized so that the contour lines indicate not the absolute but the relative tendency of grains to build up orientation scatter. Most orientations show a small tendency to undergo strong orientation subdivision. Some orientations (e.g. cube) generally reveal a strong tendency to subdivide, irrespective of neighborhood details. Some components (e.g. 90° rotated Goss) reveal a strong dependence in their subdivision behavior on neighborhood details. All results are for face-centered crystal structure with 12 slip systems.

stored elastic energy provided by the dislocations. In order to compare the predictions with analytical Avrami kinetics recovery and driving forces arising from local boundary curvature were not considered. The simulation used site-saturated nucleation conditions, i.e. the nuclei were at $t = 0$ sec statistically distributed in physical space and orientation space. The grid size was $10 \times 10 \times 10 \, \mu m^3$. The cell size was $0.1 \, \mu m$. All grain boundaries had the same mobility using an activation energy of the grain boundary mobility of $1.3 \, eV$ and a pre-exponential factor of the boundary mobility of $m_0 = 6.2 \times 10^{-6} \, m^3/(N \cdot s)$. Small angle grain boundaries had a mobility of 0. The temperature was $800 \, K$. The time constant of the simulation was $0.35 \, sec$.

Figure 15 shows the kinetics for a number of 3D recrystallization simulations with site-saturated nucleation conditions and identical mobility for all grain boundaries. The different curves correspond to different initial numbers of nuclei. The initial number of nuclei varied between 9624 (nucleation energy of $3.2 \, eV$) and 165 (nucleation energy of $6.0 \, eV$). The curves (Fig. 15a) all show a typical Avrami shape. The logarithmic plots (Fig. 15b) reveal Avrami exponents between 2.86 and 3.13 which is in good accord with the analytical value of 3.0 for site-saturated conditions. The simulations with a very high initial density of nuclei reveal a more pronounced deviation of the Avrami exponent with values around 2.7 during the beginning of recrystallization. This deviation from the analytical behavior is due to lattice effects. While the analytical derivation assumes a vanishing volume for newly formed nuclei, the cellular automaton has to assign one lattice point to each new nucleus. Figure 16 shows the effect of grain boundary mobility on growth selection.
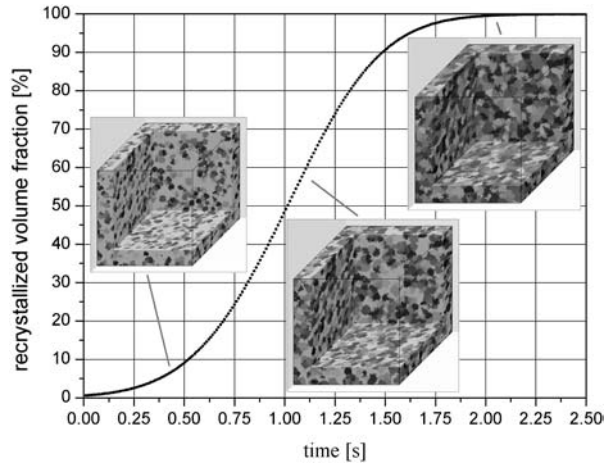
Figure 14 Simulated kinetics and microstructure of recrystallization in a plastically strained aluminum single crystal (cellular automaton). The deformed crystal had a (0 1 1)[1 0 0] orientation and a uniform dislocation density of $10^{15}\,\mathrm{m}^{-2}$. Simulation parameter: site-saturated nucleation, lattice size: $10 \times 10 \times 10\,\mu\mathrm{m}^3$, cell size: $0.1\,\mu\mathrm{m}$, activation energy of large angle grain boundary mobility: $1.3\,\mathrm{eV}$, pre-exponential factor of large angle boundary mobility: $m_0 = 6.2 \times 10^{-6}\,\mathrm{m}^3/(\mathrm{N\,s})$, temperature: $800\,\mathrm{K}$, time constant: $0.35\,\mathrm{sec}$.

While in Fig. 16a all boundaries have the same mobility, in Fig. 16b, one grain boundary has a larger mobility (activation energy of the mobility of $1.35\,\mathrm{eV}$ instead of $1.40\,\mathrm{eV}$) and, consequently, grows faster than the neighboring grains which finally cease to grow. The grains in this simulation all grow into a heavily deformed single crystal [12–15].
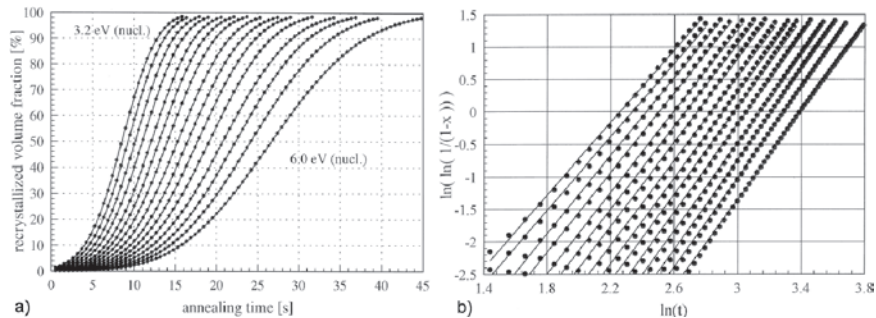
Figure 15 Simulated kinetics for various 3D recrystallization simulations with site-saturated nucleation conditions and identical mobility for all grain boundaries (cellular automaton). The different curves correspond to different initial numbers of nuclei. The initial number of nuclei varied between 9624 (nucleation energy of $3.2\,\mathrm{eV}$ and 165 (nucleation energy of $6.0\,\mathrm{eV}$); (a) Avrami diagrams; (b) logarithmic diagrams showing Avrami exponents between 2.86 and 3.13.
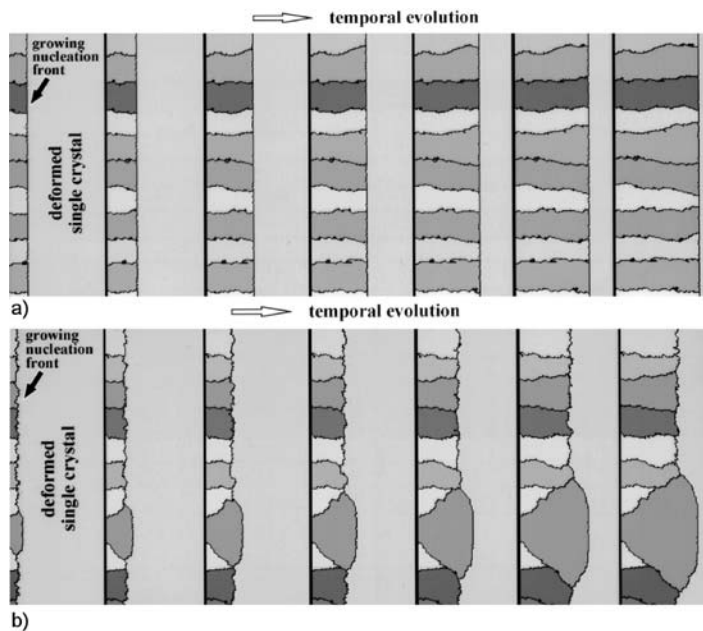
Figure 16 Effect of grain boundary mobility on growth selection. All grains grow into a deformed single crystal: (a) all grain boundaries have the same mobility; (b) one grain boundary has a larger mobility than the others (activation energy of the mobility of 1.35 eV instead of 1.40 eV) and grows faster than the neighboring grains (cellular automaton).

3.5 Examples of Coupling Cellular Automata with Crystal Plasticity Finite Element Models for Predicting Recrystallization

3.5.1 Motivation for Coupling Texture Simulation Methods

Simulation approaches such as the crystal plasticity finite element method or cellular automata increasingly gain momentum as tools for the discrete prediction of microstructure and texture. The major advantage of such approaches is that they consider material *heterogeneity* as opposed to statistical approaches which assume material *homogeneity*. Although the average behavior of materials after deformation and heat treatment can sometimes be well described without accounting for local effects, many examples exist where material response can only be predicted when considering material heterogeneity. For instance, in the field of plasticity, the quantitative investigation of ridging and roping or related surface defects observed in sheet metals requires knowledge about local texture effects. In the field of heat treatment, the origin of the Goss texture in transformer steels, and the incipient stages of cube texture formation during recrystallization can hardly be predicted without knowing local characteristics of the material. Although spatially discrete microstructure simulations have already profoundly enhanced our understanding of microstructure and texture evolution over the last decade, their potential is sometimes simply limited by an insufficient knowledge about the external boundary conditions which characterize the process and an insufficient knowledge about the

internal starting conditions which are inherited from the preceding process steps. It is, thus, an important goal to improve the incorporation of both types of information into such simulations. External boundary conditions prescribed by real industrial processes are often spatially non-homogeneous. They can be investigated using experiments or process simulations which consider spatial resolution. Spatial heterogeneities in the internal starting conditions, i.e. in the microstructure and texture, can be obtained from experiments or microstructure simulations which include spatial resolution.

3.5.2 Coupling, Scaling, Nucleation, and Boundary Conditions

In the present example, the results obtained from a crystal plasticity finite element simulation are used to map a starting microstructure and texture for a subsequent recrystallization simulation. The finite element model is used to simulate a plane strain compression test conducted on aluminum with columnar grain structure to a total logarithmic strain of $\varepsilon = -0.434$. The values of the state variables (dislocation density, crystal orientation) given at the integration points of the finite element mesh are mapped on the regular lattice of a 2D cellular automaton. In the present example, the original size of the specimen which provides the input microstructure to the crystal plasticity finite element simulations give a lattice point spacing of $\lambda_m = 61.9\,\mu m$. The maximum driving force in the region arising from the stored dislocation density is about 1 MPa. The annealing temperature is 800 K. Large angle grain boundaries are characterized by an activation energy for the mobility of 1.3 eV. Small angle grain boundaries are assumed to be immobile.

The nucleation process during primary static recrystallization has been explained for pure aluminum in terms of discontinuous subgrain growth [22]. According to this model, nucleation takes place in areas which reveal high misorientations among neighboring subgrains and a high local driving force for curvature-driven discontinuous subgrain coarsening. The present simulation works above the subgrain scale, i.e. it does not explicitly describe subgrain coarsening kinetics. Instead, it incorporates nucleation on a phenomenological basis using the kinetic and thermodynamic instability criteria known from the classical recrystallization theory [22]. The kinetic instability criterion means that a successful nucleation process leads to the formation of a mobile large angle grain boundary which can sweep the surrounding deformed matrix. The thermodynamic instability criterion means that the stored energy changes across the newly formed large angle grain boundary providing a net driving force pushing it forward into deformed areas. Nucleation in this simulation is performed in accord with these two aspects, i.e. potential nucleation sites must fulfill both the instability criteria. This nucleation model does not create new orientations. At the beginning of the simulation, the thermodynamic criterion, i.e. the local value of the dislocation density is checked for all the lattice points. If it is larger than some critical value, the lattice point is spontaneously recrystallized without any orientation change, i.e. a dislocation density of 0 is assigned to it and the original crystal orientation is preserved. In the next step, the growth algorithm starts according to Eqs. (20)–(22). This means that the kinetic conditions for nucleation are checked by calculating the misorientations among all spontaneously recrystallized cells (preserving their original crystal orientation) and their neighbor cells. If any such pair of cells reveals a misorientation

above 15°, the cell flip of the non-recrystallized cell is calculated according to its actual transformation probability. In the case of a successful cell flip, the orientation of the first recrystallized neighbor cell is assigned to the flipped cell.

3.5.3 Predictions and Interpretation

Figures 17 and 18 show simulated microstructures for site-saturated spontaneous nucleation in all cells with a dislocation density larger than 50% of the maximum value (Fig. 17) and larger than 70% of the maximum value (Fig. 18). Each figure shows a set of four subsequent microstructures during recrystallization. The upper graphs show the evolution of the stored dislocation densities. The gray areas are recrystallized, i.e. the stored dislocation content of the affected cells is dropped to 0. The lower graphs represent the microtexture images where each gray scale represents a specific crystal orientation. The fat lines indicate grain boundaries with
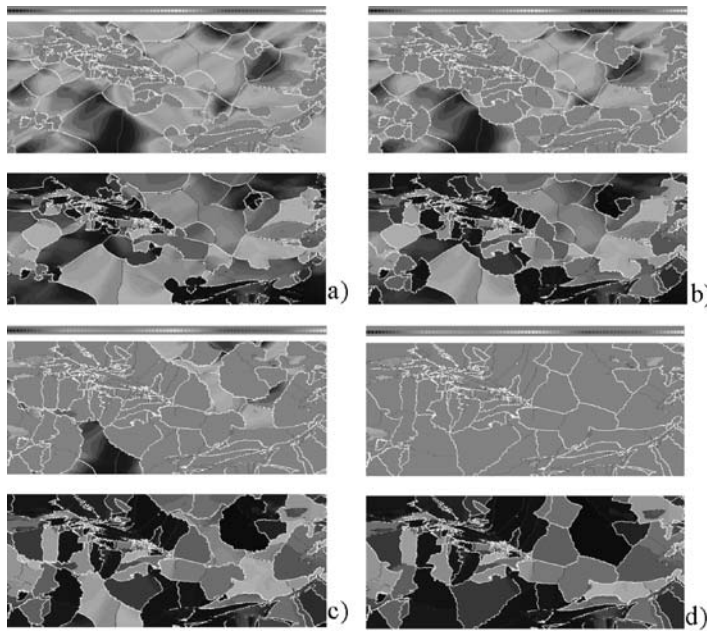
Figure 17 Consecutive stages of a 2D prediction of primary static recrystallization in a deformed aluminum polycrystal on the basis of a crystal plasticity finite element simulation in conjunction with a cellular automaton. The figure shows the change in dislocation density (a and b) and in microtexture (c and d) as a function of the annealing time during isothermal recrystallization. The texture is given in terms of the magnitude of the Rodriguez orientation vector using the cube component as reference. The gray areas in the upper figures indicate a stored dislocation density of 0, i.e. these areas are recrystallized. The white lines indicate grain boundaries. The simulation parameters are: 800 K; thermodynamic instability criterion: site-saturated spontaneous nucleation in cells with at least 50% of the maximum occurring dislocation density (threshold value); kinetic instability criterion for further growth of such spontaneous nuclei: misorientation above 15°; activation energy of the grain boundary mobility: 1.46 eV pre-exponential factor of the grain boundary mobility: $m_0 = 8.3 \times 10^{-3} \, \mathrm{m}^3/(\mathrm{N\,s})$; mesh size of the cellular automaton grid (scaling length): $\lambda_\mathrm{m} = 61.9 \, \mu\mathrm{m}$.
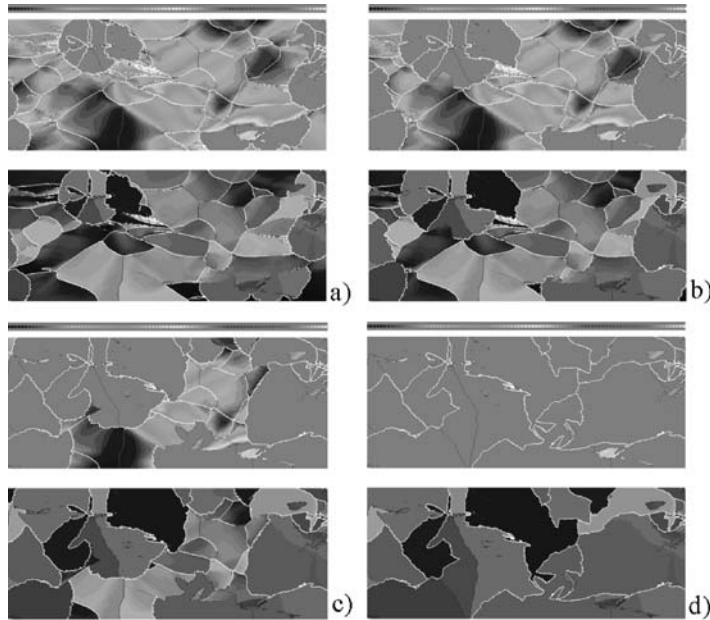
Figure 18   Parameters as in Fig. 17, but site-saturated spontaneous nucleation occurred in all cells with at least 70% of the maximum occurring dislocation density.

misorientations above 15°. The thin lines indicate misorientations between 5° and 15°. The incipient stages of recrystallization in Fig. 17 (cells with 50% of the maximum occurring dislocation density undergo spontaneous nucleation without orientation change) reveal that nucleation is concentrated in areas with large accumulated local dislocation densities. As a consequence, the nuclei form clusters of similarly oriented new grains. Less deformed areas between the bands reveal a very small density of nuclei. Logically, the subsequent stages of recrystallization reveal that the nuclei do not sweep the surrounding deformation structure freely as described by the Avrami–Johnson–Mehl theory but impinge upon each other and, thus, compete at an early stage of recrystallization. Figure 18 (using 70% of the maximum occurring dislocation density as threshold for spontaneous nucleation) also reveals strong nucleation clusters in areas with high dislocation densities. Owing to the higher threshold value for a spontaneous initial cell flip, nucleation outside of the deformation bands occurs vary rarely. It also shows an increasing grain size as a consequence of the reduced nucleation density.

Both simulations reveal further that, in most areas where spontaneous nucleation occurs due to a certain dislocation density, the kinetic instability criterion is also fulfilled stimulating further growth of the recrystallized cells. This means that high dislocation densities and large local lattice curvatures typically occur in close neighborhood. Another observation is that the nucleation clusters are particularly concentrated in macroscopical deformation bands which were formed as diagonal through-thickness instabilities. Generic *intrinsic* nucleation inside heavily deformed grains, however, occurs rarely. Nucleation is often successful at former grain boundaries where orientation changes occur naturally. This means that there might be a

transition from *extrinsic* nucleation such as inside bands or related large scale instabilities to *intrinsic* nucleation inside grains or close to existing grain boundaries. Another important result is the partial recovery of deformed material. Figures 17d and 18d reveal small areas where moving large angle grain boundaries did not entirely sweep the deformed material. An analysis of the state variable values at these coordinates and of the grain boundaries involved shows that not insufficient *driving forces* but insufficient *misorientations* between the deformed and the recrystallized areas − entailing a drop in grain boundary mobility − are responsible for this effect (orientation pinning). Details of the simulations are given in Refs. [12–15,37–39].

3.6   Relationships Between Texture, Microstructure, and Surface Properties

This section presents a study about surface roughening in aluminum and its microstructural origin. Four aluminum sheets were solution heat treated and aged to T4 condition. All samples have the same chemical composition (6xxx) but different processing and surface quality. The tensile samples were cut parallel to the transverse direction of rolling. The joint experimental investigation for the determination of surface roughness and ridging comprised the following steps: First, the microtexture at the sample surfaces was determined using orientation mapping in the SEM. The step size was $15\,\mu m$ and the scanned area amounted to $4000\,\mu m \times 4000\,\mu m$. Second, the 3D surface topography was measured using a white-light confocal microscope. Third, tensile, bending, and drawing tests were conducted. After each deformation step, digital stereo image pairs of the sample surfaces were taken. Fourth, the 3D plastic displacement fields was calculated after each straining step using photogrametry. The displacement field serves as input for deriving the surface components of the local strain tensor. After each deformation step, the surface pattern was acquired and the displacement field as well as the strain distribution was calculated. Using the experimental procedures described above, three sets of mappings were determined exactly at the same locations of each sample after each deformation step, namely, the surface topography, the microstrain distribution, and the microtexture. Figure 19 shows the surface topography of two of the four samples after tensile elongation of 7% parallel to the transverse direction (engineering strain). Before the deformation, all surfaces were polished to a mirror-finish. The gray scale code represents the local height. The presented topography maps are filtered with respect to sample bending and long range undulations. While the specimen in the upper row (sample A) shows a very heterogeneous distribution of the von Mises strain and pronounced banding of the cube orientation parallel to the rolling direction, the sample in the lower row (sample B) reveals a very good surface finish, homogeneous strain distribution and a fine grained distribution of the cube texture component without banding or related clustering effects. A comparison of the distribution of the Goss grains in the sheet plane between the two samples A and B does not give a very clear picture (Fig. 20). The overall volume fraction of grains with Goss orientation is very small. Sample A shows a somewhat more pronounced banding of the Goss grains while sample B shows a very homogeneous distribution. The wavelength of the Goss bands in sample A can be correlated with the spacing of the large bands found in the in-plane distribution of the von Mises strain (Fig. 19).
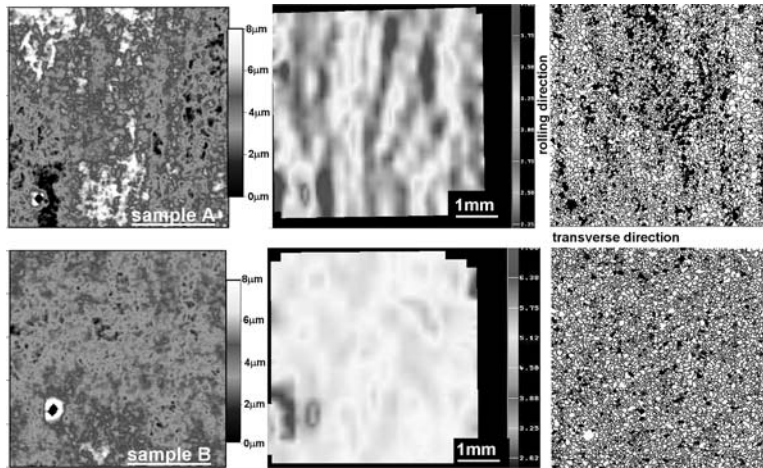
Figure 19 Two 6xxx aluminum samples in T4 state with different surface quality. The sample presented in the lower row (sample B) reveals better surface quality. The data show: surface topography (left), microstrain distribution, von Mises (center), microtexture, distribution of the cube component (right).

Although it reveals only a very small overall volume fraction, the Goss orientation is nonetheless a very interesting texture component in the context of ridging and surface roughening. This is due to its symmetry. Tensile tests conducted on flat specimens parallel to the rolling direction reveal less ridging when compared to those which are extended parallel to the transverse direction. In rolling direction, the Goss orientation has a similar Taylor factor and a similar reorientation rate as the cube orientation. However, when rotating the sample 90° about the sheet normal direction prior to testing, thereby turning also the Goss component into a Goss component rotated 90° about its normal direction, renders the Goss a highly unstable orientation with a very high Taylor factor close to 5 and a very high plastic spin.

Figure 21 shows the development of the surface roughness during sheet bending for the four different samples as a function of the bending angle. The measurements were conducted using confocal white-light microscopy. The curvature
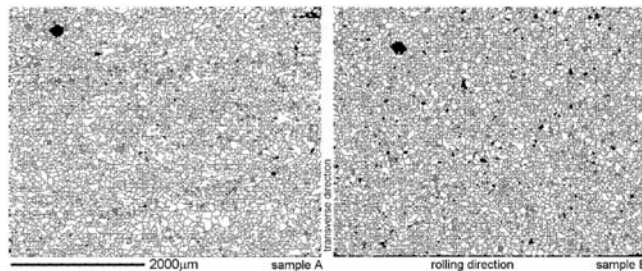
Figure 20 Two 6xxx aluminum samples in T4 state with different surface quality. Sample B reveals better surface quality than sample A. The figure shows the distribution of the Goss-oriented grains in the sheet plane.
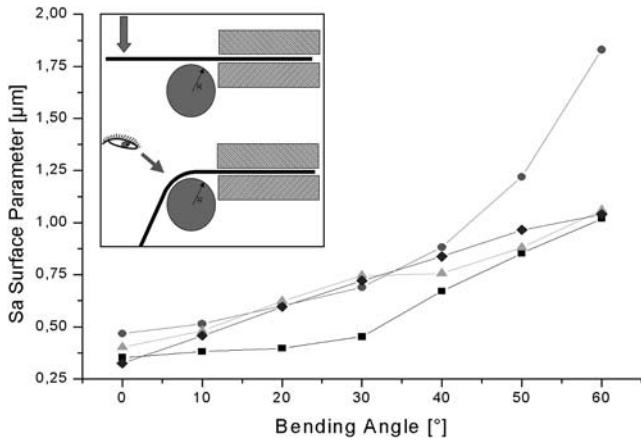
Figure 21 Development of surface roughness in four aluminum samples during bending as a function of the bending angle. The experimental set-up is presented in the upper left corner. The measurements were made using confocal white-light microscopy.

stemming from bending was corrected by a filter before analyzing the microscopic roughness. The diagram shows different behavior for the different samples. Three of the samples reveal a close to linear development of surface roughness for bending angles below 30°. One of the specimens has a much more pronounced roughening behavior particularly above 40°. Changes in roughening occur in a transition regime between 30° and 40° which is also indicated by the data shown in Fig. 22 for one of the samples. The critical regime might indicate a change in mechanism, such as the transition from ridging to ridging plus orange peel [40].
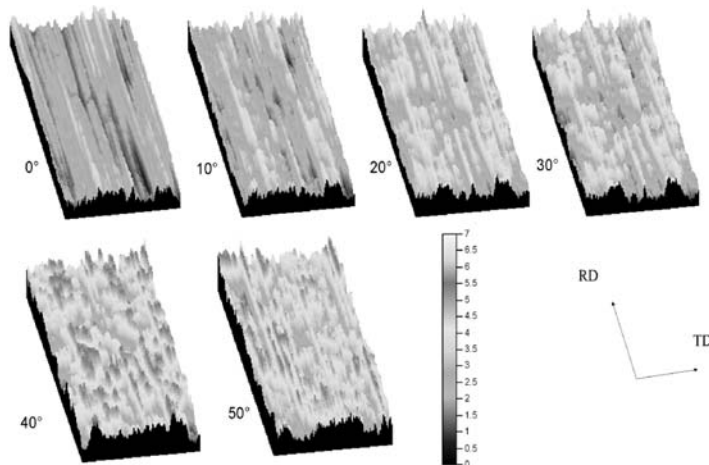
Figure 22 Example for the development of the surface roughness during bending as a function of the bending angle. The measurements were made using confocal white-light microscopy.

4   INTEGRATION OF ANISOTROPY INTO METAL-FORMING SIMULATIONS

4.1   Introduction to Anisotropy Engineering

The anisotropy concepts to be reviewed in the following are empirical yield surface approximations, yield surface formulations based on crystallographic homogenization theory, and the recently introduced texture component crystal plasticity finite element method. The present state in anisotropy engineering is naturally different between industrial applications and basic science. The use of empirical or semi-empirical polynomials for yield surface approximations is the standard procedure in the industrial practice whereas the various crystal plasticity finite element methods gradually become a standard in the basic material sciences. The importance of empirical approaches in the industrial practice is due to the fact that they provide short computation times and require only a small set of input data. An important weakness of empirical approaches lies in the absence of texture update. The prevalence of the crystal plasticity finite element method in basic research is essentially due to the natural incorporation of texture update. The major drawback of the crystal plasticity approach is the long calculation times which presently exceed those required for yield surface simulations by a factor of 50–100. An improvement in the speed of the crystal plasticity methods is attained by the recent introduction of the texture component crystal plasticity finite element method which exceeds the computation times of yield surface calculations only by a factor of 15–25. Overviews to the fundamentals of anisotropy engineering are given in Refs. [38,41–43].

4.2   From Scalar to Tensorial Material Concepts

The yield surface represents the generalization of the yield point from uniaxial tensile testing to general stress states. Expanding the yield point into a closed yield surface is only required if the material under inspection shows elastic–plastic anisotropy, i.e. if it deforms differently in different directions. However, such behavior is the rule and not the exception in real materials. Aluminum polycrystals with random and therefore quasi-isotropic behavior do practically not occur in sheet metal-forming operations.

The physical nature of elastic–plastic anisotropy in metals is the crystalline arrangement of the atoms. Commercial aluminum usually occurs in polycrystalline form where each grain has a different crystallographic orientation, shape and volume fraction. The distribution of the orientations in a polycrystalline aggregate is referred to as (crystallographic) texture. The anisotropy of the elastic tensor and the discrete nature of crystallographic slip along densely packed lattice directions on preferred crystal planes entail a highly anisotropic response of polycrystalline specimens to mechanical loading. While the elastic–plastic deformation of single aluminum crystals and bicrystals as a function of their orientation can nowadays be well predicted, plasticity of polycrystalline samples is less well understood. This is essentially due to the intricate elastic–plastic interactions occurring during co-deformation among the highly anisotropic individual grains. This interaction leads to strong heterogeneity in terms of strain, stress, and crystal orientation. Another difficulty in tackling the anisotropy of polycrystalline aluminum lies in

the fact that the crystals rotate during forming, owing to the skew symmetric portion of the displacement gradients created by crystal slip. This means that texture and anisotropy gradually change during forming, even under constant strain path conditions. In this context, it must be underlined that crystallographic orientation changes are principally non-reversible owing to the orientation sensitivity of strain path changes and the orientation dependence of strain hardening. This means that – even in the case of very simple strain paths –mechanics and texture should be treated jointly due to the strong non-linearity of the problem. These aspects which show the complexity of texture and anisotropy during forming underline that, for an engineering purpose, one major aim of polycrystal research must lie in identifying adequate measures for mapping crystallographic anisotropy into classical mathematical methods for predicting large-strain plastic deformation. The second even more challenging aim lies in developing methods for predicting also the *change* of crystal anisotropy during forming on a sound physical basis [38,41–43].

4.3 The Physical Origin of Crystalline Elastic–Plastic Anisotropy

4.3.1 Elastic Anisotropy

The elastic anisotropy of crystalline matter departs from the directionality of the electronic bond and the resulting crystal lattice structure. For small deviations of the atoms from their equilibrium positions, the reversible elastic response to loads can be approximated by a linear relationship which is referred to as Hooke's law. In this framework, the linear elastic constants can be derived as the components of the second derivative of the electronic potential. The elastic constants can be written in the form of a fourth-rank elastic stiffness tensor C_{ijkl} or in the form of a fourth-rank elastic compliance tensor S_{ijkl}. According to

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl}, \quad \varepsilon_{ij} = S_{ijkl}\sigma_{kl} \tag{23}$$

symmetry and thermodynamic considerations reduce the 81 elastic constants to a set of three independent numbers $(C_{1111},C_{1122},C_{2323})^{**}$ in the case of cubic crystal symmetry (e.g. Al, Fe, Cu) and to a set of five independent numbers $(C_{1111},C_{1122},C_{1133},C_{3333},C_{2323})^{\dagger}$ in the case of hexagonal crystal symmetry (e.g. Ti, Mg, Zn). The deviation from elastic isotropy can for cubic crystals be quantified by the so-called Zener anisotropy ratio

$$A = \frac{2C_{2323}}{C_{1111} - C_{1122}} \tag{24}$$

While aluminum has a relatively low elastic anisotropy with $A = 1.215$, iron has a larger Zener ratio of $A = 2.346$. Of all the cubic metals, tungsten has the lowest deviation from isotropy with a Zener ratio of $A \approx 1$ and lithium the largest with $A = 9.34$.

**Corresponding to (C_{11}, C_{12}, C_{44}) in reduced matrix notation.
†Corresponding to $(C_{11}, C_{12}, C_{13}, C_{33}, C_{44})$ in reduced matrix notation.

4.3.2 Plastic Anisotropy

The plastic anisotropy of crystalline matter also departs from the directionality of the electronic bond and the resulting crystal lattice structure. Both aspects determine which slip planes and which translation vectors (Burgers vectors) serve for the motion of lattice dislocations or the activation of plastically relevant athermal transformations. The main consequence of this anisotropy in the present context is that metals are deformed in a discrete rather than in a continuum fashion rendering plasticity an intrinsically anisotropic property of metals. Assuming that the normalized Burgers vectors b_j and the normalized slip plane normals n_i of the s different slip systems available in a particular crystal lattice are known, their orientation factors m_{ij} can be readily formulated as dyadic products according to

$$m_{ij}^s = n_i^s b_j^s \tag{25}$$

with the symmetric portion being

$$m_{ij}^{\text{sym},s} = \frac{1}{2}\left(n_i^s b_j^s + n_j^s b_i^s\right) \quad \text{(crystal coordinates)} \tag{26}$$

when given in crystal coordinates. One must note that all slip vectors are normalized. Transforming the latter equation into the sample coordinate system (Fig. 23) leads to

$$m_{kl}^{\text{sym},s} = \frac{1}{2}(a_{ki}n_i^s a_{lj}b_j^s + a_{lj}n_j^s a_{ki}b_i^s) \quad \text{(sample coordinates)} \tag{27}$$

where a_{ki} and a_{lj} are the transformation matrices between the crystal coordinate system and the sample coordinate system. Using these s different orientation factors, $m_{kl}^{\text{sym},s}$ of the s different available slip systems for the transformation of an external load into the slip geometry provides a simple kinematic formulation for the yield surface of a single crystal, i.e.

$$m_{kl}^{\text{sym},s}\sigma_{kl} = \tau_{\text{crit},+}^{s,\text{active}}$$
$$m_{kl}^{\text{sym},s}\sigma_{kl} = \tau_{\text{crit},-}^{s,\text{active}} \tag{28}$$
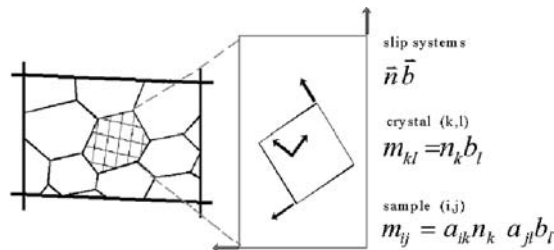
Figure 23 The plastic anisotropy of crystalline matter departs from the directionality of the electronic bond and the resulting crystal lattice structure. Both aspects determine the slip planes and translation vectors (Burgers vectors) on which lattice dislocations move during plastic deformation. The diagram shows the different coordinate system and the resulting geometrical transformation operations one has to consider in this context.

for the *active* slip systems, and

$$
\begin{aligned}
m_{kl}^{\text{sym},s}\sigma_{kl} &< \tau_{\text{crit},+}^{s,\text{non-active}} \\
m_{kl}^{\text{sym},s}\sigma_{kl} &< \tau_{\text{crit},-}^{s,\text{non-active}}
\end{aligned}
\tag{29}
$$

for the *non-active* slip systems (Fig. 24). One must note that the Einstein summation rule applies in all equations in case not stated otherwise. The slip dyads of aluminum at room temperature typically contain $\langle 1\,1\,1\rangle$ and $\langle 1\,1\,0\rangle$ vectors [38,41–46].

Most points on the single crystal yield surface describe single-slip conditions. In the graphical representation of the yield surface, single slip generally takes place when
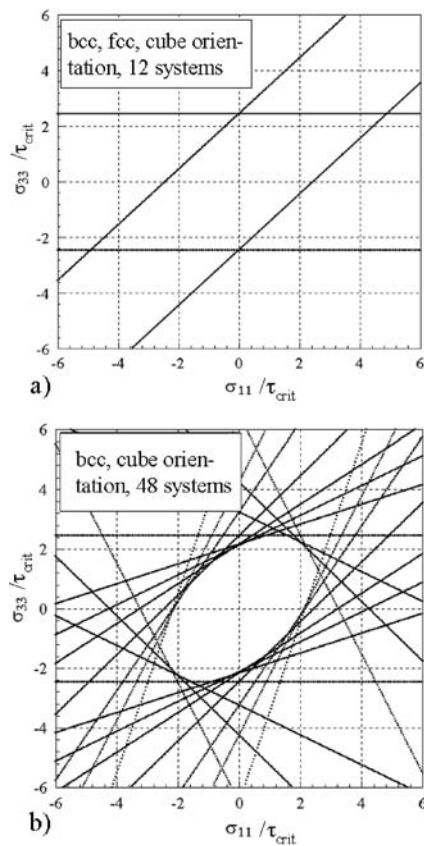
Figure 24 A simple Schmid-type formulation considering the different orientation factors of all available slip systems which essentially transform an external load into shear stresses acting on the slip systems provides a kinematic formulation for the yield surface of a single crystal. The yield surface shown in figure (a) was derived using the 12 $\{1\,1\,0\}\langle 1\,1\,1\rangle$ slip systems. The yield surface shown in figure (b) was derived using the 12 $\{1\,1\,0\}\langle 1\,1\,1\rangle$, 12 $\{1\,1\,2\}\langle 1\,1\,1\rangle$, and 24 $\{1\,2\,3\}\langle 1\,1\,1\rangle$ slip systems (body-centered cubic). The figure indicates that body-centered cubic alloys, therefore, behave plastically principally different from face-centered cubic alloys.

the stress tensor (in vector transformation notation[‡]) points at a hyperplane rather than a hyperconus (Fig. 25). Note that the cubes placed in Fig. 25 indicate the changing orientation of the external reference system, i.e. of the stress state. Polyslip conditions, as usually required for polycrystal deformation owing to the satisfaction of strain-rate compatibility among the grains, are characterized by hyperconus coordinates of the stress state (Fig. 26). The conus positions for the stress can be calculated using a conventional homogenization approach, for instance, the Taylor–Bishop–Hill theory (indicated by σ^{TBH} in Fig. 26). The corresponding multi-slip positions of the stress tensor, satisfying an externally imposed strain rate, are then denoted as Taylor positions. It must be noted in this context that the Taylor factor generally takes the form of a stress shape tensor for the crystal yield surface rather than that of a factor owing to its dependence on the strain-rate tensor. Its magnitude for a given strain rate determines the kinematic size of the yield surface in the corresponding stress direction characterizing the correct polyslip hyperconus and, thus, the kinematic portion of the corresponding stress state. The symbols $D^{s=1}$ and $D^{s=2}$ in Fig. 26 indicate the single-slip strain states from slip systems 1 and 2. Using these two slip systems allows one to realize any strain-rate state in the

$$\frac{\sigma_a m_a^{\text{sym}}}{\tau_{\text{crit}}} = \frac{\sigma_b m_b^{\text{sym}}}{\tau_{\text{crit}}} = \frac{\sigma_d m_d^{\text{sym}}}{\tau_{\text{crit}}} = \frac{\sigma_c m_c^{\text{sym}}}{\tau_{\text{crit}}} = 1$$

$$|\sigma_c| < |\sigma_b| < |\sigma_d| < |\sigma_a|$$
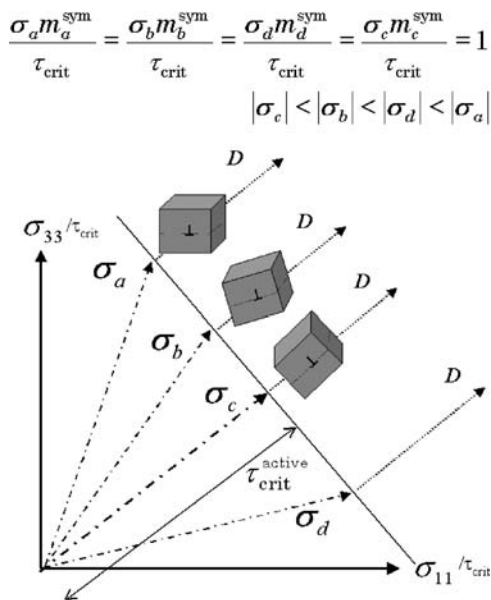
Figure 25   Most points on the single crystal yield surface describe single-slip conditions. In the graphical representation of the yield surface, single-slip generally takes place when the stress state (here given in vector notation) points at a hyperplane rather than a hyperconus. The strain-rate tensor is indicated by **D** and m is the Schmid factor, i.e. the dyadic product of the slip elements. The small cubes placed in the figure indicate the changing relative orientation between the external reference system and the crystal coordinate system.

[‡]Using the tensor–vector transformation rule.
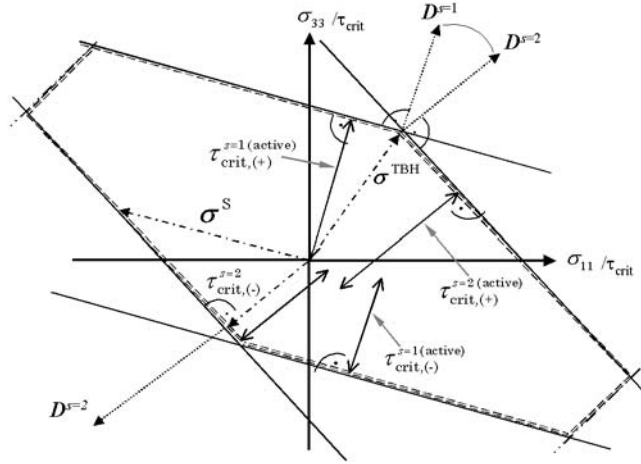
Figure 26   Polycrystal deformation requires polyslip conditions in order to satisfy strain-rate compatibility among the grains. Polyslip states are crystallographically characterized by hyperconus coordinates of the stress state. The conus positions for the stress can be calculated using a conventional homogenization approach, for instance, the Taylor–Bishop–Hill theory (indicated by σ^{TBH}). The symbols $D^{s=1}$ and $D^{s=2}$ indicate the single-slip strain states from slip systems 1 and 2. Using these two slip systems allows one to realize any strain-rate state in the respective conus by a linear combination of $D^{s=1}$ and $D^{s=2}$.

respective conus by a linear combination of $D^{s=1}$ and $D^{s=2}$. For cubic crystals, the yield surface reveals four classes of Taylor states for polyslip and one for single slip [38,41–46]. These yield states are referred to as

penta slip state (five active slip systems): $^{5}M_{pq}^{i}$(fcc, bcc (reduced): $i = 56$)

tetra slip (four active slip systems): $^{4}M_{pq}^{j}$(fcc, bcc (reduced): $j = 108$)

tri slip (three active slip systems): $^{3}M_{pq}^{k}$(fcc, bcc (reduced): $k = 135$)

bi slip (two active slip systems): $^{2}M_{pq}^{l}$(fcc, bcc (reduced): $l = 66$)

single slip (one active slip system): $^{2}M_{pq}^{n}$(fcc, bcc (reduced): $n = 24$)

$$(30)$$

where *fcc* denotes *face-centered cubic* and *bcc* denotes *body-centered cubic* crystal structure. The term *reduced* indicates that only 12 $\{1\,1\,1\}\langle1\,1\,0\rangle$ bcc slip systems were considered. The number at the end of each row gives the number of different conus cases (and single-slip cases) for the respective Taylor state. The total Taylor stress state for a polycrystalline aggregate can for a given external strain-rate state then be integrated as a volume weighted sum of all Taylor tensors derived separately for each grain for this boundary condition (Fig. 27).

$$\sigma_{\lambda}^{\text{T}} = \left\{ \frac{1}{\sqrt{6}}(2\sigma_{33} - \sigma_{11} - \sigma_{22}), \frac{1}{\sqrt{2}}(\sigma_{22} - \sigma_{11}), \sqrt{2}\sigma_{23}, \sqrt{2}\sigma_{13}, \sqrt{2}\sigma_{12} \right\} \qquad (31)$$
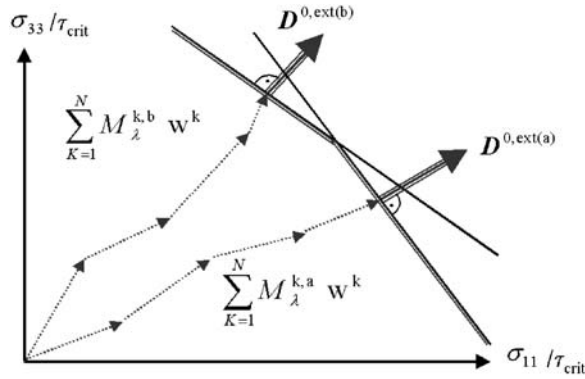
Figure 27 The Taylor stress state for a polycrystalline aggregate can for a given external strain-rate state be integrated as a volume weighted sum of all Taylor factors derived separately for each grain for the respective boundary condition. In this figure, M is the Taylor tensor, D is the strain rate, and w is the volume fraction. The counter k sums over all crystals in the aggregate.

4.4 Empirical Approximations of the Yield Surface

The first empirical mathematical description of an anisotropic plastic yield surface was suggested in 1928 by von Mises in the form of a quadratic function [47]. This approach which was originally designed to approximate the plastic anisotropy of single crystals was in 1948 rendered by Hill [48] into a generalized form using the Huber–Mises–Hencky approach:

$$f(\sigma_{ij}) = (F(\sigma_{22} - \sigma_{33})^2 + G(\sigma_{33} - \sigma_{11})^2 + H(\sigma_{11} - \sigma_{22})^2$$
$$+ 2L\sigma_{23}^2 + 2M\sigma_{13}^2 + 2N\sigma_{12}^2)^{1/2} \tag{32}$$

where F, G, H, L, M, and N are anisotropy coefficients. The above equation can be rewritten as

$$f(S_{ij}) = ((G+H)S_{11}^2 + (F+H)S_{22}^2 + (F+G)S_{33}^2 - 2HS_{11}S_{22}$$
$$- 2GS_{11}S_{33} - 2FS_{22}S_{33} + 2LS_{23}^2 + 2MS_{13}^2 + 2NS_{12}^2)^{1/2} \tag{33}$$

where S_{ij} are the deviatoric stress components. The shape coefficients of Hill's quadratic yield function can be fitted from experimentally obtained mechanical data such as the Lankford values taken in different directions of a specimen. Scaling can be provided by the yield stress obtained from uniaxial tensile testing. While the Lankford coefficients and the yield stress can be determined from tensile testing, the direct measurement of mechanical response under complex loads is an intricate task. Although Hill-based anisotropy simulations (referring to the Hill 1948 model) provide decent approximations at least of the initial plastic anisotropy in the case of certain iron textures and a number of textures in interstitial free steels, they typically fail to predict the yield shape of most aluminum alloys, high strength steels, austenitic steels, copper, or hexagonal materials. Typical examples where the Hill 1948 yield criterion is not applicable are cup drawing operations of aluminum

crystals with sixfold slip symmetry, i.e. with a crystal {1 1 1} plane parallel to the sheet surface. In this case, six slip systems have identical Schmid factor relative to the surface which cannot be modeled by the Hill polynomial owing to its quadratic form. Due to this principle shortcoming, a number of optimized empirical anisotropic yield surface concepts with higher-order polynomial forms have been proposed in the last decade, such as those introduced later by Hill [49] and by Barlat [50] which are better suited for face-centered cubic alloys and many body-centered cubic steels. In the last years, various authors have presented improved empirical yield surface approaches where the yield function can be fitted using both mechanically obtained and texture-based data.

The chief advantage of using an empirical anisotropic yield surface function as a constitutive law in metal-forming finite element simulations is time efficiency and the simple mechanical methods with which it can be derived. The dominant limitation of yield surface functions is that the anisotropy of polycrystals generally *changes* during forming owing to the change of texture. This evolution of anisotropy is not mapped by a corresponding change of the shape of the yield surface. In other words, the same yield surface shape is used throughout one finite element simulation without making a physically meaningful update of its steadily changing shape. Although empirical constitutive laws can be used to gradually change the yield surface shape during forming, their capability is typically constrained by a lack of physical information about the actual development of the crystallographic texture during forming.

4.5 Crystallographic Approximations of Elastic–Plastic Anisotropy

4.5.1 Approximation of Elastic Anisotropy Using Polycrystal Homogenization Theory

A typical problem in the field of anisotropy engineering is the approximation of the integral elastic response of a polycrystalline sample under an external load. Although various aspects can principally contribute to the anisotropy of the overall elastic stiffness, we concentrate in the following on the influence of the crystallographic texture. The macroscopic elastic properties of a textured polycrystal can be calculated by formulating appropriate volume-weighted means of the individual elastic single crystal tensor, rotated parallel to the respective local coordinate system of each individual crystal. This average value of the integral elastic tensor must therefore take into account all individual orientations of the grains which are described by the orientation distribution function.

An early homogenization approach for the elastic response under an external load was suggested by Voigt, who assumed that in the case of a macroscopically prescribed strain-rate state, each material portion is in the same strain-rate state as the entire sample, irrespective of its spatial position in the specimen. The strain rate would then be homogeneous throughout the sample. However, in a polycrystalline sample, the elastic response typically varies from grain to grain, due to the spatially changing crystal orientation. Since in the Voigt model the prescribed strain rate is the same everywhere in the sample, the stress must vary. The Voigt limit for the elastic response of a polycrystalline sample can, thus, be calculated by weighting the tensor of the elastic stiffness as a function of orientation with the orientation distribution function. A different approach to treating the homogenization problem in an

elastically loaded polycrystalline sample was suggested by Reuss. He suggested that in the case of a macroscopically prescribed stress state each material portion is in the same stress state irrespective of its spatial position in the specimen. The stress would then be homogeneous throughout the specimen. The elastic response may then vary from grain to grain, in accord with the local orientation of the crystal. Since, in the Reuss model, the prescribed external stress is constant throughout the specimen, the strain must vary according to the local grain orientation. Consequently, the elastic Reuss limit can be calculated for a polycrystal by weighting the tensor of the elastic compliance as a function of orientation with the orientation distribution function. Since neither the Voigt nor the Reuss prediction provides reliable approximations to the elastic modulus of a polycrystal, Hill defined an average modulus which consists of the equally weighted results of both above models.

4.5.2   Approximation of the Yield Surface Using Polycrystal Homogenization Theory

Polycrystalline alloys subject to metal-forming operations typically develop or inherit morphological textures as well as crystallographic textures. While the former are often less relevant in typical commercial sheet material, the latter strongly determine the overall anisotropy. In the following, we will hence concentrate on texture effects on yield anisotropy. Orientation distributions can directly serve as input data for the calculation of the crystallographically determined portion of the yield surface shape using Taylor–Bishop–Hill or self-consistent type approaches [41–43,51,52]. This applies for a single crystal yield surface as well as for the homogenization bounds of the polycrystal yield surface (Fig. 28). The major spirit and advantage of the crystallographic yield surface over
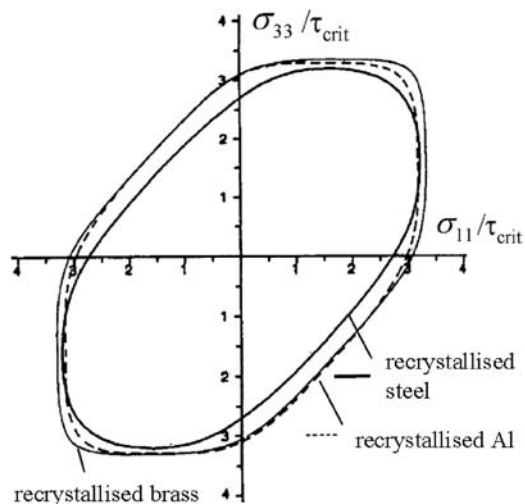
Figure 28   Some examples of yield functions for different materials calculated by use of the homogenization bounds for their respective polycrystal yield surface. The figure shows yield surface sections for aluminum and steel.

empirical concepts consists in the fact that it reduces the individual anisotropic behavior of large sets of individual grains comprising a polycrystalline aggregate (10^6–10^{12} grains for a typical large scale forming operation) to a simple crystallographic shape function. It is, thus, an ideal example of a scale-bridging simulation approach which reduces the tremendous complexity inherent in real microstructures to a simple anisotropic function. Since texture-based yield surface approximations use the complete crystallographic anisotropy information of a specimen, they are often superior to empirical approaches which are fitted only to a small set of mechanical parameters. Modern approaches for the approximation of the yield surface typically use both, texture-based and mechanical data to fit the complete anisotropy function.

   More complex yield surface functions incorporate kinematical and kinetic plasticity effects. In this context, it must be noted that the crystallographic texture only gives the respective anisotropic *shape* function for a particular polycrystalline sample, but the texture dependence of the *internal stress* and the individual hardness of the different grains are typically ignored by the constitutive laws employed in homogenization approaches. However, it is principally feasible to generalize the crystallographic yield surface concept by enriching it with the individual strength of each grain. This leads to a formulation of the following kind:

$$f(S_{ij}) = \frac{1}{V} \int_V M_{ij}(\boldsymbol{g}, D_{ij})\tau_{\text{crit}}(D_{ij}, \boldsymbol{g})\mathrm{d}V \approx \sum_{k=1}^{N} M_{ij}^k \tau_{\text{crit}}^k w^k \tag{34}$$

where $f(S_{ij})$ is the yield surface, V the sample volume, M_{ij} the Taylor shape function obtained by the homogenization theory as a function of strain rate D_{ij} and rotation matrix $\boldsymbol{g}$, τ_{crit} the flow stress of each individual grain, and w the volume fraction of each grain. An example where kinetic information about the local texture-dependent hardness of the various grains has been used to approximate a yield surface is given in Fig. 29 [38,53]. The left diagram shows a portion of the yield surface as it anisotropically shrinks during partial recrystallization of aluminum. The right-hand side shows three subsequent time steps of a coupled crystal plasticity FEM–cellular automaton simulation where the upper figure gives the texture in terms of the magnitude of the Rodriguez vector and the lower figure the strength in terms of the dislocation density (black areas are recrystallized). The data from this discrete simulation served as an input to the kinematic–kinetic yield surface model.

4.6   A Texture Component Crystal Plasticity Finite Element Method For Scalable Large-Strain Anisotropy Simulations

4.6.1   The Basic Challenge of Constitutive Anisotropy Modeling

Aluminum polycrystals typically have crystallographic textures which form and inherit during processing. The intrinsic elastic and plastic anisotropy of crystalline matter entails an overall anisotropic response of such specimens when mechanically loaded. This behavior imposes two basic requirements in the context of forming simulations. The first one is the mapping of the *initial* anisotropy into mathematical formulations for predicting large-strain deformation. The second even more demanding goal is the description of the *change* of crystalline anisotropy during
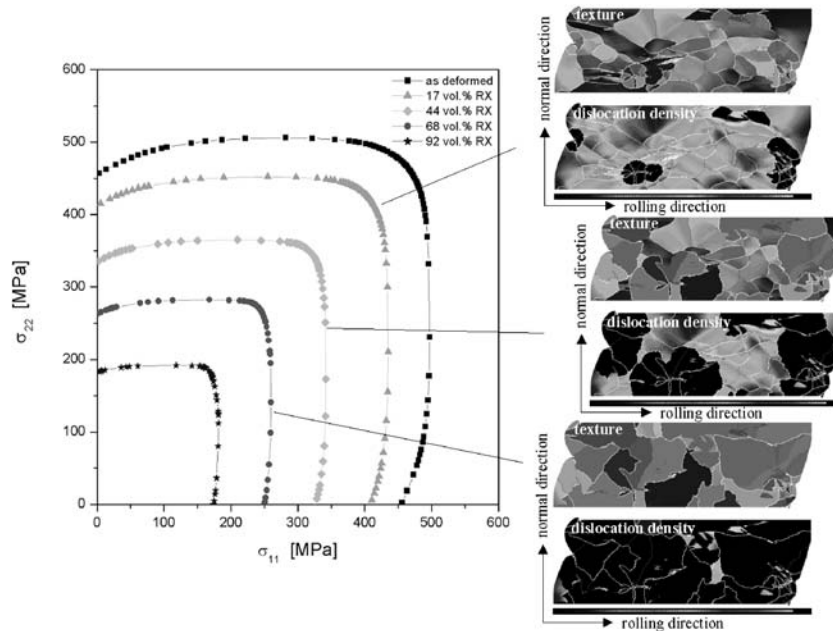
Figure 29 Calculation of the yield surface of a recrystallizing aluminum sample. The simulations were performed using a crystal plasticity finite element simulation in conjunction with a cellular automaton.

forming. This is necessary since the crystals rotate during deformation owing to the antisymmetry of the crystalline displacement gradients. The first problem can be solved using adequate yield surface functions. The second problem cannot be solved in a straightforward fashion by the use of a statistical constitutive law since each crystal can take an individual reorientation path during forming. Translating this into the yield surface concept means that any constitutive evolution law for the shape function of the yield surface depends on the individual behavior of at least 10^3 possible discrete texture components as a function of their path-dependent local stress and strain-rate states. For solving this problem a new efficient and physically rooted prediction polycrystal plasticity simulation method has recently been developed which accounts for these two crystallographic aspects [9,10]. It is based on directly feeding discrete orientation components onto the integration points of a crystal plasticity finite element model. The texture components can be extracted from experimental data, such as pole figures stemming from x-ray or electron diffraction.

4.6.2 Using Texture Components in Crystal Plasticity FE Constitutive Models

As outlined above, the main issue of large-strain, large-specimen crystal plasticity metal-forming simulations is to tackle both the anisotropy due to the *starting texture* and due to *texture evolution*. One method could be to assign each grain orientation to one integration point. However, such a brute force approach is not feasible when

simulating larger parts containing $\sim 10^{10}$ grains (e.g. automotive parts made of 6xxx aluminum). This essential limitation suggests the use of a more compact mathematical form to map and update textures of large parts properly during metal-forming simulations. Crystal plasticity finite element approaches which update the texture require a *discrete* representation of the orientation distribution function or a portion of it at each integration point. Mapping such a discrete portion of the global texture requires the reduction of the information content to a level at which complex deformation processes can be simulated at reasonable computation costs. Such a form is offered by the texture component method. It approximates the orientation distribution function by a superposition of sets of simple standard functions with individual coordinates, orientation density, and scatter in orientation space. Such a representation of a preferred orientation is referred to as a texture component. In contrast to the use of global symmetric Wigner functions, for instance, in the Fourier-type series expansion methods, the texture component method is based on using spherical, normalized, localized standard functions [54–56]. The superposition can be expressed by

$$f(g) = F + \sum_{c=1}^{C} I^c f^c(g) = \sum_{c=0}^{C} I^c f^c(g), \quad \text{where } I^0 = F,\ f^0(g) = 1 \tag{35}$$

where g is the orientation, $f(g)$ the orientation distribution function, F the volume portion of all randomly oriented crystals (random texture component), and I^c the volume portion of all crystals which belong to the texture component c (Fig. 30) [56]. The orientation distribution function is defined by

$$f(g)\mathrm{d}g = 8\pi^2 \frac{\mathrm{d}V_g}{V} \quad \text{which implies } f(g) \geq 0 \tag{36}$$

where V is the sample volume, and $\mathrm{d}V_g$ the volume of all crystals with an orientation g within the orientation portion $\mathrm{d}g = \sin(\phi)\,\mathrm{d}\phi\,\mathrm{d}\varphi_1\,\mathrm{d}\varphi_2$. Normalization requires

$$\oint f^c(g)\mathrm{d}g = 1 \quad \text{which implies } \sum_{c=0}^{C} I^c = 1 \tag{37}$$
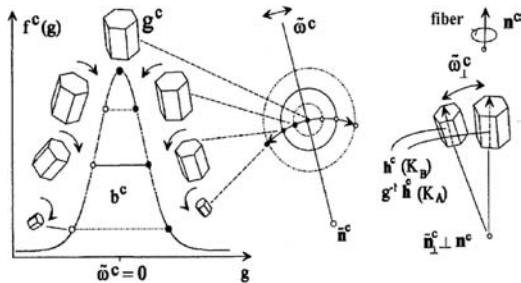
Figure 30 Schematical sketch of a spherical texture component c with a preferred orientation g^c and scatter width b^c. $f^c(g)$ only depends on $\tilde{\omega}^c = \tilde{\omega}(g^c, g)$, i.e. it is independent of the rotation axis $\tilde{n}^c$. (From Refs. 55 and 56.)

As a rule, texture components require positivity, i.e.

$$f^c(g) \geq 0 \quad \text{for all } g \in G \text{ and } I^c > 0 \tag{38}$$

where G is the orientation space. Distribution functions which have a maximum at a preferred orientation g^c and decrease with increasing orientation distance $\tilde{\omega}^c = \tilde{\omega}(g^c, g)$ are referred to as central functions. Such functions, including corresponding pole figures, can be generally represented in the form of series expansions of χ functions or, respectively, Legendre polynomials. More practical approximations of texture component have been introduced on the basis of spherical Gauß- and Lorentz-functions [9,10,54–56]. Although these functions are not exactly in accord with the central limit theorem, they can be represented in an analytically closed form. In most cases, ordinary spherical Gauß functions are well suited to decompose a texture (Fig. 31). The texture component method provides a small set of compact functions which are characterized by simple parameters of physical significance (Euler angles, scatter, volume fraction). Usually, only a few texture components are required for representing textures of aluminum in a precise mathematical form. The most important texture components in aluminum are the cube-component ($\{0\,0\,1\}\langle 1\ 0\ 0\rangle$, $\varphi_1 = 0°$, $\phi = 0°$, $\varphi_2 = 0°$), the Goss-component ($\{0\,1\,1\}\langle 1\ 0\ 0\rangle$, $\varphi_1 = 0°$, $\phi = 45°$, $\varphi_2 = 0°$), the brass-component ($\{0\,\underline{1}\,1\}\langle 2\ 1\ 1\rangle$, $\varphi_1 = 35°$, $\phi = 45°$, $\varphi_2 = 0°$), the copper-component ($\{\underline{2}\,1\,1\}\langle 1\ 1\ 1\rangle$, $\varphi_1 = 90°$, $\phi = 35°$, $\varphi_2 = 45°$), and the S-component ($\sim \{1\,2\,3\}\langle 6\ 3\ 4\rangle$, $\varphi_1 = 60°$, $\phi = 32°$, $\varphi_2 = 65°$).

The texture components are extracted from experimentally obtained pole figures (x-ray diffraction, neutron diffraction) or single orientation data sets (electron diffraction) by identifying the main texture maxima and the suggested scatter width and by minimizing the deviation between the original (experimental) texture and the one created by the texture component functions in an iterative fashion. Depending on the experience in interpreting crystallographic textures, the user can specify the position, height, and scatter of the texture components within certain bounds during the minimization. This makes sense, when the number of texture components
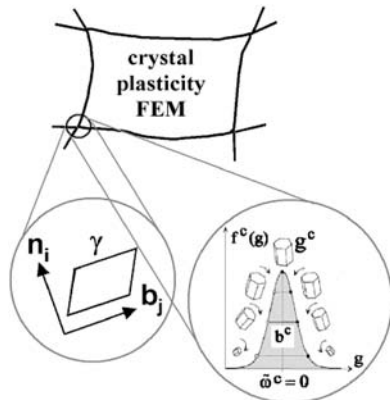
Figure 31 Basic set-up of the texture component crystal plasticity finite element method. (From Refs. 9 and 10.)
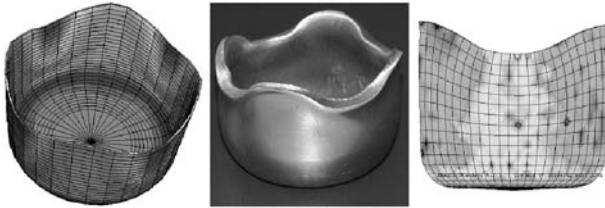
Figure 32 Simulation (left, right) and experiment (center) of cup drawing of aluminum. Left: the gray scale indicates the sheet thickness. Right: the gray scale indicates the orientation change during drawing.

initially prescribed to match an experimental texture is small or when a certain scatter width of the components should not be exceeded. Subsequently, the so-extracted texture components must be mapped onto the integration points of a finite element mesh. This is conducted in two steps. First, the discrete preferred (center) orientation of each texture component is assigned in terms of its respective Euler triple (φ_1, ϕ, φ_2) onto each integration point. It is important in this context that the use of the Taylor assumption locally allows one to map more than one texture component on an integration point. Second, all preferred orientations of the components that were initially assigned to the integration points are rotated in a fashion that the resulting distribution of all the rotated orientations deviating from the initial preferred orientation reproduces exactly the desired texture component function. This means that the scatter which was originally only given in orientation space is now matched by an equivalent scatter both in real space and in orientation space. Figures 32 and 33 show some results of the texture component crystal plasticity finite element method.
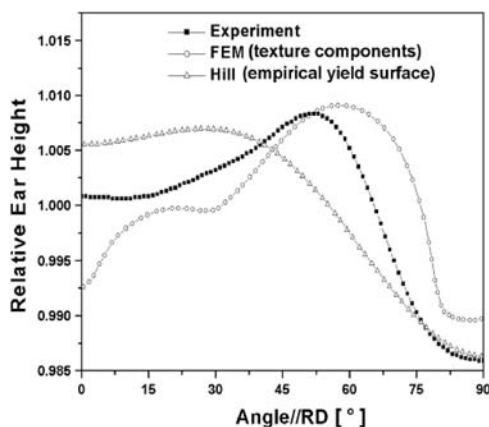
Figure 33 Simulations and experiment of earing for cup drawing of aluminum. Results of the texture component finite element simulation [9,10] and a finite element simulation obtained by the use of a Hill yield surface which was fitted to the texture.

REFERENCES

1. D. Raabe, M. Sachtleber, Z. Zhao, F. Roters, and S. Zaefferer, Acta Mater. 2001, 49, pp. 3433–3441.
2. M. Sachtleber, Z. Zhao, and D. Raabe, Mater. Sci. Eng. A 2002, A336, pp. 81–87.
3. F. Basson and J. H. Driver, Acta Mater. 2000, 48, pp. 2101–2116.
4. J. Hjelen, H. Weiland, J. Butler, J. Liu, W. H. Hu, and E. Nes, Textures Microstruct. 1991, 14–18, pp. 983–988.
5. O. Engler and G. Gottstein, Steel Res. 1992, 63, pp. 413–422.
6. B. L. Adams, S. I. Wright, and K. Kunze, Metall. Trans. A 1993, 24, pp. 819–832.
7. H. Weiland, JOM 1994, 46, pp. 37–44.
8. S. R. Kalidindi, C. A. Brounkhorst, and L. J. Anand, J. Mech. Phys. Solids 1991, 40, pp. 537–554.
9. Z. Zhao, F. Roters, W. Mao, and D. Raabe, Adv. Eng. Mater. 2001, 3, pp. 984–990.
10. D. Raabe, Z. Zhao, and F. Roters, Steel Res. 2001, 72, pp. 421–426.
11. V. Marx, D. Raabe, O. Engler, and G. Gottstein, Textures Microstruct. 1997, 28, pp. 211−218.
12. D. Raabe, Mater. Sci. Forum 1998, 273−275, pp. 169−174.
13. D. Raabe, Proceedings of the Third International Conference on Grain Growth in Polycrystalline Materials (H. Weiland, B. L. Adams, and A. D. Rollett, eds.), Carnegie Mellon University, Pittsburgh, TMS, 1998, pp. 179–185.
14. D. Raabe, Phil. Mag. A 1999, 79, pp. 2339–2358.
15. D. Raabe, Adv. Eng. Mater. 2001, 3, pp. 745–752.
16. J. von Neumann, "The General and Logical Theory of Automata (1963)," in: *Papers of John von Neumann on Computing and Computer Theory* (W. Aspray and A. Burks, eds.), Vol. 12 in the Charles Babbage Institute Reprint Series for the History of Computing, MIT Press, Cambridge, 1987.
17. S. Wolfram, "Theory and Applications of Cellular Automata," in: *Advanced Series on Complex Systems*, selected papers 1983–1986, Vol. 1, World Scientific, Singapore, 1986.
18. J. H. Conway, *Regular Algebra and Finite Machines*, Chapman and Hall, London, 1971.
19. D. Raabe, *Computational Materials Science*, Wiley–VCH, Weinheim, 1998.
20. D. Turnbull, Trans. AIME 1951, 191, pp. 661–692.
21. G. Gottstein and L.S. Shvindlerman, *Grain Boundary Migration in Metals – Thermodynamics, Kinetics, Applications*, CRC Press, Boca Raton, 1999.
22. F. J. Humphreys and M. Hatherly, *Recrystallization and Related Annealing Phenomena*, Pergamon Press, Oxford, 1995.
23. H. W. Hesselbarth and I. R. Göbel, Acta Metall. 1991, 39, pp. 2135–2144.
24. C. H. J. Davies, Scripta Metall. Mater. 1995, 33, pp. 1139–1154.
25. V. Marx, D. Raabe, and G. Gottstein, Proceedings of 16th RISØ International Symposium on Mat. Science: Materials: Microstructural and Crystallographic Aspects of Recrystallization (N. Hansen, D. Juul Jensen, Y. L. Liu, and B. Ralph, eds.), RISØ Nat. Lab, Roskilde, Denmark, 1995, pp. 461–466.
26. V. Marx, F. R. Reher, and G. Gottstein, Acta Mater. 1998, 47, pp. 1219–1230.
27. C. H. J. Davies, Scripta Mater. 1997, 36, pp. 35–46.
28. C. H. J. Davies and L. Hong, Scripta Mater. 1999, 40, pp. 1145–1152.
29. D. Raabe, Z. Zhao, and W. Mao, Acta Mater. 2002 (in press).
30. R. C. Becker, Acta Metall. Mater. 1991, 39, pp. 1211–1230.
31. R. C. Becker and S. Panchanadeeswaran, Acta Metall. Mater. 1995, 43, pp. 2701−2719.
32. J. H. Driver, A. Skali, and M. Wintererger, Phil. Mag. 1984, 49, pp. 505–526.
33. A. Skalli, R. Fortunier, and J. H. Driver, Acta Metall. 1985, 33, pp. 997–1012.
34. L. Delannay, O. V. Mishin, D. Juul Jensen, and P. Van Houtte, Acta Mater. 2001, 49, pp. 2441–2462.

35. D. Raabe, Z. Zhao, S. J. Park, and F. Roters, Acta Mater. 2002, 50, pp. 421–440.
36. Z. Zhao, W. Mao, and D. Raabe, Materials Science Forum, Proceedings of the 13th International Conference on Textures of Materials ICOTOM 13, Seoul, South Korea, 2002, 408, pp. 281–286.
37. D. Raabe and R. Becker, Modell. Simul. Mater. Sci. Eng. 2000, 8, pp. 445–462.
38. D. Raabe, Comput. Mater. Sci. 2000, 19, pp. 13–26.
39. D. Raabe and R. C. Becker, Proceedings of the 12th International Conference on Textures of Materials ICOTOM 12, Montreal, Canada, 1999 (Jerzy A. Szpunar, ed.), Vol. 2, NRC Research Press, National Research Council of Canada, Building M-55, Ottawa, Canada, ON K1A 0R6, 1999, pp. 1112−1117.
40. M. Sachtleber, Z. Zhao, and D. Raabe, Acta Mater., (in press).
41. W. F. Hosford, *The Mechanics of Crystals and Textured Polycrystals*, Oxford University Press, Oxford, 1993.
42. U. F. Kocks, C. N. Tóme, and H.-R. Wenk, *Texture and Anisotropy*, Cambridge University Press, Cambridge, 1998.
43. H.-J. Bunge, Kristall u.Technik 1970, 5, pp. 145–162 (in German).
44. G. I. J. Taylor, Inst. Metals 1938, 61, pp. 307–322.
45. J. F. W. Bishop and R. Hill, Phil. Mag. 1951, 42, pp. 414–432.
46. G. Z. Sachs, Z. Verein Deutscher Ing. 1928, 72, pp. 734–746.
47. R. V. von Mises, Zeitschr Angew Math. Mech. 1928, 8, pp. 161–180.
48. R. Hill, Proc. Roy. Soc. 1948, 193, pp. 281–296.
49. R. Hill, J. Mech. Phys. Solids 1990, 38, pp. 405–422.
50. F. Barlat and J. Lian, Int. J. Plasticity 1989, 5, pp. 51–64.
51. P. Van Houtte, K. Mols, B. Van Bael, and E. Aernoudt, Textures Microstruct. 1989, 11, pp. 23–44.
52. A. Van Bael, P. Van Houtte, E. Aernoudt, F. R. All, L. Pillinger, P. Hartley, and C. E. N. Sturgess, Textures Microstruct. 1991, 14–18, pp. 1007–1013.
53. D. Raabe, Ann. Rev. Mater. Res. 2002 (in press).
54. K. Lücke, J. Pospiech, K. H. Virnich, and J. Jura, Acta Metall. 1980, 29, pp. 167–182.
55. K. Helming, R. A. Schwarzer, B. Rauschenbach, S. Geier, B. Leiss, H. Wenk, K. Ullemeier, and J. Heinitz, Z. Metallkd. 1994, 85, pp. 545–562.
56. K. Helming, *Texturapproximation durch Modellkomponenten*, Cuvillier Verlag, Göttingen, Germany, 1996 (in German).

9

Property Prediction

JAMES T. STALEY

Consultant, Durham, North Carolina, U.S.A.

ROBERT E. SANDERS, JR.

Aluminum Corporation of America, Alcoa Center, Pennsylvania, U.S.A.

1 INTRODUCTION

Methods have been developed to predict the properties of aluminum alloy products for a number of reasons. One reason is to minimize the number of iterations required to develop a new material. Another is to predict effects of process changes dictated by the installation of proposed new equipment. Yet another reason is to determine the relative sensitivities of different unit processes on the resultant properties. Almost every physical property is structure-sensitive. Consequently, one could argue that the ideal model to predict mechanical properties would consist of several parts that could be chained together. The first would predict structure in the cast state. It would have as inputs, chemical composition and thermal history. The output would be morphology and chemical composition of the intermetallic particles that formed during solidification and the gradient caused by both microsegregation and macrosegregation. Successive models would use the input from the preceding model along with thermomechanical sequences to predict the resultant structure with respect to all relevant chemical and topological characteristics including quantity and spatial distribution. The microstructural data would be inputs to models that would predict such things as the stress-strain curve, K_{Ic}, and the forming limit diagram. The state of the art, however, is that few structural features can be modeled from composition and thermomechanical inputs, and the outputs are not detailed to the extent required. Consequently, the procedures used now to predict mechanical properties rely heavily on empirically derived relationships. In many instances, the data are merely related to composition or processing parameters with no regard to mechanisms. In others, mechanisms are postulated, equations describing these relationships are taken from the literature, and constants in the more fundamental equations are calculated from the data. Another approach is to combine measured

microstructural information and either predict certain properties directly or to combine the microstructural information with data from simpler tests to predict properties such as the yield surface. Effects of particular structural features depend on whether the alloy is heat-treatable or non heat-treatable. In general, the smallest feature that interacts with dislocations has the major effect. Microstructural features affecting properties and examples of several approaches are presented below.

1.1 Non Heat-Treatable Aluminum Alloy Products

The property most often quoted for non heat-treatable (NHT) alloy products is some form of strength parameter. Practically any NHT alloy product will be required to meet some strength standard, either to insure structural performance or to provide some estimate of formability. An ability to predict strength from the input alloy and process parameters allows the aluminum producer to provide a more robust and consistent product to the end customer. The 0.2% yield strength is targeted where the onset of material deformation is critical. Examples are beverage can sheet, fin stock, or sheet used in transportation applications requiring dent resistance. The tensile strength may be more useful to predict in other applications where formability as well as strength is required. Some O-temper products specify maximum tensile strengths. In fact the Aluminum Association system for defining property limits for NHT alloys in H1x (work-hardened) tempers is based on percentages of tensile strength achieved by the working operation.

1.1.1 Strength of NHT Products

These alloys are strengthened by a combination of solid solution strengthening, dispersion strengthening, and cold work. Grain size has a modest effect. Examples of graphical and statistical methods to predict properties are presented.

Strength of Material in Annealed Temper
Early workers determined effects of impurity and alloying elements on the strength of aluminum. The principal impurities remaining after smelting and refining are iron and silicon. These impurities increase the strength of 99.999% Al by solid solution strengthening at low levels and dispersion strengthening at higher levels. Effects of these elements on yield and ultimate tensile strengths of material in the annealed temper are shown in Fig. 1. While these effects are undoubtedly significant in the low solute 1xxx alloys, they do not affect strength markedly in the higher strength structural alloys.

Managanese and magnesium form the basis of the 3xxx and 5xxx non-heat treatable alloys. Manganese levels up to about 1% progressively increase strength by a combination of solid solution and dispersion strengthening. Additions of magnesium provide solid solution strengthening. Combined effects of magnesium between 2% and 7.3% and manganese up to 0.9% are shown in Fig. 2.

Regression analyses of data has been used to predict effects of chemical composition on strengths of non heat treatable alloy products. Shabel [1] considered data from 27 experimentally produced compositions of Al-Mg-Mn sheet containing 1.4–4.5% Mg, up to 0.43% Mn, and total Fe + Si levels from 0.01% to 0.65%.
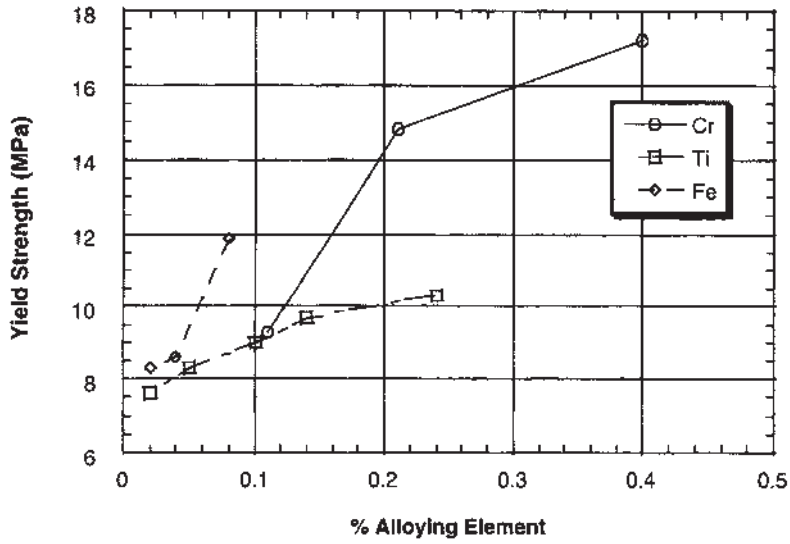
Figure 1 Effects of selected impurities on mechanical properties of high purity aluminum. (Courtesy HC Stumpf, Alcoa Technical Center, Alcoa Center, PA)
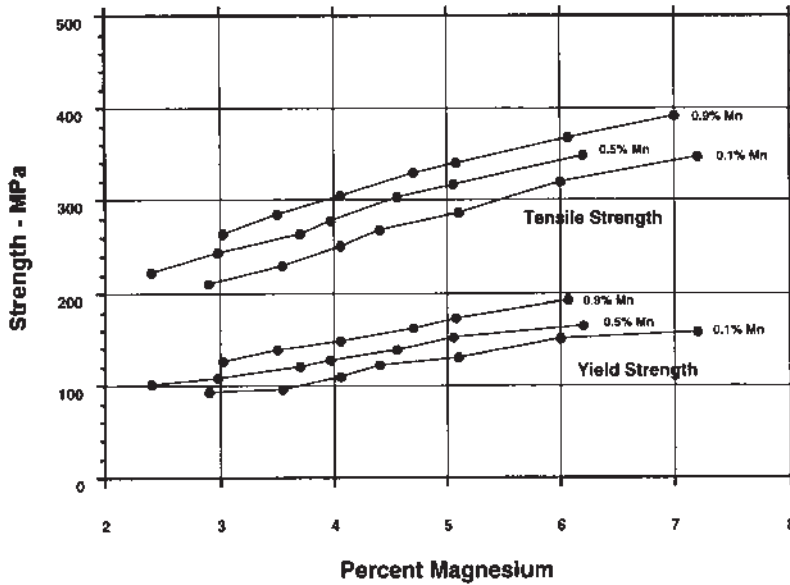
Figure 2 Effects of Mg and Mn on strengthening of annealed NHT aluminum alloys. (Courtesy of WA Anderson, Alcoa Technical Center, Alcoa Center, PA)

The following equation best fits the data:

$$YS(ksi) = 1.1608 + 5.9149\,Mg - 0.27475\,Mg^2 + 1.9059\,logten\,(Fe+Si) + 5.9410\,Mn$$
$$\sigma = 1.093 \qquad\qquad R^2 = 92.64\%$$

$$(1)$$

About 78% of the values fell within one ksi predicted by this equation, and 99% fell within 3 ksi.

Another approach to predicting strength is to consider typical properties and compositions of commercially produced material. Lyle [2] used typical values from the Aluminum Association and ASM Handbooks for 16 3xxx and 5xxx alloys. He did not include Fe and Si in his analysis. The following equation best fits the data:

$$YS(ksi) = 3.444 + 3.299\,Mg + 2.750\,Mn + 6.688\,Cr \qquad\qquad (2)$$
$$\sigma = 0.923 \qquad\qquad R^2 = 97.55\%$$

About 81% of the data fell within one ksi of the prediction.

These types of regression analyses have been useful in providing practical guidance for metallurgists to select or modify alloys, but they illustrate the dangers of relying on a composition-strength relationship instead of a composition-microstructure-strength understanding. These types of correlations typically are much less reliable when attempts are made to apply them to situations where the prior casting or deformation history is dramatically different from the data base. Hence, the need to base the models on microstructure comes into play once again. One example of this problem is exemplified by the impact of grain size on yield strength of O-temper aluminum alloys. While the Hall–Petch effect is normally thought of as rather weak in high stacking fault energy alloys such as Al, Fig. 3 shows that grain size between 14 and 30 microns can affect yield strength 3xxx-O and 5xxx-O sheet [3]. This kind of variation could make the difference between a product failing or passing a minimum yield strength specification.

Strength of Material in H1 Temper

To achieve useful engineering properties the NHT aluminum alloys are most often strengthened by the high levels of cold work which introduce dislocations into the microstructure. The work hardening may be accomplished by stretching, cold rolling, drawing, or by the customer's metal working operation. The traditional way of looking at work hardening in aluminum alloys has been the use of a percentage cold work vs strength curve such as that shown in Fig. 4. Mathematical descriptions of these curves have usually taken the form of either the Holloman equation:

$$\sigma = K\,\varepsilon^n \qquad\qquad (3)$$

or the Ludwig equation:

$$\sigma = \sigma_a + K\,\varepsilon^n \qquad\qquad (4)$$
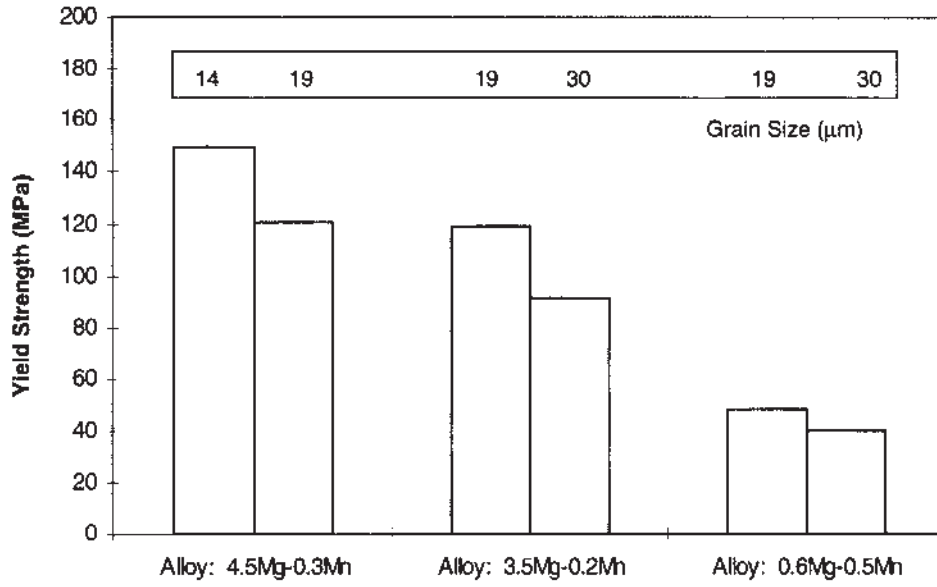
Figure 3 Effect of grain size on strength of selected Al-Mg-Mn alloys in the annealed temper. (Courtesy of SL Palmer, Alcoa Technical Center, Alcoa Center, PA)
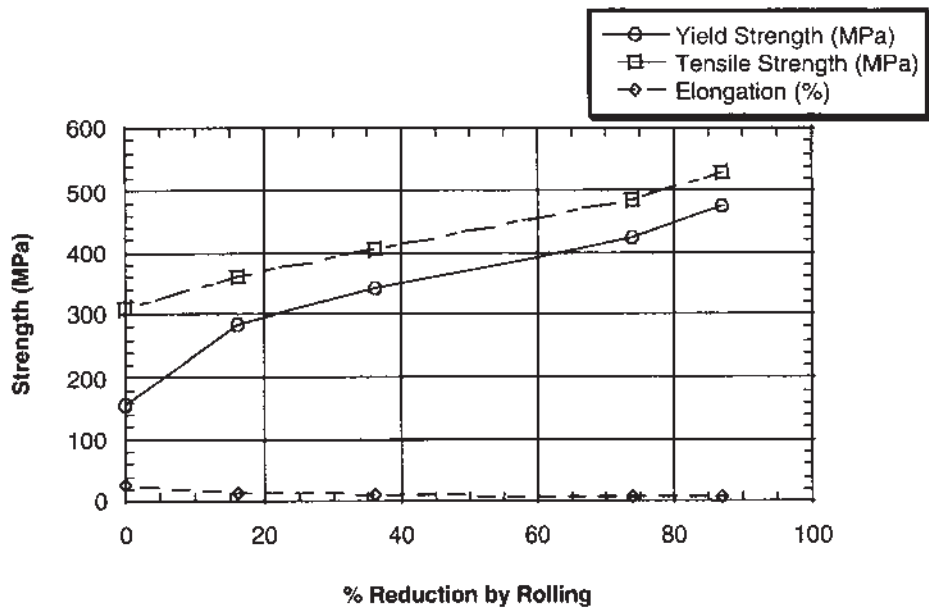
Figure 4 Typical work hardening (strength vs % cold reduction by rolling) curve for a commercial actually 5xxx aluminum alloy.
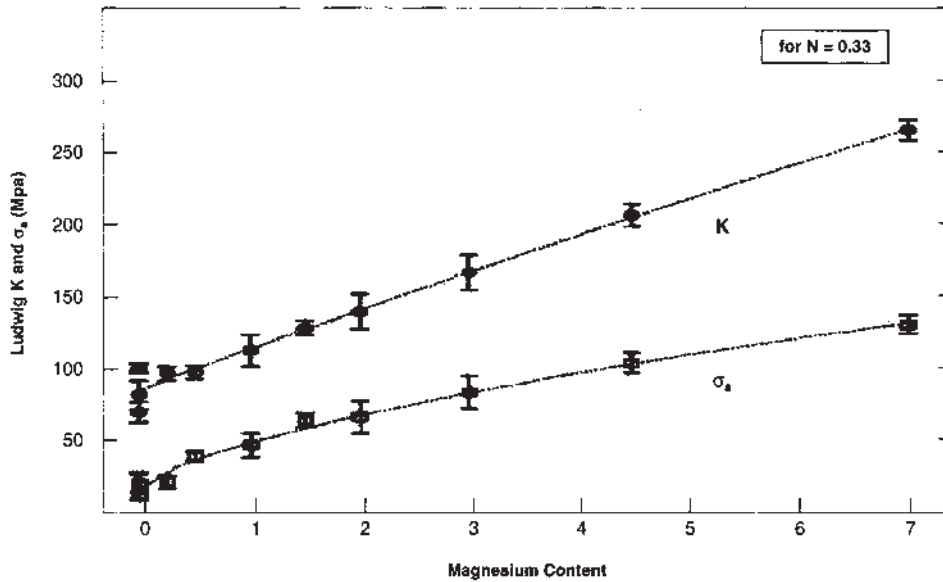
Figure 5 Impact of Mg content on *K* values (at constant *n* value) during work hardening of aluminum sheet. (Courtesy of JP Suni, Alcoa Technical Center, Alcoa Center, PA)

The results can be most conveniently plotted as a log stress vs log strain curve. They have been quite useful in predicting mill loads and rolling forces for engineering purposes. The later formulation is conceptually appealing in that the work hardened contribution to the total strength can be separated from the underformed (O temper) strength.

When data for numerous aluminum alloys are compared the conclusion is that the K and σ_a values are increased much more than the in value. Magnesium and other elements in solid solution have a dramatic impact on these values as pointed out by Suni et al. [4]. The impact of Mg on strength development with cold work can be seen in Fig. 5. The qualitative aspects of microstructure responsible for these rankings are pretty well understood. The relative effects of the different solute elements on dynamic recovery during cold deformation can explain the observed behavior [5]. However, quantitative understanding of solute-dislocation interactions has not progressed to the point where it can contribute significantly to improve the accuracy of strength predictions based on the above expressions.

One area of progress noted in the quantitative predictions which emerge is the interaction of solutes on the work hardening behavior. Doherty and McBride [6] pointed out the enhancement in work hardening behavior observed when certain combinations of solute are present. In industrial practice, a practical problem arises in dealing with models which invoke solute enhanced work hardening: the knowledge of whether a solute is in or out of solution prior to cold working. Alloys containing Mg and Cu, when processed to allow precipitation prior to work hardening will not increase in strength as predicted in the models. The data in Table 1 illustrate the point for two 3004-H38 alloy sheet materials processed with various annealing practices. Anneal 1 was designed to maximize the amount of Cu and Mg in solution

Table 1 Influence of Cu content and thermal treatment on solute retention and strength development in 3xxx cold rolled and stabilized sheet

	423°C Anneal		371°C Anneal	
Annealing Conditions	0.13 Cu Yield Strength (MPa)	0.22 Cu Yield Strength (MPa)	0.13 Cu Yield Strength (MPa)	0.22 Cu Yield Strength (MPa)
Soak 2 hr/CWQ	250	261	255	269
Soak 2 hr Cool 36°C/hr	250	250	253	266
Soak 2 hr Cool 8.3° C/hr	250	241	255	262

Base Alloy: 0.2 Si, 0.4 Fe, 1.1 Mn, 1.2 Mg. Cold rolled 88% after annealing. Baked for 20 min at 204°C.

before cold rolling, where the other two anneals allowed precipitation of an Al-Cu-Mg phase by soaking below the alloy solvus and/or slow cooling. The alloy with the higher Cu level is much more sensitive to the annealing conditions.

This concept of using processing and alloy additions to maximize the strength of 3xxx alloy sheet has been exploited in the introduction of "bake-hardened" can body stock [7]. The process retains higher amounts of Mg, Cu, and Si in solution by means of a continuous anneal and rapid cool. This allows high strength to be obtained with lower amounts of cold work, and, after simulated can baking, a larger percentage of strength of the bake hardening sheet is retained compared to conventionally processed sheet. This bake hardening effect has been explained by either of two mechanisms: precipitation of A1-Cu-Mg-Si phases which compensates for recovery during the bake [8] or by an enhanced resistance to recovery brought about by the extra solute [9].

Researchers have recognized for some time the important influence of prior dislocation structure on the further work hardening of an aluminum product. Morris [10] showed that extremely high levels of strength could be reached in NHT alloys by progressive cycles of work hardening-recovery-work hardening etc. A schematic of this "recovery-enhanced work hardening" process is shown in Fig. 6. This phenomenon, while undoubtedly related to the ability of the microstructure to store dislocations, has not been studied to the point where it can be used as a predictive model for industrial alloys or commercial processes.

Strength of Material in H2 and H3 Tempers
Because of the loss in ductility in work hardened alloys, many NHT products are given a partial anneal (H2x) or stabilization (H3x) treatment to improve the workability of products requiring forming by the end customer. The reduction in dislocation density that occurs during recovery is well known and has been the subject of many academic studies since the late 1940s. The use of meta-recovery (annealing of statistically necessary dislocations) and ortho-recovery (more sluggish process of subgrain growth) to describe the two different regimes has been helpful in understanding how to approach the problem of predicting strength in partially annealed aluminum alloys.
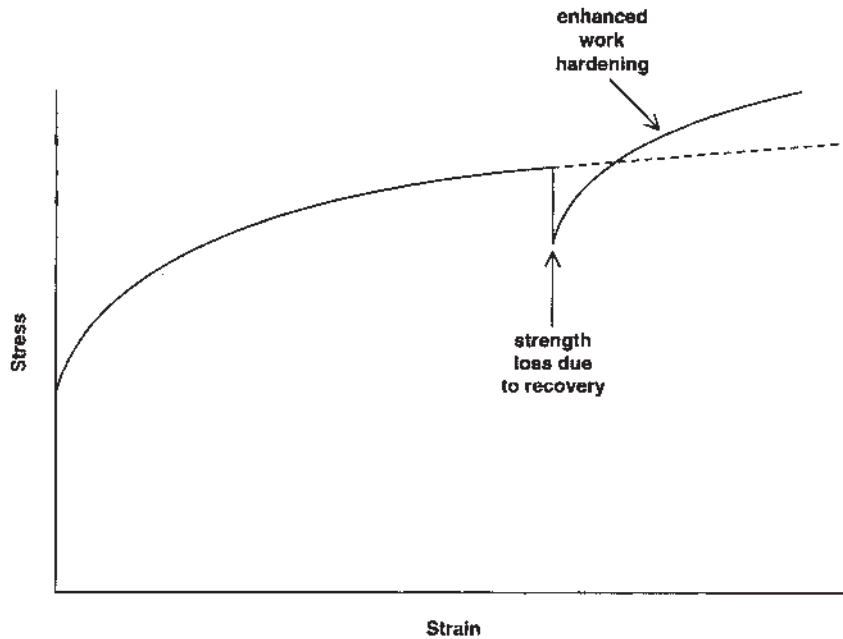
Figure 6   Schematic diagram showing enhanced work hardening of all alloy sheet subjected to an intermediate recovery treatment.

Because meta-recovery occurs very rapidly with the onset of thermal exposure, its importance must be in the prediction of properties after relatively low strains (where a subgrain substructure has not fully developed) followed by thermal exposure. To these authors' knowledge this recovery regime has not been studied extensively for the prediction of strength in aluminum alloys.

On the other hand the ortho-recovery process is more useful for predicting properties where high strains have resulted in well-developed substructures. Kuhlmann [11] demonstrated that the decay of mechanical properties in work hardened aluminum alloys progressed logarithmically with time. The thermal activation of the process kinetics is well modeled, but not always completely understood. Nes [12] and others argue that an explanation of ortho-recovery due to subgrain coalescence is too simplistic and that some account must be made of the "sharpening" of the deformation texture by the recovery process.

Even without complete microstructural agreement on mechanical mechanisms, the logarithmic nature of recovery can be quite useful in predicting properties of cold worked aluminum after thermal treatments. Partial anneal curves have been used within the industry for years to design furnace practices to hit certain properties. More recently examples of models [13,14] which successfully predict strength for an alloy with a given thermal history are common. However, the "complete" model that can account for variations in upstream processing history through the microstructure does not currently exist.

1.1.2 Predicting Texture of NHT Products

The deformation processes used to produce NHT alloy flat rolled products almost invariably result in anisotropy of strength and ductility parameters. This result may impact the structural application of the product, but most often has consequences for the forming behavior of the product in a customer's manufacturing operation. While the types of forming processes used to manufacture aluminum final products vary dramatically, the goal of finding a mechanism for predicting forming behavior of NHT alloys has been elusive since many aspects of the microstructure (in addition to those which impact strength) come into play. Second phase constituent particles which affect fracture can limit elongation and reduce formability without affecting uniaxial strength parameters. The entire thermo-mechanical history can influence the grain size and texture (preferred orientation) of a sheet product and change formability with little impact on strength. The challenge: to get beyond predicting strength, you must understand the entire microstructure.

Statistical models such as the composition/strength equations presented earlier may be useful for process control but, since they are not physically based, do not offer much predictive capability when process parameters are significantly changed. Thus, in going beyond simple strength/cold work correlations, based on dislocation density, the relationship of the product microstructure to the process must be completely understood to effectively predict formability.

The task of predicting texture must address the problems of predicting both recrystallization behavior and the strengthening of deformation texture components which occurs when NHT alloys are cold worked. These problems have been the subject of numerous fundamental scientific investigations. Interestingly, the industrial problem driving many of these studies has been the desire to predict the earing behavior of rigid container sheet. The earing phenomenon, while not strictly a forming issue, is an irritant to the makers of aluminum cans in that the uneven peaks and valley present after drawing (Fig. 7) may require extra trim scrap or lead to loss of productivity during canmaking.

Using the example of earing prediction for aluminum cans as an example, we can examine the state of our ability to use physically based models to make an industrially relevant prediction for earing in alloy 3004 (Al-Mg-Mn) can body stock. Most commercial processes for making aluminum can sheet involve casting a large DC (direct chill) ingot, homogenizing at a high temperature (550–600°C), hot rolling to an intermediate thickness, annealing and cold rolling to final thickness (typically in the 0.25–0.30 mm range). The process is low cost and efficient but requires extremely close process control to insure that the product supplied to the canmaker is consistent in thickness, strength, earing, etc. The microstructure of the sheet, shown in Fig. 8, is a result of all of the steps in the process.

Efforts of aluminum producers [15] to provide consistent attributes in can sheet point to the need to understand structure evolution throughout the process quite clearly. The efforts of workers in the Pechiney organization to link together models for property predictions are described in detail in reference [13]. Though the example used is not 3×04 can sheet, the prediction used in the reference clearly demonstrates the logical linking together of models for predicting end properties of NHT aluminum alloy sheet.
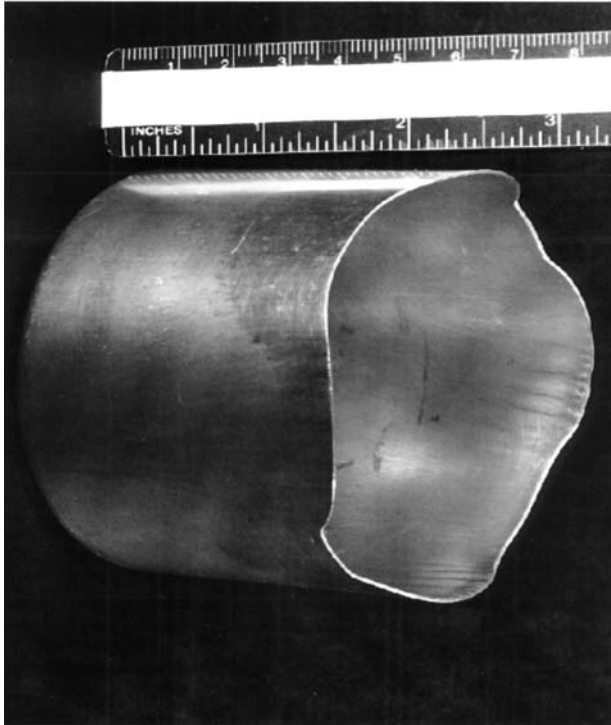
Figure 7   Photograph of earing phenomenon in 3xxx aluminum alloy drawn cups.
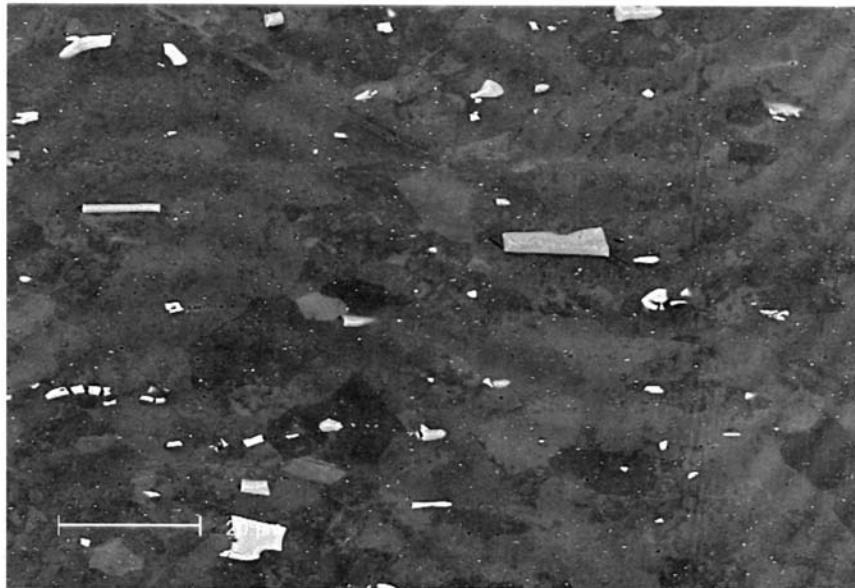
Figure 8   Typical microstructure of alloy 3004-H19 sheet showing constituent and dispersoid intermetallic phases. (Courtesy of TN Rouns, Alcoa Technical Center, Alcoa Center, PA)

Casting models such as ALSTRUC [16] have helped describe the as-cast microstructure of ingot based on solidification rate and composition. It provides a starting point for other models by mapping out the distribution and size of second phase constituent particles after casting and quantifying the solid solution content of and ingot. These models also have the important capability to predict through thickness differences in microstructure caused by variation in cooling rate. These models have been validated primarily for solidification rate typical of large DC ingot, and whether they would be able to predict successfully structure in a much higher rate process such as roll casting is not known.

Extensions of ALSTRUC [17] and other work [18,19] have made it possible to evolve the microstructure through preheating or homogenization processes. The 3xxx alloys have been especially well studied due to their importance in the container industry. The dissolution of phases, transformations of existing particles, and solute changes within the microstructure can be predicted by models such as those developed by Suni and Shuey [18,19]. These models use known diffusion relationships and considerable quantities of experimental data to tune the predictions. While dissolution of existing phases is reasonably easy to handle, heating paths that require significant nucleation of new dispersoids are more difficult to predict. The other problem inherent in these models is that solute species such as Mn which form dispersoids during preheating are not uniformly distributed in the cast ingot. The low diffusivity of Mn and coring from the ingot combine to produce extremely inhomogeneous spatial distributions of particles.

The key to predicting anisotropy of the final H19 sheet is understanding the type of preferred orientation that exists after hot rolling is completed. The earing pattern is a result of the annealing texture developed after the hot mill anneal and the rolling texture introduced by cold rolling the sheet to the H19 temper. Understanding of the cube (100)[100] texture is critical: maximizing this texture after rolling is desirable because it mitigates the strong deformation texture due to cold rolling.

The origin of the cube texture itself was in considerable debate for many years, but recently the concept that pre-existing bands of near-cube orientation are stable during rolling, recover quickly, and gain a size advantage during annealing seems to be nearly universally accepted. The problem of cube texture prediction in commercial processing would not be so troublesome were it not for the fact that hot rolling is a complex series of deformation passes interrupted by opportunities for recrystallization. Studies by Sellars [20], Nes and Vatne [21], and Hirsch [22], have addressed the problem and have begun to show some promise for multi-pass rolling histories. The success of these models has come from the understanding that different texture components nucleate and grow from different sites in the microstructure. Vatne and Nes, [23] for instance, use the competition among three sites (cube, constituent particles, and grain boundaries) under different deformation conditions to predict the texture which evolves from a multi-pass hot rolling process.

However, the studies cited above are typically good only for a narrow range of composition and processing conditions. When process conditions or alloy compositions vary significantly from the standard product, most of the models tend to lose their accuracy. The successful prediction of properties (based on microstructure) is going to require that the entire fabrication process be successfully modeled from casting to final deformation passes.

Numerous workers have pursued the next link in the chain of predicting NHT sheet properties: the development of cold rolling textures [24]. While there has been considerable progress made in understanding deformation textures on a dislocation substructure level at relatively low amounts of strain, the ability to actually predict the cold rolled texture in highly deformed (> 80% reductions) sheet of commercial alloys has not yet been accomplished. The presence of large amounts of shear and strain gradients, high volume fractions of non-deformable particles, and variations in incoming textures make this part of the property prediction chain perhaps the most difficult of all.

While our ability to predict texture evolution is perhaps suspect to date, our abilities to measure and quantify preferred orientations in rolled sheet products has progressed rapidly over the last 20 years. The development of ODF (orientation distribution functions) and other means of describing in detail the texture components present in sheet are now much more well-developed. Models by Barlat and co-workers [25] have utilized the capabilities to predict yield surfaces of sheet product, and with the use of mechanical property data construct forming limit diagrams or earing height predictions. These capabilities represent the endpoint in connecting the microstructure as it has evolved through the process to final properties. In making such predictions from texture data, it must be recognized that through thickness texture gradients exist in most sheet products. Thus, some means of measuring those gradients and integrating them to make the prediction are obvious improvements for the future. The next stages of texture measurement, local orientation imaging techniques, are now making it even more feasible to understand the deformation processes on a grain by grain scale.

1.1.3 Summary of Predicting Properties of NHT Products

Existing relationships provide good rules of thumb to describe work hardening for engineering use. Interactions among solutes in more complex alloys and the contribution of grain size to strengthening should be understood more fully to make reliable strength predictions. For industrial rolling, where exit temperatures can be appreciably above room temperature, recovery in large coils must be taken into account for multi-pass schedules.

While the ability to predict simple strength properties for NHT alloys based on work hardening and recovery phenomena are reasonably well developed, the capability to predict other properties is not yet comparable. Improvements in our capabilities must be rooted in the capability to predict microstructure and, in turn, make the connections to properties. Significant progress has been made in various aspects of the microstructure prediction problem, but efforts to link them together have been few. Examples are shown from the case of 3×04 can stock to illustrate the current level of capability and the possibilities for future development.

1.2 Heat-Treatable Alloys

When considering the factors which affect properties of heat-treatable alloy products, i.e. chemical composition, grain structure, and intermetallic particles (constituents, dispersoids, and precipitates) precipitates have, by far, the major effect on most properties. The effect of precipitates is substantially higher than that possible by solution strengthening. For example, the strength increase attributable to

solid solution strengthening with increasing the solute content of Al-Zn-Mg-Cu alloys from $<4\%$ to $>12\%$ was 130 MPa, while strength attributable to precipitation hardening was at least 330 MPa. All models predict that strength increases with increasing volume fraction of precipitates, so it follows that strength increases with increasing solute content. The precipitates may either be cut or looped by dislocations depending on their size, aspect ratio, elastic properties, strength, and distance between particles. Strength increases with increasing size when the particles are being cut, and decreases with increasing spacing when particles are being looped. At peak strength, some particles are being cut while others are being looped. Models based on first principles are inadequate to predict properties to the degree of accuracy required for industrial use. Examples of phenomenological modeling are presented.

1.2.1 Predicting Effects of Quench Path on Properties of Heat-Treatable Alloy Products

The quenching process has significant effects on properties of heat-treatable alloy products. Because of the sloping solvus curve, the alloying elements are supersaturated below the solution heat treatment temperature. This supersaturation imposes a driving force for precipitation during the quench which **increases** with decreasing temperature. Diffusion kinetics **decrease** with decreasing temperature, however, so the rate of precipitation is highest at temperatures between the solution heart treatment temperature and room temperature. The temperature range for most rapid precipitation extends from about 500°F to 800°F for most commercial aluminum alloys, and it is unique for each alloy. The precipitates formed in this temperature range nucleate at grain corners and edges and at the interfaces between the matrix and intermetallic particles, particularly dispersoids. They are large, equilibrium precipitates and are incoherent with the matrix, so they exert little influence on the motion of dislocations. Therefore, they have little effect on strength. Solute atoms which are in the form of coarse equilibrium precipitates are unavailable to participate in subsequent precipitation (aging) treatments to produce fine, coherent precipitates. Consequently, quenching practices which permit precipitation to occur during the quench result in products with lower strength. Because equilibrium precipitates nucleate at grain boundaries, they deplete the nearby matrix of solute atoms. This microstructural feature can lead to differences in electrochemical potential between regions near the grain boundaries and the matrix. These differences in potential can render the material susceptible to intergranular corrosion and stress-corrosion cracking. The presence of grain boundary precipitates can also to precipitate-free zones (PFZ) near the grain boundaries. The combination of coarse, brittle particles at the boundaries and a weak PFZ can lead to increased tendency for intergranular fracture. Fracture toughness generally decreases with increasing proportion of intergranular fracture, so quenching practices which allow the formation of grain boundary precipitates leads to decreased fracture toughness.

Effects of quenching on properties were first examined as follows [26,27]. The most critical temperature ranges for affecting properties were determined using interrupted quench techniques. Thin specimens were quenched to temperatures intermediate to the solution temperature and room temperature, held for a series of times, then quenched to room temperature, and aged using the standard practice.

Properties were determined, and the results were plotted on time vs temperature graphs to develop Time-Temperature-Property C-curves. For corrosion resistance, the results were expressed as either pitting, mixed pitting and intergranular, or intergranular. The C-curve described the maximum time that the material could be held at a particular intermediate temperature without becoming susceptible to intergranular corrosion. To develop strength C-curves, strength (either tensile ultimate or yield strengths) was plotted versus hold time for each isothermal temperature and times to develop particular levels of strength were obtained from the strength-hold time curves by interpolation. Iso-strength C-curves were plotted on temperature vs time axes by connecting the interpolated points. A critical temperature range was arbitrarily established as the temperature range where time was 2½ times the time at the nose of the C-curve. To predict properties, a series of specimens were quenched in a number of media, and cooling curves were obtained. From these cooling curves, an average rate to cool within the critical quenching temperature range was obtained. The property of interest was plotted versus this average quench rate.

Work in the US [27] and, particularly, the former USSR [28] has revealed that the shape, position both laterally and vertically of T-T-P C-curves is a strong function of the major alloying elements. In general, increasing their level moves the C-curves to longer times because of the higher supersaturation. Effects on vertical displacement depend on the phase diagram. Dispersoid type also has a large effect. Compared to alloys containing Al_3Zr dispersoids, alloys with $Al_{20}Cu_2Mn_3$ and, particularly, $Al_{12}Mg_2Cr$ translate the strength C-curves to longer times because the dispersoids serve as nucleation sites for equilibrium precipitates. Ease of nucleation depends on the relative lattice parameters of matrix and precipitate, so effects of dispersoid type and species of major alloying element may have a synergistic effect.

Predicting Effects of Quench Path on Corrosion Resistance

Need for a more quantitative method for predicting effects of quench path on properties led to a process known as quench factor analysis. J. W. Cahn [29] has shown that transformation kinetics during continuous cooling can be predicted from an isothermal Time-Temperature-Transformation, TTT, diagram:

$$\xi = \int \frac{dt}{C_{ttt}(T)} \tag{5}$$

where:

$dt =$ differential time

$C_{ttt}(T) =$ critical time from a TTT diagram

When $\xi = 1$, the extent of transformation during continuous cooling will be the amount described by the TTT diagram.

This concept was adapted to predict corrosion mode of 2024-T4 using a TTP C-curve instead of a TTT diagram [30]. Later work showed that the method was able to predict corrosion mode of 7×50 W [31]. As illustrated in Fig. 9, Quench Factor, Q, is determined by summing the time intervals, δt, during quenching that the material was in a temperature range which had an average critical time, Ct,

from the corrosion mode TTP C-curve.

$$Q = \sum \frac{\delta t}{Ct} \qquad (6)$$

When $Q < 1$, the material will be susceptible to corrosion by pitting.

Predicting Effects of Quench Path on Strength

This approach was extended to predict strength of precipitation hardenable aluminum alloy products. The inputs to the model are an isothermal Time-Temperature-Property (TTP) C-curve and either a measured or postulated cooling curve. The C-curve may be described mathematically using constants determined by regression analysis of data obtained from isothermal quenching experiments. The expression below is a version of the one formulated by Evancho and Staley [30] who modified an equation by Cahn [32] describing nucleation:

$$C_t(T) = k_2 \exp\left(\frac{k_3 k_4^2}{RT(k_4 - T)^2}\right) \exp\left(\frac{k_5}{RT}\right) \qquad (7)$$

where:

$C_t(T)$ = critical time at temperature T for attainable strength σ

k_2 = constant which includes the reciprocal of number of potential nucleating sites

k_3 = constant which includes the change in free energy associated with formation of a nucleus

k_4 = constant related to solvus temperature

k_5 = mobility term

R = gas constant.

To describe a particular level of σ, another term, $k_1(T)$, is added to give:

$$C_t(T)_x = k_1(T) k_2 \exp\left(\frac{k_3 k_4^2}{RT(k_4 - T)^2}\right) \exp\left(\frac{k_5}{RT}\right) \qquad (8)$$

where:

$C_t(T)_x$ = critical time for attainable strength σ to decrease to a level where σ equals σ_x

$k_1(T) = -\log_e(\sigma_x)$

σ_x = usually chosen to equal $0.995\,\sigma_{max}$ for plotting

σ_{max} = maximum level of σ

$\sigma_{min}(T)$ = minimum level of σ at T.

The model to predict strength assumes that the rate of loss of solute during either quenching or holding below the solvus is a first order reaction. The phenomenon can be structured in the form of a differential equation [33]:

$$\frac{ds}{dt} = -A(T)(s - s_{\min}(T)) \tag{9}$$

where:

$$s = \text{concentration of solute in solution}$$
$$t = \text{time at temperature, } T$$
$$A(T) = \text{kinetic constant}$$
$$s_{\min}(T) = \text{equilibrium solute concentration at } T.$$

Based on the assumption that attainable yield strength depends linearly on solute concentration, one can rewrite (Eq. (9) as follows:

$$\frac{d\sigma}{dt} = -k(T)(\sigma - \sigma_{\min}(T) \tag{10}$$

For constant temperature, solving this equation gives:

$$\sigma = \sigma_{\min}(T) + (\sigma_{\max} - \sigma_{\min}(T)\exp(-k(T)t) \tag{11}$$

where:

$$\sigma = \text{strength capability after time } t \text{ at temperature } T$$
$$\sigma_{\max} = \text{maximum value of } \sigma$$
$$\sigma_{\min}(T) = \text{minimum value of } \sigma \text{ at temperature } T$$
$$k(T) = \text{kinetic constant at temperature } T.$$

For continuous cooling, the term $k(T)$ can be replaced by (5):

$$k(T) = \left(\frac{1}{C_t(T)}\right) \tag{12}$$

where $C_t(T)$ is as defined in Eq. (8).

This leads to the following equation which describes the attainable property after cooling over any path:

$$\frac{d\sigma}{dt} = -\frac{1}{C_t(T)}(\sigma - \sigma_{\min}(T)) \tag{13}$$

For implementation on a personal computer, the differential equation can be simplified. By analogy with Eq. (5) terms t and $-k(T)$ in the foregoing equations can be replaced by a new term Q, known as quench factor:

$$\sigma = \sigma_{\min}(T) + (\sigma_{\max} - \sigma_{\min}(T))\exp(Q) \tag{14}$$

where:

$$Q = \sum_{t_0}^{t_f} \frac{\Delta t}{C_t(T)} \qquad \text{and} \qquad \begin{array}{l} t_0 = \text{time at the start of the quench} \\ t_f = \text{time at the end of the quench} \end{array}$$

In the computer program an increment ΔQ_i is calculated for each time increment during continuous cooling Δt_i:

$$\Delta Q_i = \frac{\Delta t_i}{C_t(T_i)} \tag{15}$$

The incremental loss in ability to develop strength $\Delta \sigma_i$ is then calculated:

$$\Delta \sigma_i = (\sigma_{\max} - \sigma_{\min}(T_i))(1 - \exp(-\Delta Q)) \tag{16}$$

This value is subtracted from the current value of σ; the new value is output, and the process continue until the quench is complete.

Predicting Effects of Quench Path on Fracture

Fracture toughness generally decreases with decreasing quench rate. This phenomenon has been correlated with an increase in the proportion of intergranular fracture. A quench factor model was developed to predict the ability to develop toughness on aging after continuous cooling. This model is based on the assumption that the instantaneous rate of loss in ability to develop toughness during quenching is related to the amount remaining in solution and to the minimum toughness that would be developed at the current temperature by isothermal quenching. Because the solute in solution is related to the ability to develop yield strength, the toughness model includes terms from the yield strength model. The model assumes that the effects of the morphology of the precipitates which serve to decrease toughness by increasing the proportion of intergranular fracture are accounted for by the loss of toughness determined empirically by isothermal quenching experiments and, therefore, contains no terms which represent microstructure. As described in Ref. 34, rate of loss in the ability to develop toughness given these assumptions can be described by the following differential equation:

$$\frac{dK}{dt} = -\frac{1}{D_t(T)}(K_{\max} - K_{\min}(T))\left[\frac{\sigma - \sigma_{\min}(T)}{\sigma_{\max} - \sigma_{\min}(T)}\right]^{\frac{C_t(T)}{D_t(T)}} \tag{17}$$

where:

$$K = \text{measure of fracture toughness}$$
$$K_{\min}(T) = \text{minimum level of toughness at temperature, } T$$
$$K_{\max} = \text{maximum level of toughness}$$

and

$$D_t(T) = m_2 \exp\left(\frac{k_3 k_4^2}{RT(k_4 - T)^2}\right) \exp\left(\frac{k_5}{RT}\right) \tag{18}$$

where:

$D_t(T) =$ critical time at temperature, T for attainable toughness, K

$m_2 =$ constant which includes reciprocal of number of potential nucleating sites of precipitates which influence K.

For plotting purposes, a term $m_1(T)$ analogous to term $k_1(T)$ can be used.

To implement the model on a personal computer, the incremental loss in toughness ΔK_1 for the initial time increment Δt_1 during continuous cooling was calculated using the following equation:

$$\Delta K_1 = (K_{\max} - K_{\min}(T_1))\left(1 - \exp\left(\frac{\Delta t_1}{D_t(T_1)}\right)\right) \tag{19}$$

and following the reasoning in Reference 34, another form of the equation was used to describe the incremental loss in toughness ΔK_i during continuous cooling over succeeding time increments Δt_i:

$$\Delta K_i = (K_{\max}K_{\min}(T_i))\left[\frac{\sigma_{i-1} - \sigma_{\min}(T_i)}{\sigma_{\max} - \sigma_{\min}(T_i)}\right]^{\frac{C_t(T_i)}{D_t(T_i)}}\left[1 - \exp\left(\frac{\Delta t_i}{D_t(T_i)}\right)\right] \tag{20}$$

1.2.2 Predicting Effects of Aging on Properties of Heat-Treatable Alloy Products

Immediately after quenching, the properties of heat-treatable alloy products are similar to those of NHT products in O tempers. Within minutes, however, properties start to change as G-P zones develop at room temperature. Aging at elevated temperatures not only accelerates the rate of change in properties; because the nature of the precipitates are different from those developed at room temperature, combinations of properties are different.

Predicting Effects of Aging on Strength of Heat-Treatable Alloy Products
The generally higher ratio of strength-to-weight of heat-treatable alloy products relative to those of NHT products prompted early efforts to predict strength.

PREDICTING EFFECTS OF AGING ON STRENGTH OF MATERIAL IN T4 TEMPERS.   Many aluminum alloy products are cold worked after the quench from the solution heat treatment temperature either to decrease the level of residual stress or to straighten the material. This operation becomes more difficult with increasing time at room temperature because of the strength increase due to the nucleation and growth of GP zones. Predicting the strength after various periods of natural aging can be important when factors such as safety and capability of the stretching equipment are considered. Natural aging curves for tensile properties of several alloys are available in the literature [35,36], and stress to initiate yielding can be estimated from these curves. When the stress to produce higher levels of plastic strain must be known accurately, however, chemical composition must be considered and the stress strain curve must be estimated. Regression analysis of a number of compositions in a statistically designed experiment has been used to predict effects of chemical composition and natural aging time on flow stress of 7xxx-W plate, but no predictive model has been published.

PREDICTING EFFECTS OF AGING ON STRENGTH OF MATERIAL IN T3 TEMPERS.   The interaction between cold work and precipitates in natural aged tempers manifests itself in strength properties. Metallurgists have observed since the early days of producing 2024-T3 and 2024-T351, and more recently in producing 2324-T39, that time at room temperature between quenching and stretching affects strength, but the mechanism is uncertain. Consequently, experiments must be used to determine the magnitude of the effects.

PREDICTING EFFECTS OF AGING ON STRENGTH OF MATERIAL IN T6 AND T8 TEMPERS.   Regression analyses have been used to predict properties as function of alloying content and processing parameters to reduce the variability of existing materials and to speed alloy development. These results are proprietary to the producers, so the models are not generally available. For incremental alloy development, statistically designed experiments are performed when the existing data base is insufficient to enable sufficiently accurate predictions. Interval between quenching and aging to peak strength can have a large effect which is alloy dependent. Theories based either on thermodynamics or kinetics can predict such effects qualitatively, but no model has been advanced which allows one to predict, a priori, the magnitude of the effect.

PREDICTING EFFECTS OF AGING ON STRENGTH OF MATERIAL IN T7 TEMPERS.   The T7 tempers for 7XXX alloys were developed to provide improved resistance to stress-corrosion cracking (SCC) and exfoliation corrosion. Early experiments had shown that these characteristics were improved in 7075 products aged above about 150°C. Strengths were less than those attainable with a T6 practice, but an initial low temperature aging step minimized the degradation in strength. Subsequent work showed that an extended natural aging interval and heating rate provided the same benefit (31). The effects are attributed to the same phenomenon. During natural aging, slow heating, or preaging near 110°C, G-P zones nucleate and grow. If they attain a size which is stable at the aging temperature, they will grow and transform into $\eta\prime$ particles having an average size and spacing which provides high strength. If they develop a size distribution which includes many particles which dissolve during aging, however, the resulting coarser $\eta\prime$ size and spacing will provide lower strength. At the higher aging temperatures, some $\eta\prime$ transforms to η. If the $\eta\prime$ is finely dispersed, however, the η will be more finely dispersed than it would have been if nucleated directly. Production workers soon discovered that strength variability was considerably greater than that observed for 7075–T6, so work was done to find and eliminate the cause. Experiments showed that relatively small deviations from nominal aging temperature which had insignificant effects on strength of 7075 aged at 110–120°C had appreciable effects on strength of 7075 aged at 165–175°C. A device which integrated effects of time above about 150°C on loss of strength by overaging was developed using empirically based relationships, and this method was used at Alcoa plants to control aging of 7075–T7 products.

Subsequent research on 7075 and the newer alloy 7050 showed that averaging kinetics of both alloys follow a simple Avrami type relationship [37]

$$\sigma = \sigma_{\max}\exp-\left(\frac{t}{F_{ys}(T)}\right)^{n} \tag{21}$$

where:

$$\sigma = \text{yield strength after overaging for time, } t$$
$$\sigma_{\max} = \text{maximum strength assuming overaging begins at time equals 0}$$
$$F_{ys}(T) = \text{aging time when } \log_e\left(\frac{\sigma}{\sigma_{\max}}\right) = -1$$
$$n = \text{determined by experiment to be 1}$$

The change in $F_{ys}(T)$ with overaging temperature follows an Arrhenius relationship:

$$F_{ys}(T) = A_{ys}\exp\left(\frac{Q_{ys}}{RT}\right) \tag{22}$$

where:

$$A_{ys} = \text{constant}$$
$$Q_{ys} = \text{activation energy for loss in strength}$$
$$R = \text{gas constant}$$
$$T = \text{overaging temperature, degrees absolute}$$

Unpublished work by M. Przystupa of U.C.L.A. has shown that Eq. (21) is consistent with a coarsening mechanism of η during overaging and for an Orowan strengthening mechanism by the η particles.

   Later work with 7075 showed that low-high-low temperature aging practices further minimized the loss in strength by aging at temperatures to develop η while maintaining desired corrosion and SCC characteristics [38,39]. The first step nucleates fine η' which promotes the development of finely dispersed η during the second step which is at a higher temperature than that used in conventional two-step T7-type aging treatments. Consequently, any solute remaining can precipitate during the third step to form η' particles which provide additional strengthening.

Predicting Effects of Aging 7XXX Alloy Products to T7 Tempers on Electrical Conductivity

For similarly processed 7XXX material in T7 tempers, electrical conductivity has been shown to correlate with resistance to exfoliation corrosion and to SCC. The correlation is so useful that electrical conductivity minima have been written into many specifications as lot release criteria for products of 7075 and 7050 in T7 tempers. Experimentation revealed that the change in electrical resistivity (inversely proportional to conductivity) followed a relationship similar to that for yield strength except for the exponent [37]

$$\rho = \rho_{\min} + (\rho_{\max} - \rho_{\min})\exp - \left(\frac{t}{F_\rho(T)}\right)^{0.5} \tag{23}$$

where:

$$\rho = \text{electrical resistivity after overaging for time } t \text{ at temperature } T$$
$$\rho_{min} = \text{minimum resistivity}$$
$$\rho = \text{maximum restivity}$$
$$F_{\rho(T)} = \text{aging time when } \log_e \left(\frac{\rho - \rho_{min}}{\rho_{max} - \rho_{min}} \right) - 1.$$

To account for aging during heating from the first to the second aging step, the effects can be integrated [40].

1.2.3 Predicting Effects of Grain Structure on Properties of Heat-Treatable Alloy Products

Heat treatable alloys which contain dispersoids (almost all) dynamically recover during hot work. During the working operation they deform to produce a grain shape and crystallographic texture which is characteristic of the deformation process. Products which are not subsequently cold worked, i.e. plate, forgings, and extrusions, retain this unrecrystallized structure during solution heat treatment provided that the hot working temperature is high enough during bulk deformation. Products which are cold worked prior to solution heat treatment, such as sheet and tubing, recrystallize during solution heat treatment and lose the deformation texture. Despite recrystallization, many cold worked products, particularly the high-strength members of the 7XXX family, retain the elongated grain structure produced during bulk deformation because the dispersoids pin the grain boundaries. In general, the grains become finer and more equiaxial as the amount of cold work prior to solution heat treatment increases.

Predicting Effects of Grain Structure on Resistance to Exfoliation Corrosion and SCC

Grain shape has a major effect on resistance to exfoliation and stress-corrosion cracking (SCC) of products which are in susceptible tempers. Corrosion attack in susceptible materials proceeds along grain boundaries. When the structure is fine and equiaxial, the intergranular corrosion leads to loss of a few grains. When the structure is elongated, whether recrystallized or not, corrosion follows along the extended grain boundaries and the corrosion products exert a wedging action on the material. This leads to severe exfoliation damage as much of the material peels off.

Resistance to SCC of susceptible materials such as 7075-T6 products is a strong function of grain shape (aspect ratio) and stress orientation. Material with equiaxial grains has low resistance in any stress direction because intergranular cracks can readily follow the boundaries. Material with elongated grain structure, however, is highly resistant in the long-transverse and longitudinal directions because the long dimensions of the grains are perpendicular to the stress. No published information is available relating grain aspect ratio to exfoliation or SCC resistance.

Predicting Effects of Grain Structure on Strength

The major influence of grain structure is its effect on yield stress anisotropy. Strength can be significantly increased in the direction of deformation by development of particular deformation textures, particularly a fiber texture in which a large fraction of grains have their ⟨111⟩ directions oriented parallel to the stress direction [41]. Aging can either increase or decrease the anisotropy through an interaction between precipitates and texture. The extent of the change cannot be predicted at this time.

Predicting Effects of Grain Structure on Fracture Toughness

Grain structure has an effect by influencing the amount of intergranular fracture. In general, fracture toughness of hot worked products decreases with increasing degree of recrystallization. No model is available, however, to predict either the amount of intergranular fracture or the fracture toughness as a function of grain morphology.

1.3 Summary

Predicting properties by combining predictions of chemical composition, solidification and thermomechanical processing on microstructure with predictions of microstructure on properties is still a goal. Currently, however, properties are being predicted based both on statistical analysis and by phenomenological modeling.

REFERENCES

1. B. S. Shabel, Unpublished Alcoa research, 1975.
2. J. P. Lyle, Jr., Unpublished Alcoa research, 1985.
3. S. L. Palmer, Unpublished Alcoa research, 1994.
4. J. P. Suni, R. T. Shuey, and R. D. Doherty, "Solute dependence of cold work strengthening in aluminum alloys," in *Aluminum Alloys – Their Physical and Mechanical Properties*, (T. H. Sanders, Jr. ed.), (ICAA4). Georgia Tech, Atlanta, 1994, pp. 521–528.
5. R. E. Sanders, Jr., H. C. Stumpf, and S. F. Baumann, "Non Heat-Treatable Aluminum Alloys," in *Aluminum Alloys – Their Physical and Mechanical Properties*, (E. A. Starke, Jr. ed.), EMAS, London, 1986, pp. 1441–1484.
6. R. D. Doherty and J. K. McBride, "Solute Enhanced Strain Hardening of Aluminum Alloys," in *Aluminum Alloys for Packaging*, (J. G. Morris, ed.), TMS, 1993, pp. 347–368.
7. Y. Kitao, E. Usui, and T. Inaba, "Can Body Stock for Gage Down," Kobe Res. Dev. 1982, 32, p. 2.
8. C. Sigli, H. Vichery, and B. Grange, "Computer Assisted Metallurgy for Packaging Alloys," in *Aluminum Alloys for Packaging II*, (J. G. Morris et al., ed.), TMS, 1996, pp. 189–198.
9. R. D. Doherty and S. F. Baumann, "Structural Origin of Solute Enhanced Work Hardening of Aluminum Alloys," in *Aluminum Alloys for Packaging*, (J. G. Morris, ed.), TMS, 1993, pp. 369–391.
10. J. G. Morris, Superstrength Aluminum Alloys, Mat. Sci. Eng. 1972, 9, pp. 361–367.
11. D. Kuhlmann, G. Masing, and J. Raffelsieper, "Zur theorie der erholung", Z. Metallkde 1949, 40, pp. 241–246.
12. E. Nes, "Recovery revisited", Acta Metall. Mater. 1995, 43, pp. 2189–2207.
13. G. M. Raynaud, B. Grange and C. Sigli "Structure and Property Control in Non Heat-Treatable Alloys," in *Aluminum Alloys – Their Physical and Mechanical Properties*, (L. Arnberg, et al., ed.), Proceedings ICAA3, 1992, pp. 169–213.
14. L. A. Lalli, Unpublished Alcoa research, 1986.

15. G. J. Marshall, "Microstructural Control During Processing of Aluminum Canning Alloys," in *Aluminum Aloys – Their Physical and Mechanical Properties*, J. H. Driver, et al., ed.), Proceedings ICAA5, 1996 pp. 19–30.

16. Y. Langsrud, A. L. Dons, E. K. Jensen, and S. Brusethaug, "ALSTRUC – A Model for the Calculation of as Cast Microstructures in Industrial Aluminum Alloys," in *Aluminum Alloys – Their Physical and Mechanical Properties*, (L Arnberg, et al., ed.), Proceedings ICAA3, 1992, pp. 15–20.

17. E. Tromberg, A. L. Dons, R. Orsund, E. K. Jensen, and L. Arnberg, Computer Modelling as a Tool for Understanding the Precipitation of Mn in AA3xxx alloys. 1992.

18. J. P. Suni, R. T. Shuey, and R. D. Doherty, "Dispersoid Modeling in 3xxx Alloys," in *Aluminum Alloys for Packaging II*, (J. G. Morris et al., ed.), TMS, 1996, pp. 9, 145–160.

19. J. P. Suni and R. T. Shuey, "Modelling of Dispersoid and Constituent Particle Evolution in 3xxx alloys," in *Aluminum Alloys for Packaging III*, (S. K. Das, ed.), TMS, 1997, pp. 21–38.

20. M. R. van der Winden and C. M. Sellars, "Laboratory Simulation of Breakdown Rolling of AA3104," Mat. Sci. Forum, 1996, pp. 385–390.

21. H. E. Vatne, T. Furu R. Orsund and E. Nes, "Modelling Recrystallization after Hot Deformation of Aluminum," Acta Mater. 1996, 44, pp. 4463–4473.

22. J. Hirsch and R. Grenz, Microstructure Control on an Aluminum Hot Reversing Finishing Mill," in *Hot Deformation of Aluminum Alloys II*, TR Bieler et al., ed.), TMS, 1997, pp. 305–316.

23. H. E. Vatne, T. Furu, and E. Nes, "Microstructural Modelling of Recrystallization Applied to Industrial Processing of Metals," in *Proceedings of Rex '96 Third Int. Conf. on Recrystallization and Related Phenomena*, (T. R. McNelley, ed.), Monterey, Inst. For Advanced Studies, 1997, pp. 95–108.

24. A. D. Rollett and S. I. Wright, "Typical Textures in Metals," in *Texture and Anisotropy*, (U. F. Kocks et al., ed.), Cambridge University Press, UK 1998, pp. 179–238.

25. F. Barlat "Crystallographic Texture, Anisotropic Yield Surfaces and Forming Limits of Sheet Materials," Mat. Sci. Eng. 1987, 91, pp. 55–72.

26. W. L. Fink and L. A. Willey, Trans. AIME, 1948, 175, pp. 414–427.

27. L. A. Willey, in *Aluminum: Vol. I, Properties, Physical Metallurgy, and Phase Diagrams*, (K. R. VanHorne ed.), Cleveland, OH, ASM, 1967, p. 140.

28. G. Vander Voort, ed., *Atlas of Time-Temperature Diagrams for Nonferrous Alloys*. Metals Park, OH, ASM, 1991, pp. 1–42.

29. J. W. Cahn, "Transformation Kinetics During Continuous Cooling", Acta Met. 1956, 4, pp. 572–575.

30. J. W. Evancho and J. T. Staley, "Kinetics of Precipitation in Aluminum Alloys During Continuous Cooling," Metall. Trans. 1974, 5, pp. 43–47.

31. J. T. Staley, S. C. Byrne, E. L. Colvin, and K. P. Kinnear, "Corrosion and Stress-Corrosion of 7XXX-W Products," Proceedings of ICAA5: Aluminium Alloys–Their Physical and Mechanical Properties, Part 3, Grenoble, France, 1–5 July 1996, Materials Science Forum, Switzerland, 1996, pp. 217–222.

32. J. W. Cahn, "The Kinetics of Grain Boundary Nucleated Reactions," Acta Met. 1956, 4, pp. 449–459.

33. J. T. Staley, "Quench Factor Analysis of Aluminum Alloys," Mater. Sci. Tech., 1987, 3, pp. 923–935.

34. J. T. Staley, R. D. Doherty, and A. P. Jaworski, "Improved Model to Predict Properties of Aluminum Alloy Products after Continuous Cooling," Met. Trans. 1993, 24A, pp. 2417–2427.

35. J. E. Hatch, ed., *Aluminum: Properties and Physical Metallurgy*. ASM, Metals Park, OH, 1983.

36. J. R. Davis, ed., *ASM Specialty Handbook Aluminum and Aluminum Alloys*, OH, ASM, Metals Park, 1993.
37. J. T. Staley. Ageing Kinetics of Aluminium Alloy 7050. Metal. Trans. 5:929–932, 1974.
38. B. Cina, USP 3, 8566, 584.
39. M. H. Brown, USP 4, pp. 477–292.
40. J. T. Staley, ''Kinetics for Predicting Effects of Heat Treating Precipitation-Hardenable Aluminum Alloys'', Ind. Heat. 1977, 44, pp. 6–9.
41. K. R. Van Horn, ''Factors Affecting Directional Properties in Aluminum Wrought Products,'' Am. Soc. Met. Trans. 1955, 47, pp. 38–76.

10

Mechanical Properties

D. SCOTT MacKENZIE

Houghton International Incorporated, Valley Forge, Pennsylvania, U.S.A.

1   INTRODUCTION

Mechanical testing is necessary to ensure that components will not fail in service. Testing is performed so that the designer can predict the performance of a part or a component in the field. In this chapter, a brief description of design theories and selected testing methods that are available to the designer to predict component serviceability under realistic conditions are provided. Often, mechanical testing is done for quality control. Since mechanical testing is a dynamic field, new tests are continually being devised, and older tests are being revised, and reinterpreted, for cost or performance reasons. Therefore, a review of the literature should be conducted to conserve materials, achieve reliability, and minimize environmental impact.

The prediction of the behavior of a stressed material is dependent on the applied loading and the relative magnitudes of the principal stresses and strains. The material can fail in either a ductile or brittle fashion. The predictive theory used is dependent on the type of loading and the expected response.

When a component is loaded, it responds either elastically or plastically depending on the material, amount of stress, strain rate, and geometry. In elastic deformation, there is a temporary change in the distances between atoms or crystallographic planes. In plastic deformation, there is a permanent change in the relative position of the atoms. This is caused by the displacement of atoms within the crystal lattice.

If the loads are sufficiently high to cause failure or fracture, the material behaves in either a brittle or ductile fashion. Gensamer [1] summarized the terms used to characterize fracture. These terms and the described behavior are shown in Table 1.

Table 1   Terms Used to Charectorize Fractures

Crystallographic Mode	Shear, Cleavage
Appearance	Fibrous, granular
Strain to fracture	Ductile, brittle

1.1 Types of Loading

The method of loading affects the manner in which a body reacts under stress. A body can experience four types of loading:

1. Static load, short time. The load is gradually applied until failure occurs. The loads on the body are in equilibrium. A tensile test is typical of this type of loading.
2. Static load, long duration. The load is applied gradually until it reaches a maximum and is then maintained. Creep testing is done in this manner.
3. Dynamic loading. The load is applied rapidly, and the loads do not have time to reach equilibrium within the body. The part momentum is considered as the pan is accelerated. The stresses are considered static but are applied in a short time impulse. Charpy impact testing is typical of this type of loading.
4. Repeated loading. Loads are applied to the body repetitively, with the load being either completely or partially removed after each cycle. There are also two main types of loading: high stress applied for a few cycles or low stress applied for many cycles.

Loading of various types may be applied at the same time or sequentially. The order is unimportant unless it is necessary to determine the source of failure. The local stress distribution and the relative magnitudes of the stresses and strains present are also important.

1.2 Stress and Strain

1.2.1 Plane Stress

If a small differential body is stressed (Fig. 1), there are three main stresses normal to the faces of the cube (σ_x, σ_y, and σ_z) and six shear stresses along each face (τ_1, τ_2, τ_3, τ_4, τ_5 and τ_6). If σ_z is much smaller than σ_x and σ_y, then the stress state is a condition called *plane stress*, since the normal stresses of significant magnitude lie in only one plane.

Assume that a triaxial stress state occurs at a uniaxially loaded notched plate (Fig. 2). Because no load is applied to the surface or at the root of the notch, the stress in the z direction is zero. Stress in the z direction rises to a maximum at the center of the plate. If the plate is thin, then the increase in this stress is small relative to stress in the other directions. A plane stress condition then occurs (Fig. 1(a)). This is very common in thin materials such as sheets.
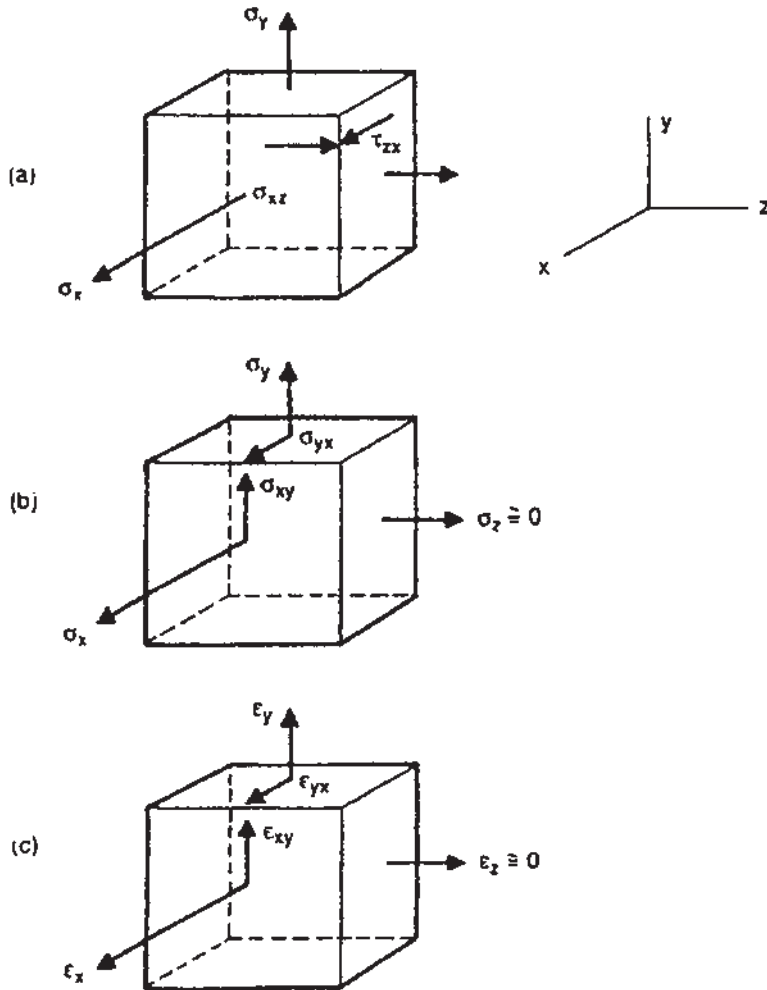
Figure 1 Stresses on a small differential element (a) Principal Stresses; (b) Plane Stress; and (c) Plane Strain.

While the stress in the z direction is small compared to that in the x and y directions, a large maximum shear stress that lies at a 45° angle to the load direction occurs. Plastic deformation occurs by shear along this inclined plane, creating a shear-lip or slant fracture.

1.2.2 Plane Strain

If the strain e in one direction is much smaller than in the other two directions, then a condition of plane strain exists (Fig. 1(c)). If the notched plane is thick, then the strain in the z direction is constrained, resulting in a plane strain condition. Because strain in the z direction is limited, the plastic zone at the notch tip is small. Therefore

Direction of uniform load

Magnitude of Stress in the Z direction
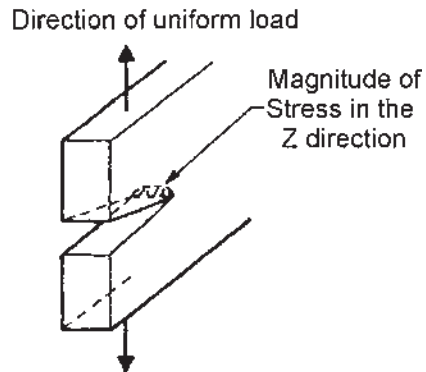
Figure 2   Stress distribution in a notched thin plate, showing stress distribution at a notch.

the axial stresses at the notch tip are large, and the shear stresses are small. Since the axial stresses are much larger than the shear stresses, the axial stresses control the fracture. A flat fracture results that is normal to the axial stresses.

2   PLASTICITY AND DUCTILE FRACTURE

When a material behaves in a ductile or plastic manner, gross deformation occurs. This may manifest as a failure to obey Hooke's law throughout the load history or as plasticity in an initially elastic region, as is typical in a common tensile test. It occurs by portions of grains sliding over one another. Energy is absorbed by the plastic deformation. Ductile fractures are characterized by yielding along the edges of the fracture. The final fracture is usually at 45° to the original plane of fracture. This final fracture is usually called the shear lip.

2.1   Ductile Fracture Appearance

On a macroscopic scale, a ductile fracture is accompanied by a relatively large amount of plastic deformation before the part fails. After failure, the cross-section is reduced or distorted. Shear lips are observed at the latter part of the fracture and indicate the final failure of the part. The fracture surface is dull, with a fibrous appearance.

Microscopically, ductile fracture is characterized by several distinct stages [2–5]:

1. The creation of a free surface from a small particle. This particle can be a second phase particle, dispersoid, or inclusion. This free surface can be formed by the metal matrix separating from the small particle at the matrix/particle interface, or the free surface can be formed by the fracturing of the small particle.
2. The free surface around the small particle creates a void. This void grows by plastic strain and hydrostatic stress.
3. Finally, the voids grow to a size that they join or coalescence with adjacent voids.

Inclusions in Ductile Matrix

Void Nucleation

Void Growth

Strain Localization between voids

Necking Between Voids

Void coalescence and Fracture
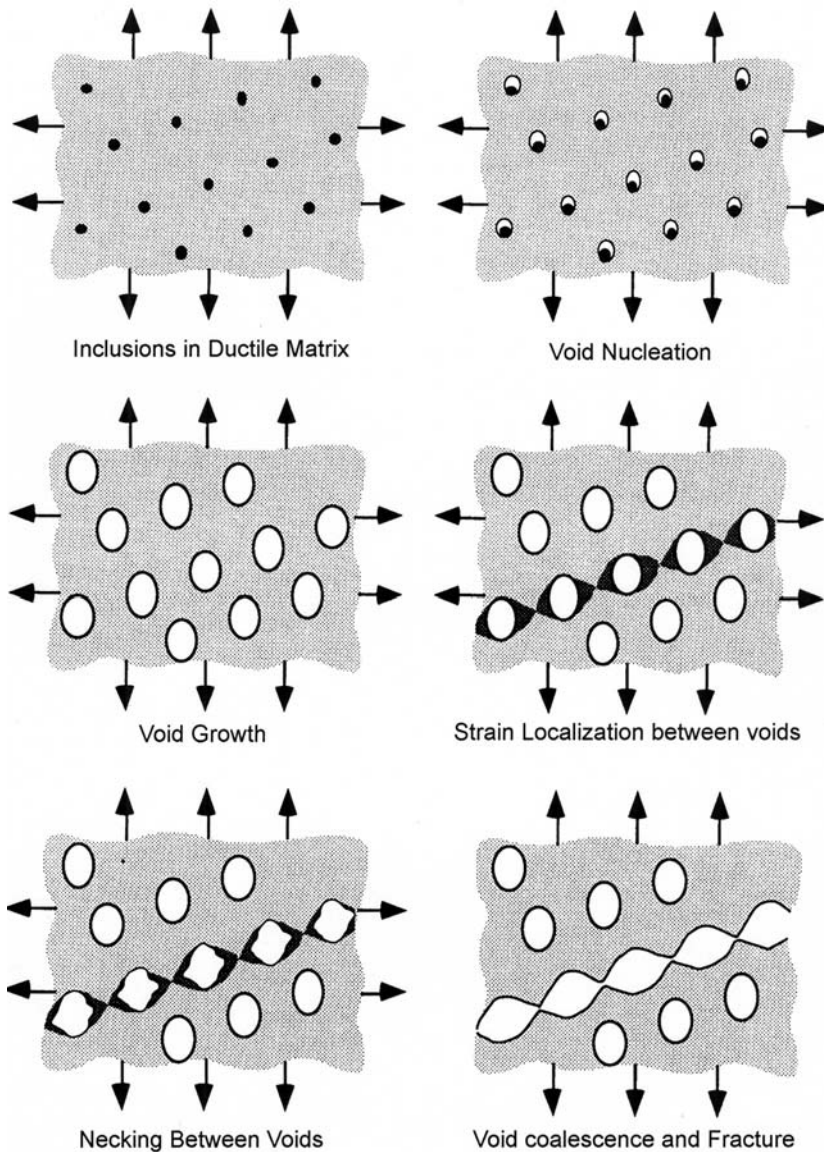
Figure 3   Creation of dimples by void formation, growth and coalescence.

This process of void formation, growth and coalescence is shown schematically in Fig. 3. If the particles are well matched to the matrix, and form a strong interface between the matrix and the particle, then the initial formation of voids is the critical step. Fracture occurs shortly after void formation [6]. If the interface between the particles and the matrix is weak, then voids form and grow readily. Substantial plastic deformation occurs. Fracture occurs when the voids reach a critical size. These voids substantially reduce the cross-section, with the resulting local plastic instability. Depending on the applied stresses, the shape and configuration of the
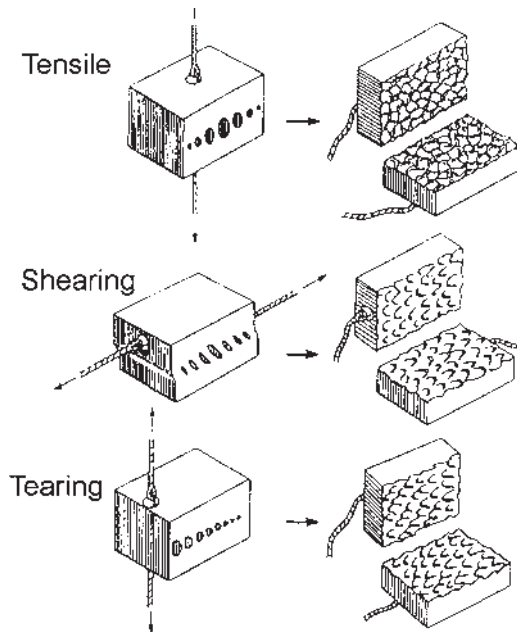
Figure 4 Schematic representation of the creation of dimples in a loaded member by (a) simple tension; (b) shear loading; and (c) tearing.

dimple shape can be changed (Fig. 4). This fact is important in determining the type of loading during a post-fracture investigation. Dimples are small, and can only be detected by using electron microscopy (Fig. 5).

2.2 Theories of Plasticity and Ductile Fracture

Many theories have been put forth in an attempt to explain plasticity and ductile fracture, but only a few have survived rigorous examination. Three of these are discussed below.

2.2.1 Maximum Normal Stress Theory

According to the Maximum Normal Stress Theory, the failure of a pan occurs when the largest principal stress equals the yield strength. Consider a small material element that is loaded as in Fig. 6. If $s_1 > s_2 > s_3$, then failure occurs when $s_1 = s_{ys}$ where S_{ys} is the yield strength of the material. This is illustrated schematically in Fig. 7. This theory is applicable for simple stress states such as tension. It is not applicable to torsion. In torsion, this theory implies that failure occurs when $t_{max} = S_{ys}$. This does not fit experimental data.

2.2.2 Maximum Shear Stress Theory

The maximum shear stress theory is easy to use and errs on the conservative side. It has been used as the basis for design codes and adequately predicts yielding. This
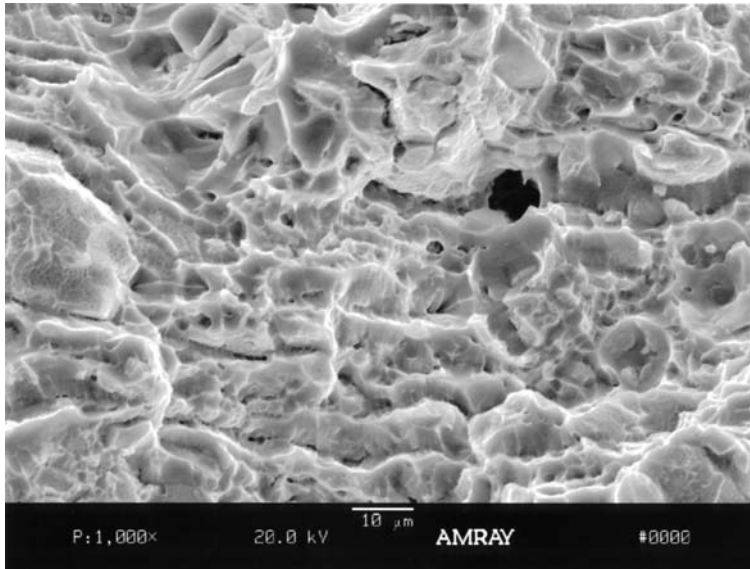
Figure 5 Scanning Electron Microscope photomicrograph of dimples in 7050–T7451 aluminum.
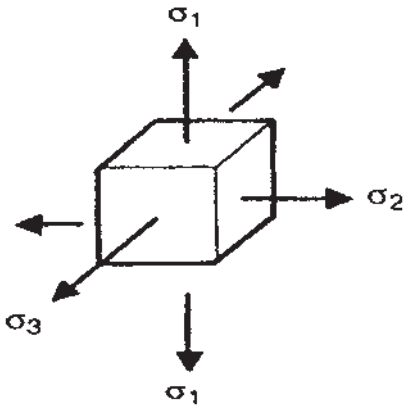
Figure 6 Principal stresses on a small differential element.

theory says that yielding occurs whenever the maximum shear stress in a small element is equal to the maximum shear stress in a tensile specimen at the onset of yielding.

If the Mohr's circle for a tensile test is considered (Fig. 8), then yielding occurs when

$$\tau_{max} = \frac{\sigma}{2} \qquad (1)$$
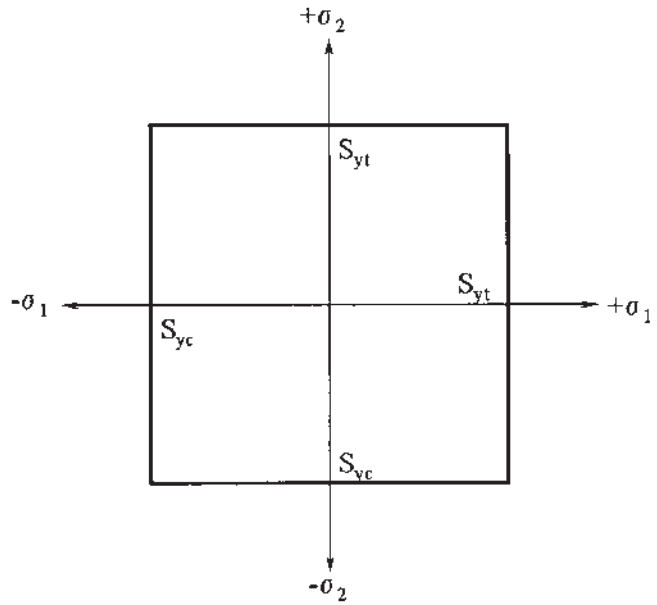
Figure 7   The Maximum Normal Stress Theory of failure based on the yield strength of a material.
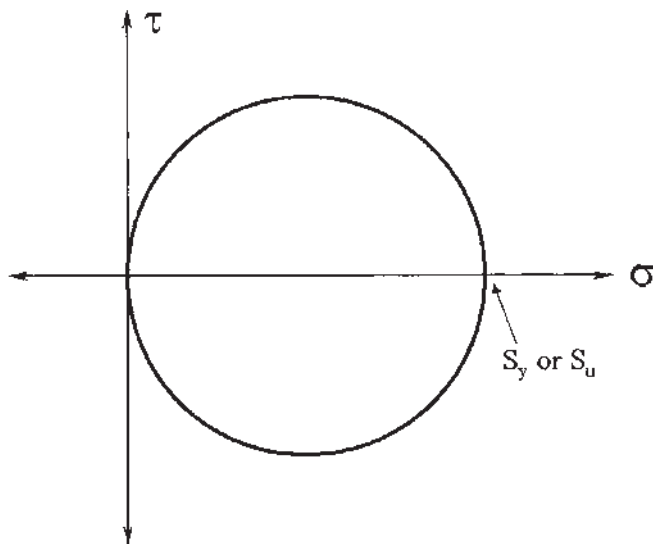
Figure 8   Mohr's Circle for a tensile stress.
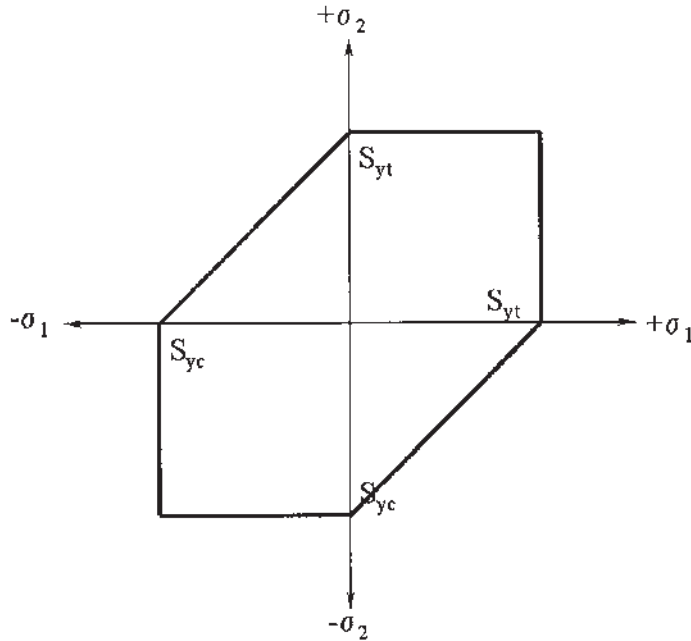
Figure 9 The Maximum Shear Stress Theory of failure.

For a three-dimensional mechanical element, the shear stresses are:

$$\tau_{12} = \frac{\sigma_1 - \sigma_2}{2} \tag{2}$$

$$\tau_{12} = \frac{\sigma_2 - \sigma_3}{2} \tag{3}$$

$$\tau_{12} = \frac{\sigma_1 - \sigma_3}{2} \tag{4}$$

The largest of the above stresses is designated t_{max}; whenever $t_{max} = S_{ys}$, then yielding will occur. This is illustrated in Fig. 9. The Maximum Normal Stress Theory and the Maximum Shear Stress Theory are identical when the principal stresses have the same sign.

2.2.3 Distortion Energy Theory (von Mises–Hencky Theory)

The von Mises–Hencky theory is the best theory to use for ductile materials because it defines accurately the beginning of yielding. This theory was originally proposed by von Mises [7] in 1913. In essence, the von Mises yield criterion states that all the principal stresses (tensile and shear) must be considered to create the von Mises stress, s_0:

$$\sigma_0 = \frac{1}{\sqrt{2}} \left[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 + 6(\tau_{12}{}^2 + \tau_{23}{}^2 + \tau_{13}{}^2) \right]^{\frac{1}{2}} \tag{5}$$
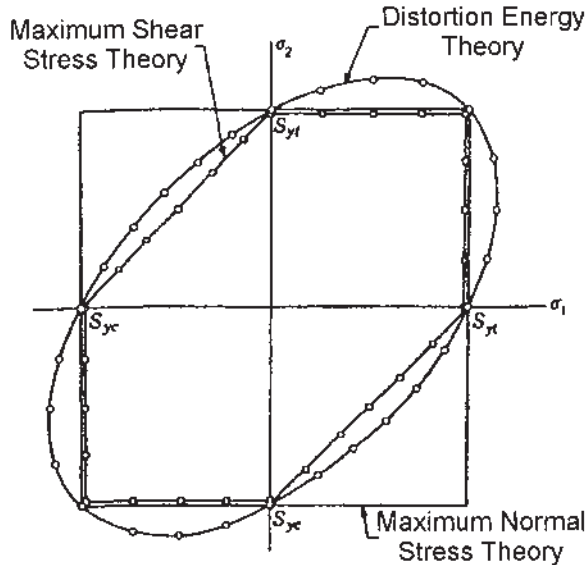
Figure 10   Comparison of the Distortion Energy Theory of failure and the Maximum Shear Stress Theory.

Yielding is predicted whenever the von Mises stress $s_o = s_{ys}$. If the stress state is biaxial ($s_3 = 0$) and the member is in torsion ($s_2 = -s_1$ and $t = s_1$), then Eq. (3) shows that

$$\tau_{\max} = 0.577\sigma_{ys} \tag{6}$$

This definition of yielding is not dependent on any normal stress or shear stress but depends on all components of the stress. Because the terms are squared, it is independent of the direction of the stress. It is also not necessary to know the largest principal stress to learn if yielding will occur.

2.2.4   A Comparison

The three yielding criteria are shown schematically in Fig. 10. Comparing the Distortion Energy Theory and the Maximum Shear Stress Theory, the distortion energy theory predicts that significantly higher stresses must be experienced before yielding occurs. The Maximum Normal Stress Theory predicts results equivalent to those predicted by the maximum shear stress theory whenever the directions of the principal stresses are the same. However, this theory fails to accurately predict yielding when the signs of the stresses are opposite. The Maximum Shear Stress Theory always gives conservative results. Of the three theories, the Distortion Energy Theory predicts yielding with the greatest accuracy in all four quadrants. The Maximum Normal Stress Theory should not be used in design.

3 ELASTICITY AND BRITTLE FRACTURE

Since the early 1940s, there has been tremendous growth in the number of large welded structures. Many of these structures have failed catastrophically in service – most notably the ''Liberty ships'' [8] used to transport war material during World War II. Analysis of the fracture surfaces of the failures [9] indicated that they initiated at a notch and propagated with no plastic deformation. These notches were of three types:

1. *Design features*. Structural members that were rigidly joined at angles less than 90° and then welded.
2. *Fabrication details*. Procedures used during the manufacture of the part caused the formation of notches. Welding are strikes, gouges, and fitting procedures created physical notches. Weld procedures and heat treatment caused metallurgical or micro-structural notches to occur from abrupt changes in microstructure or the production of microstructures that were brittle. Features such as porosity from welding or casting also caused brittle fracture initiation.
3. *Material flaws*. These flaws resulted from melt practice at the mill and appeared as large inclusions, internal oxidation, porosity, or segregation.

In brittle fractures, limited energy is absorbed by the fracture. Energy is absorbed through regions of small plastic deformation. Individual grains separate by cleavage along specific crystallographic planes.

3.1 Brittle Fracture Appearance

Visually, brittle fractures are characterized by little or no plastic deformation or distortion of the shape of the part. The fracture is usually flat and perpendicular to the stress axis. The fracture surface is shiny, with a grainy appearance. Failure occurs rapidly, often with a loud report. Because the brittle cleavage is crystallographic in nature, the fracture appearance is faceted. Often other features are present, such as river patterns [10]. These are shown schematically in Fig. 11.

One other form of brittle fracture is called intergranular cracking. In this fracture mechanism, failure occurs by decohesion along grain boundaries, and not on specific crystallographic planes like in cleavage fracture. Intergranular cracking can have several different causes. It is beyond the scope of this chapter to describe in detail all the different causes of intergranular cracking, but typical causes of intergranular cracking in aluminum alloys include:

1. Precipitation of brittle phases at the grain boundaries and the formation of grain boundary precipitate free zones.
2. Stress Corrosion Cracking
3. Intergranular Corrosion
4. Grain Boundary decohesion at elevated temperatures (creep rupture).

The fracture surface appearance of intergranular cracking is generally shiny, and faceted. It has the appearance of ''rock-candy.'' Often, when the mechanism is from corrosion, there is the corrosion product present. This can dull the appearance of the facets. The appearance of intergranular fracture is most clearly seen in the electron microscope, and an example is shown in Fig. 12.
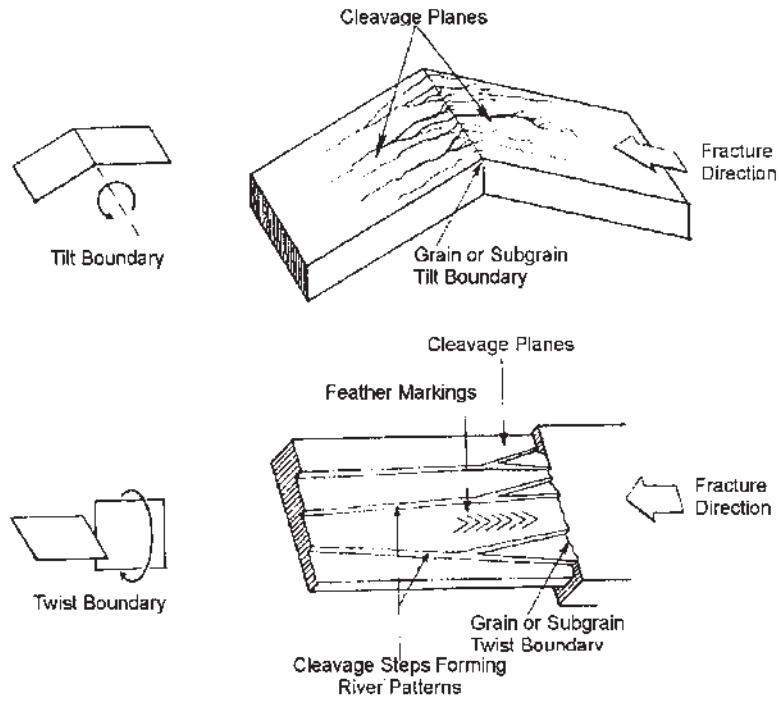
Figure 11 Schematic of river patterns formed in brittle materials (after McCall and French).
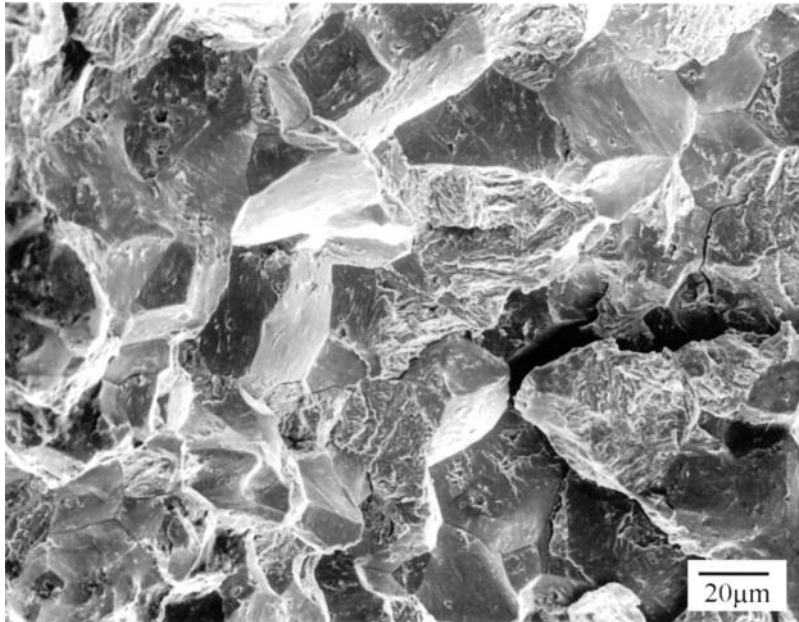
Figure 12 Intergranular Fracture

3.2 Theories of Elasticity and Brittle Fracture

When calculating stresses on a body, it is assumed that the body is elastic and homogeneous, and that it conforms to Hooke's Law,

$$\varepsilon = \frac{\sigma}{E} \tag{7}$$

where e is the strain (in/in); s is the stress (lb/in^2); and E is Young's Modulus. However, a material is not completely elastic or completely homogeneous. It is a collection of fibers or microstructures, either randomly aligned or with an organized orientation.

This orientation can provide directionality to the properties of a material. A material is isotropic (no directionality of properties) if the component is much larger than the constituent parts. If the organization of the constituent parts is nonrandom, the material is anisotropic and will show directionality in properties such as tensile, impact, or electrical conductivity.

If a material is perfectly elastic, there is no permanent deformation, and fracture occurs at the maximum stress, s_{UTS}. Brittle materials do not undergo plastic deformation, i.e. $s_y = s_{UTS}$. They follow Hooke's law until fracture. Brittle materials also have a compressive strength that is much greater than their tensile strength. In addition, a brittle material will have its torsional strength equal to their tensile strength. Many of the theories proposed to explain brittle fracture assume that the material is perfectly elastic. A few of the more important theories explaining brittle fracture are described below.

3.2.1 Coulomb–Mohr Theory

The Coulomb–Mohr theory, proposed by Coffin [11], is based on tensile properties and compressive strength. It says that failure will occur for any stress state for an elastic material whenever

$$\frac{\sigma_1}{\sigma_{UTS}} + \frac{\sigma_3}{\sigma_{UCS}} \geq 1 \tag{8}$$

This theory was confirmed by Grassi and Cornet [12], who stressed gray iron cast iron tubes biaxially until failure. Note that the maximum normal stress theory produces similar results when both of the principal stresses, s_1 and s_2, are positive. This theory was modified by Burton [13], who produced similar results. The results are not as conservative as the Coulomb–Mohrs theory in that fracture was more closely predicted. These theories are shown schematically in Fig. 13.

3.2.2 Griffith Microcrack Theory

The theoretical strength of a material is based on the cohesive force between atoms. If the cohesive strength is a sine curve, the following is obtained:

$$\sigma = \sigma_{max} \sin\left(\frac{2\pi x}{\lambda}\right) \tag{9}$$
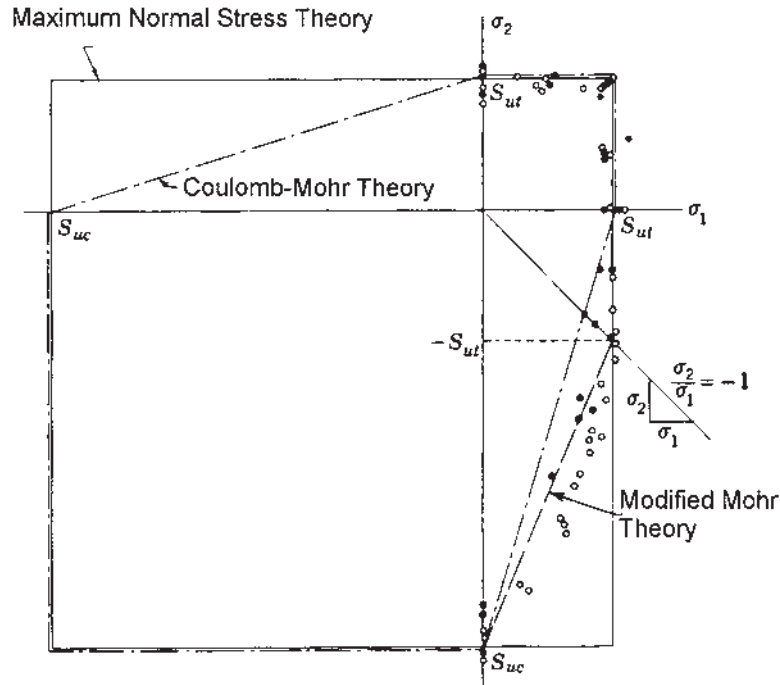
Figure 13   Columb–Mohr Theory and Maximum Normal Stress Theory.

Where $s_{\max}$ is the maximum theoretical cohesive strength, x is the atomic displacement due to the applied force, and ? is the wavelength of the lattice spacing. Since the changes in the atomic displacement are small, $\sin(x) = x$, resulting in a modification of Eq. (7):

$$\sigma = \sigma_{\max}\left(\frac{2\pi x}{\lambda}\right) \tag{10}$$

Since only brittle materials are being considered,

$$\sigma = E\varepsilon = E\frac{x}{\lambda} \tag{11}$$

Combining the above equations yields

$$\sigma_{\max} = \frac{\lambda}{2\pi}\left(\frac{E}{a_0}\right) \tag{12}$$

When the part fractures, two surfaces are created. Each of these surfaces has a surface energy γ_s, so that the work done per unit area in creating the fracture surface

is the area under the stress-displacement curve:

$$U_0 = \int_0^{1/2} \sigma_{\max} \sin\left(\frac{2\pi x}{\lambda}\right) dx = \frac{\lambda \sigma_{\max}}{\pi} \qquad (13)$$

Since energy is required to create two fracture surfaces,

$$\lambda = \frac{2\pi \gamma_s}{\sigma_{\max}} \qquad (14)$$

The maximum theoretical stress from atomic cohesive forces is

$$\sigma_{\max} = \sqrt{\frac{E_{\gamma_s}}{a_0}} \qquad (15)$$

Using typical values for the variables above and expressing results in terms of the elastic modulus E, estimates of $s_{\max}$ vary between $E/4$ and $E/15$. Common aluminum alloys typically have a fracture stress of approximately $E/200$. The only materials that approach the theoretical values are defect-free metallic whiskers or ceramic fibers. Therefore, this shows that small flaws or cracks are responsible for the tremendous decrease in the theoretical strength. These small flaws or cracks reduce the fracture strength due to stress concentrations [14].

Griffith [15] proposed that the difference between the theoretical strength and the strength realized in practice was due to a population of fine cracks that produce stress concentrations. These stress concentrations cause the theoretical cohesive strength to be achieved in local regions. Griffith created the criterion that "a crack will propagate when the decrease in strain energy is at least equal to the energy required to create a new crack surface" [16]. This statement is used to establish when a flaw of a specific size exposed to a tensile stress will propagate in a brittle fashion.

Using the crack model in Fig. 14, the flaw or crack is assumed to have an elliptical cross section of length $2c$. This shape is typical of many types of flaws. The thickness of the plate is very small compared to the width and length of the plate (conditions of plane stress predominate). Using the stress concentration of this elliptical crack (determined by Inglis [17]), the maximum stress at the tip of the crack is:

$$\sigma_{\max} = \sigma\left[1 + 2\left(\frac{c}{\rho t}\right)^{\frac{1}{2}}\right] \approx 2\sigma\left(\frac{c}{\rho t}\right)^{\frac{1}{2}} \qquad (16)$$

Where γ_t is the radius of the crack tip. The reduction of the elastic strain energy is equivalent to

$$U_E = \frac{-\pi c^2 \sigma^2}{E} \qquad (17)$$

Where s is the applied tensile stress normal to the crack. Since energy is released by propagation of the crack, the term to the right of the equal sign is negative. The theoretical stress between the atomic planes needs to be exceeded at only one point,
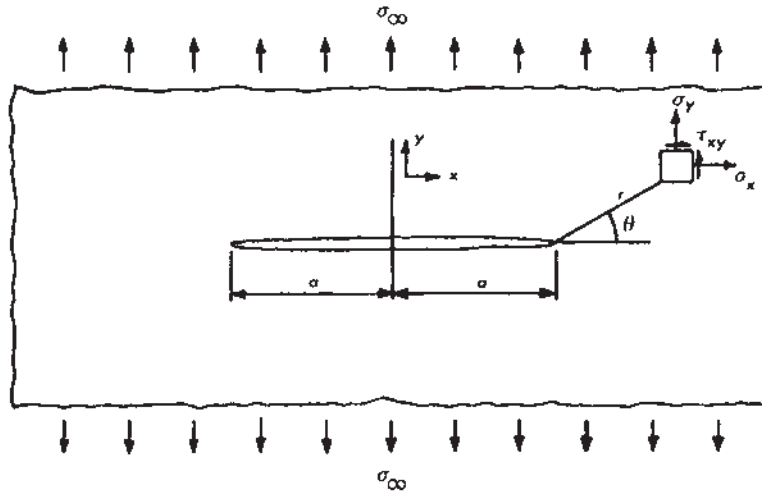
Figure 14   Griffith crack model used to determine the stress required to propagate a crack.

and the applied stress will be significantly lower than the theoretical stress. If the crack has a length of $2c$, and it is elliptical in shape and is relieved of stress in a roughly circular area of radius c, then the increase in surface energy is made up for by the decrease in strain energy. The condition when elastic strain energy equals the increase in surface energy due to propagation of the crack is provided by

$$\frac{d\Delta U}{dc} = 0 = \frac{d}{dc}\left(4c\gamma_s - \frac{\pi c^2 \sigma^2}{E}\right) \tag{18}$$

where

$$4\gamma_s = \frac{\pi c^2 \sigma^2}{E} = 0 \tag{19}$$

This leads to the Griffith formula,

$$\sigma = \left(\frac{2E\gamma_s}{\pi c}\right)^{\frac{1}{2}} \tag{20}$$

This is the stress required to propagate a crack of size c in a brittle material. If the material is thick in relation to the crack, and plane strain conditions predominate, then the Griffith equation is:

$$\sigma = \left(\frac{2\,E\gamma_s}{(1-v)^2 \pi c}\right)^{\frac{1}{2}} \tag{21}$$

where v is Poisson's Ratio.

4 FRACTURE MECHANICS

4.1 Basic Relationships

Metals that fail in a brittle manner experience some plastic deformation before failure [18–20]. There are three types of loading common in most engineering structures. They are shown in Fig. 15. Mode I is the most common and will be discussed here. Because of plastic deformation prior to fracture in metals (even when failing in a brittle manner), the Griffith microcrack theory does not generally apply to metals. One method of making the Griffith criterion of brittle fracture compatible with the plastic deformation evident in metals was suggested by Orowan [21]. He suggested the inclusion of a surface energy term due to the plasticity at the crack tip, γ_p. This results in a modification of the Griffith equation (for plane strain):

$$\sigma = \left(\frac{2E(\gamma_s + \gamma_p)}{\pi c}\right)^{\frac{1}{2}} \approx \left(\frac{2E\gamma_p}{\pi c}\right)^{\frac{1}{2}} \text{(plane stress)} \tag{22}$$

$$\sigma = \left(\frac{2E\gamma_p}{\pi c(1 - v^2)}\right)^{\frac{1}{2}} \text{(plane stress)} \tag{23}$$

The elastic surface energy term, γ_s is neglected because $\gamma_p \gg \gamma_s$.

Irwin [22] proposed that stress at the crack tip were a function of the applied stress and the crack size. He developed the relationship.

$$K = \sigma\sqrt{\pi c} \tag{24}$$

where K is the stress intensity factor. K is completely defined by the crack geometry, applied stress and specimen geometry. The value of the stress intensity factor when unstable crack growth occurs is the critical stress intensity factor, K_{ic} (for mode I), where the value of K_{ic} is a material property. While the above is for an elliptical flaw, other flaw shapes have also been calculated [23,24].
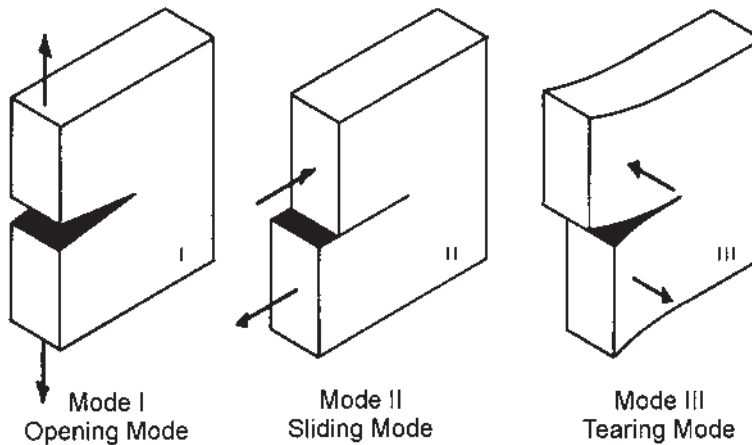
Figure 15 Types of loading typically experienced by engineering materials.

This assumes that plane strain conditions have been realized. If plane stress conditions are present, then the stress is relaxed by the increased plastic zone at the crack tip. Further, the state of stress is no longer triaxial and is diminished. The cases of plane stress and plane strain for modes I, II, and III are more fully described by Hertzberg [25] and Rolfe and Barsom [26].

Toughness is a measure of the energy required to resist fracture in a material. Often this property is more important than the actual tensile properties, particularly if the part is to be used in a dynamic environment. The term "impact strength" is used to denote the toughness of the material. This term is actually a misnomer; it should really be "impact energy." However, the term impact strength is so established that it makes little sense to change it. Toughness is strongly dependent on the rate of loading, temperature, and the presence of stress concentrations. Several standardized tests have been developed since World War II to measure the resistance to brittle fracture, notably the Charpy V notch test [27], the dynamic tear test [28], and the plane strain fracture toughness test [29]. Essentially, these tests are attempts to quantify the behavior of the material in service and how the material is expected to fail in service. The first two are discussed more fully later.

Under impact loading, there is a limited time for uniform plastic flow to occur. Locally, the deformation may exceed the fracture stress required at the grain boundaries, geometric irregularities, or other discontinuities. Once the crack has initiated, the crack itself becomes a stress riser, propagating until it is blunted or complete failure has occurred. Fracture toughness is dependent on a variety of variables that affect the mechanical properties such as test, temperature, chemistry and melt practice, strain rate, section size, notch acuity. Microstructure plays an important role. As these variables are changed, the fracture behavior may change.

Irwin [30] devised a series of tests to verify his theory of fast fracture, in that there was a critical stress intensity, which if exceeded, caused catastrophic failure. Thin sheets of 7075-T6 aluminum were used. Since the sheets were thin, plane strain conditions were not achieved, so he used a relationship between applied stress and stress intensity:

$$K = \sigma \left[W \tan\left(\frac{\pi c}{W}\right) \right]^{\frac{1}{2}} \tag{25}$$

where W is the width of the specimen. When (c/W) becomes small, the equation becomes identical to that for plane strain. In these experiments, Irwin was able to demonstrate that the stress at fracture was inversely proportional to c. It was these experiments that led to the intense effort in the aerospace industry to understand the underlying principles of fracture, and apply these principles to design fracture tolerant structures.

4.2   Effects of Section Thickness

Using 7075-T6, additional experiments were used to examine the effects of section size [31]. The results [32,33], showed that there was considerable variation of toughness with the thickness of the specimen. Further, at large thickness, the toughness appeared to reach a constant value (Fig. 16). Within this curve, there
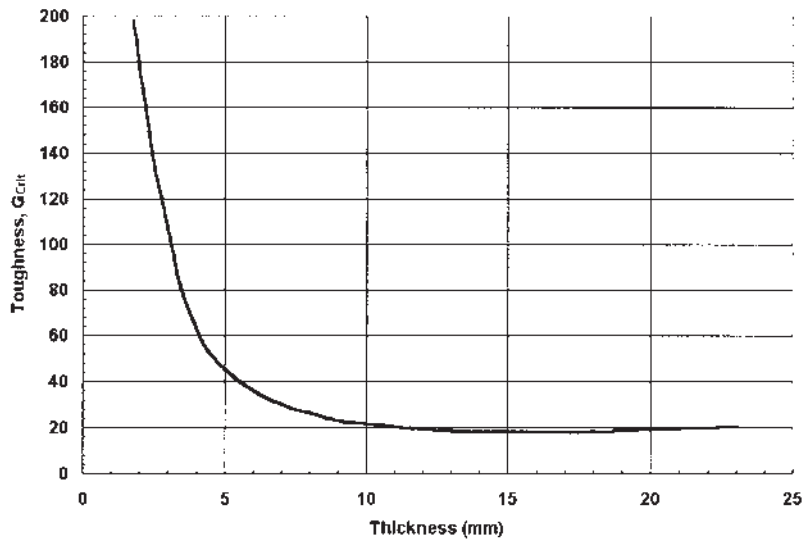
Figure 16   Variation of Toughness with thickness for 7075–T6 (From Ref. 132.)

are three regions apparent. First, there is the region where maximum toughness is obtained (thin sections). Second, there is the region of intermediate toughness, with the final region, a region with relatively constant toughness (thick sections).

In the first region, the fracture appears to consist entirely of a shear lip, or in other words, the fracture surface is inclined at angle of approximately 45° angle to the tensile axis. In this situation, the stress in direction of the thickness of the specimen tends toward zero, and a state of plane stress is achieved. As the specimen is pulled, it experiences buckling. Because of this buckling, yielding occurs on the through thickness planes at an angle of 45° to the tensile axis. Crack extension occurs by sliding. This sliding motion is achieved by the movement of a number of screw dislocations [34,35] on the 45° plane as shown by Fig. 17.

In the intermediate range, the fracture behavior is complicated. The fracture does not consist of entirely "slant" type fracture, nor does it contain entirely a "flat" plane strain type fracture. Instead the regions of "flat" and "slant" fracture are approximately equal. At the thin end of the thickness range, the "slant" ligaments on either side of the test piece carry most of the load. At the thick end of the range, the side ligaments carry a much smaller percentage of the load. The amount of "flat" fracture increases. This is shown schematically in Fig. 18. It has been found [31] that the amount of "flat" fracture depends only on the thickness of the test specimen, and was independent of crack length.

In the third region, the fracture consists of predominately "flat" fracture. Some evidence of very small shear lips may be present at the later part of fracture. Fracture is catastrophic and rapid. No plastic deformation is evident. In this third region, any increase in the thickness of the test piece causes no further decrease in the toughness.

Figure 17 Mode of separation in a thin sheet.

4.3   Crack Opening Displacement

In the above discussions, it has been assumed that the plastic region at the crack tip is small compared to the crack dimensions. When this occurs, the material can be considered using linear-elastic fracture mechanics (LEFM). When the material used is ductile, and the plastic region at the crack tip is large compared to the crack dimensions, then a different tool must be used. Elastic-Plastic Fracture Mechanics applies to materials that show nonlinear behavior as a function of time (plastic deformation). This method can be used when LEFM is no longer valid, or when the specimen thickness does not meet plane-strain conditions. It is apparent, that the use of LEFM is appropriate to the use of large thick test specimens, or to the testing of high-strength brittle materials. Because of cost reasons, the use of small test pieces is of great interest. This necessitates the use of EPFM. In addition, the use of a smaller test specimen may more accurately reflect the actual size of the parts, and are easily tested in the laboratory.

4.3.1   First Principles

The crack-tip stress or strain can not be characterized when plastic deformation or yielding is extensive. However Wells [36], when attempting to measure K_{Ic} in a

Figure 18   Schematic of fracture in the intermediate range.

number of steels, found that the steels did not exhibit behavior that could be modeled using LEFM. The steels were too tough. During examination of the specimens, he noticed that the crack faces had separated prior to fracture. This was caused by extensive plastic deformation prior to fracture. This deformation had effectively blunted the notch. He also noticed that the amount of blunting of the crack that occurred was proportional to the apparent toughness of the material. He proposed that the crack opening displacement be used to characterize materials that do not exhibit LEFM behavior. This became known as the crack tip opening displacement (CTOD) or crack opening displacement (COD).

An initial approximation that relates the COD to the stress intensity factor can be developed if the crack tip plasticity makes the crack act as if it were longer [37]. Therefore the effective crack tip opening displacement can be determined by assuming an effective crack length of $c + r_y$.

$$\delta = K_I \left[\frac{k+1}{G} \right] \sqrt{\frac{r_y}{2\pi}} \tag{26}$$

The plastic zone correction for plane stress is

$$r_y = \frac{1}{2\pi}\left(\frac{K_I}{\sigma_y}\right)^2 \tag{27}$$

Substituting provides

$$\delta = \frac{4}{\pi}\frac{K_I^2}{\pi\sigma_y E} \tag{28}$$

where δ is the COD, G is the shear modulus, and K is the stress intensity. As an alternative, the COD can be related to the energy release rate:

$$\delta = \frac{4}{\pi}\frac{\Gamma}{\sigma_y} \tag{29}$$

where Γ is the energy release rate. A more general form of the above relations can be expressed as

$$\delta = \frac{\Gamma}{m\sigma_y} = \frac{K_I^2}{m\pi\sigma_y E} \tag{30}$$

where m is a dimensionless constant that is approximately 1.0 for plane stress and 2.0 for plane strain.

4.3.2 Measuring Crack Opening Displacement

In the beginning, the crack opening displacement was measured using a spring loaded paddle device [38]. This spring loaded paddle was located inside the crack (which was not extended by fatigue or other method). As the specimen was loaded, and the crack faces would separate, the paddle would rotate in an amount proportional to the crack opening displacement. This rotation was sensed by a trans-ducer, and displayed on a strip chart or other recording device. This device was useful for specimens that had no pre-cracking, or when insertion depths were shallow. This device is shown schematically in Fig. 19.

There are different test standards available for COD testing. The British Stan-dard [39] was first, followed by the US Standard [40]. Experimentally, the crack opening displacement is divided into two parts – the elastic and plastic components. The elastic COD is obtained from:

$$\delta_{elastic} = \frac{K^2(1 - v^2)}{2\sigma_y E} \tag{31}$$

The plastic component assumes that the test specimen rotates about a plastic hinge. The displacement from the plastic deformation of the test specimen is related to the displacement at the mouth of the crack, V_p:

$$\delta_{plastic} = \frac{r_p(W - a)V_p}{r_p(W - a) + a + z} \tag{32}$$
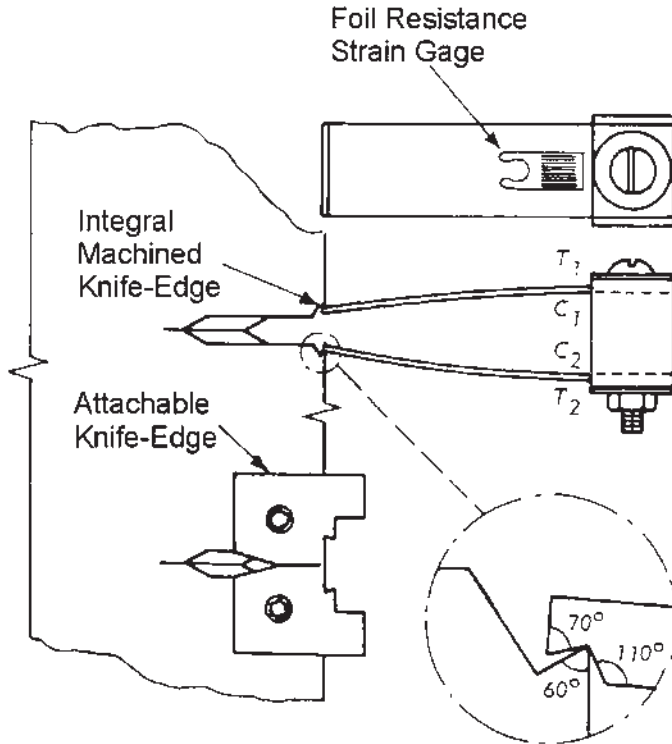
Figure 19   Schematic of a paddle-type COD gage.

where r_p is the plastic rotational factor. From BS 5762, $r_p = 0.40$; from ASTM E1290)$r_p = 0.44$ for the single edge cracked beam specimen (SENB). For the compact specimen, the plastic rotational factor is:

$$r_p = 0.4 \left[1 + 2 \left[\left(\frac{a_0}{b_0} \right)^2 + \frac{a_0}{b_0} + 0.5 \right]^{\frac{1}{2}} - 2 \left[\frac{a_0}{b_0} + 0.5 \right] \right] \tag{33}$$

Measurement of the crack mouth opening displacement is easily accomplished with the use of modern clip gages. This gage consists of four strain gages on a pair of cantilever beams. Any deflection of the beams causes a change in the resistivity in the strain gages, with is linear with the mouth displacement. These clip gages can be either clipped directly to knife edges attached to the test specimen, or clipped directly to the specimen. If clipped directly to the specimen, the test specimen must be machined to accommodate the clip gage.

4.4   The J-Integral

This parameter has enjoyed great success when characterizing ductile, or nonlinear materials. Originally proposed by Rice [41], this method extends the usefulness

of fracture mechanics beyond the limiting constraints of LEFM. Rice looked at the amount of nonlinear elasticity, and showed that the nonlinear energy release rate could be written as a path-independent line integral [42]:

$$ J = \int_{\Gamma} \left(w dy - T_i \frac{\partial u_i}{x} ds \right) \tag{34} $$

where w is the strain energy density, T_i are the components of the traction vector, μ_i are the displacement vector components, and ds is the length increment along path Γ. The strain energy density is defined as:

$$ w = \int_0^{\varepsilon_{ij}} \sigma_{ij} \, d\varepsilon_{ij} \tag{35} $$

where σ_{ij} and ε_{ij} are the stress and strain tensors. T_i defines the normal stresses acting along the contour Γ:

$$ T_i = \sigma_{ij} n_j \tag{36} $$

where n_j are the components of the vector normal to Γ.

It has also been shown that J can be used as a stress intensity factor [43,44]:

$$ \sigma_{ij} = K_1 \left(\frac{J}{r} \right)^{\frac{1}{n+1}} $$
$$ \varepsilon_{ij} = K_2 \left(\frac{J}{r} \right)^{\frac{n}{n+1}} \tag{37} $$

where k_1 and k_2 are proportionality constants. In the special case of linear-elastic materials, $n = 1$, and the above equation predicts that the stress is proportional to $1/r$, which is consistent with LEFM.

4.4.1 Relationship Between J and COD

Shih [45] described the relationship between J and COD by first assuming that

$$ J = m\delta\sigma_y \tag{38} $$

where m is a function of σ_y/E and the strain hardening component n, that varies between 1.15 and 2.95 [46,47]. This is shown in Fig. 20. This equation is an extrapolation from LEFM. He examined the displacement at the crack tip to J and mechanical properties. Shih determined that the crack opening displacement was related to J by:

$$ \delta = \frac{d_n J}{\sigma_y} \tag{40} $$
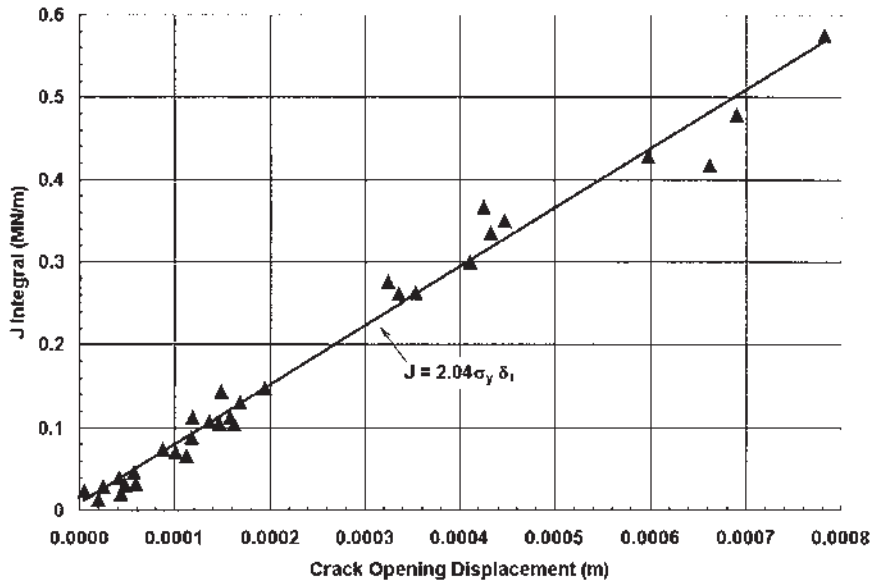
Figure 20 Variation of J-integral with the crack opening displacement (COD) for notched bend (SENB) test with a steel with $\sigma_y = 370$ MPa.

4.4.2 Usefulness of the J-Integral

The path independence of the J-Integral allows calculation of the plastic strains and stresses from a description of a remote contour. What this means is that a contour can be chosen that only contains elastic stresses and strains. This greatly simplifies calculation of J.

As was shown above, the J-Integral can be also considered an expression of the elastic-plastic energy release rate or stress intensity factor. Based on this, and in analogy with LEFM, it can be expected that there is some critical value of J, which would predict the start of crack extension. This would be similar to K_{lc} in LEFM. As an example, calculated values of J could be compared to the critical value of J, J_{Ic}, to determine is crack extension would occur. Crack extension would not occur if J were less than J_{lc} (for mode *I* loading).

The idea of the J-integral is not easy to understand. But the concept is useful to understand what is occurring in ductile or thin materials. In principle, the use of the *J*-Integral is only useful to determine the occurrence of cracking, and not predict crack growth.

Solutions for J-integral are quite difficult. Usually for real test specimens, or for real parts, it is necessary to use computer finite element analysis to determine J. However, for common test pieces, solutions to J have been developed. For the Single Edge Notched Bend Specimen (SENB):

$$J = \frac{2U_t}{Bb} = \frac{2U_t}{B(W-a)} \tag{41}$$

For the Compact Test Specimen (CT):

$$J = \frac{2U_t}{B(W-a)} f\left(\frac{a}{w}\right) \tag{42}$$

where:

$$a = \left[2\sqrt{\left(\frac{a}{b}\right)^2 + \left(\frac{a}{b}\right) + \frac{1}{2}}\right] - 2\left(\frac{a}{b} + \frac{1}{2}\right) \tag{43}$$

and

$$f\left(\frac{a}{W}\right) = \frac{(1+a)}{(1+a^2)} \tag{44}$$

4.5   Applications of Fracture Mechanics

In most situations, the size of a defect is large and can be readily detected before a piece of equipment ever sees service or catastrophic failure can occur. However, catastrophic failures still occur. This usually occurs by the growth of a small flaw by fatigue or other growth mechanism, until the critical flaw size is exceeded. Once this flaw size is exceeded, catastrophic failure occurs. This has occurred in ships, aircraft [48] and bridges – practically in anything of value. In this section, several examples of the use of fracture mechanics in predicting catastrophic failure will be illustrated.

4.5.1   Fatigue Crack Growth

Fracture mechanics has been used to study crack propagation [49–51]. When fatigue crack growth has been measured and plotted against the stress intensity, three regions are apparent (Fig. 21). In the first region, fatigue behavior is just beginning. There is a threshold stress intensity, K_{th}, where stress intensity values below the threshold do not cause crack growth. This region is controlled by the mean stress, microstructure and the environment. In the second region, fatigue striations are formed. Crack growth rates in this region are influenced largely by environment, mean stress and the frequency of testing. Microstructure plays a minor role in influencing crack growth in this region. This region is often linear between the change in the stress intensity factor and the crack growth rate. In the final region, final fast fracture occurs. This region is characterized by final catastrophic failure. This region is largely influenced by microstructure, thickness and mean stress. There is little influence of environment on final fracture. Final fracture occurs when the fatigue stress intensity factor becomes equal to the fracture toughness of the material.

The crack propagation rate in a material follows an equation of the form:

$$\begin{aligned}
\frac{\mathrm{d}a}{\mathrm{d}N} &= C(\Delta K)^m - \text{Paris} \\
\frac{\mathrm{d}a}{\mathrm{d}N} &= \frac{C(\Delta K)^m}{(1-R)K_c - \Delta K} - \text{Forman}
\end{aligned} \tag{45}$$
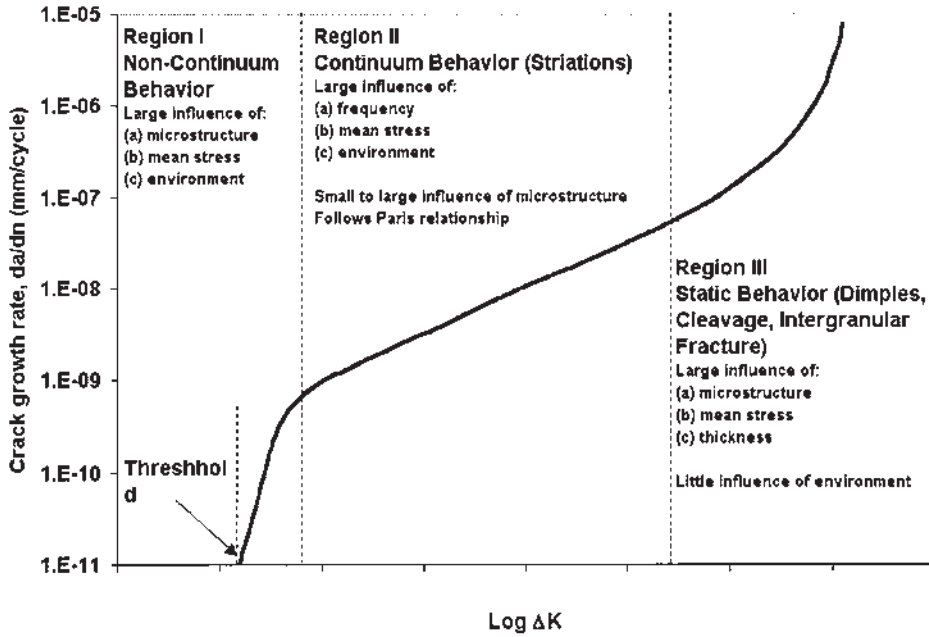
Figure 21 Characteristics of the fatigue crack growth rate curve.

where C is a constant based on the geometry of the crack and test; a the alternating stress; a is the crack length, m is a material constant found during testing; R is the stress ratio; K is the change in the stress intensity factor, K_c is the critical stress intensity, and N is the cycles to failure. These two equations form the basis for prediction of the cycles to catastrophic failure from a flaw of some finite size. The first equation describes only the linear region in Region II, while the Forman equation also describes Region III.

While the above provides information regarding how a crack propagates in a solid, it can also be used to predict fatigue life. Assume a solid, with some elliptical flaw. The critical size of the flaw is given by Brock [52] as

$$a_{cr} = \frac{K^2 Q}{1.21 \pi \sigma^2} \tag{46}$$

where α_{cr} is the critical flaw depth, K is the critical stress intensity, Q is the flaw shape factor, and σ is the applied stress. Using the Paris equation, where:

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C K_1^m \tag{47}$$

where α is the depth of the flaw, N is the number of cycles sustained, and K_1 is the stress intensity range,

$$K_1 = \sigma \left(\frac{1.21 \pi a}{Q} \right)^{\frac{1}{2}} \tag{48}$$
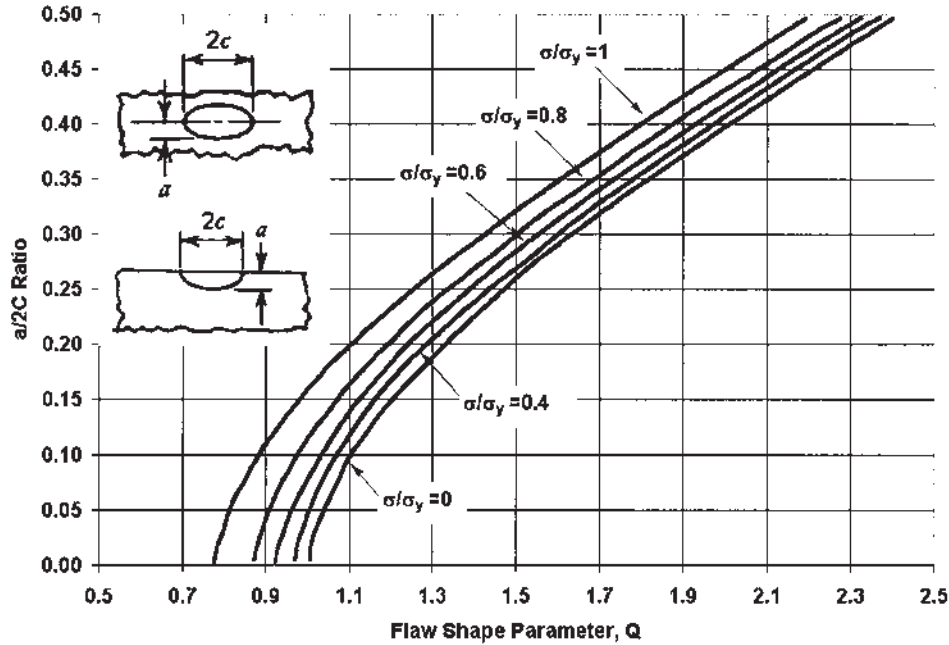
Figure 22 Flaw-Shape parameter, Q, curves for surface and internal cracks (From Ref. 133.)

The stress range is obtained using specially designed specimens related to a specific geometry [53–55]. The material constants C and m are obtained through fatigue crack growth testing. The value of the flaw shape factor, Q, is obtained from the references cited above as well as others [56–58], as well as shown in Fig. 22. By combining the above equations and rearranging, the following equation obtained:

$$dN = \frac{da}{C\left[\sigma\left(\frac{121\pi a}{Q}\right)^{\frac{1}{2}}\right]^m} \tag{49}$$

The cycle life is obtained by integrating the above equation between the limits of the initial flaw depth α_o, and the critical flaw depth α_{cr}:

$$N = \int_{a_0}^{a_{cr}} da \left\{ c\left[\sigma\left(\frac{1.21\pi a}{Q}\right)^{\frac{1}{2}}\right]^m \right\} \tag{50}$$

The initial flaw depth, a_o, could be either the detectable flaw size from nondestructive testing or an assumed flaw size. Typically, for most practical applications, a flaw size of 0.010 inches is usually adequate, and is the lower limit of delectability for most NDT methods.

Based on the above relationship, it is possible to determine the crack growth rate for any type of da/dN -ΔK relationship. However, this assumes that the correlation between crack length, and the stress intensity factor is known. Sometimes, because of the crack geometry, it is difficult to analytically integrate the crack growth rate. Instead, it is necessary to use numerical methods to determine the crack growth rate of a part or built-up structure. Often the difficulty occurs because of the complex nature of geometry of the crack, or because of the occurrence of load shedding. Often, cracked members in built-up structures will tend to shed loading to other uncracked members. This is because cracking causes a decrease in stiffness. The displacement of each of the members is also mutually constrained. The technique above is strictly valid for constant amplitude loading.

As often occurs in real structures, the loading is generally not constant amplitude. The fatigue environment is complex, with many different loading, and possibly random cycles. This means that each different cycle can have a different ΔK, with different crack growth rates. One way to predict the crack growth rate is to sum the effect of each cycle. However, this will provide a very conservative answer if there are any large positive peak loads. These peak loads will cause retardation in cracking [59].

4.5.2 Stress-Corrosion Cracking (SCC)

Stress-Corrosion-Cracking is the simultaneous action of a corrosive environment and a tensile stress. In the absence of either a corrosive environment or a tensile stress, stress-corrosion-cracking will not occur. In traditional stress-corrosion-cracking testing, a test specimen is exposed to a corrosive environment by applying a large stress. In most instances, the time to crack initiation is the primary interest. Because the stresses are typically high, propagation and subsequent SCC failure is usually rapid. In general, data is presented in terms of the applied stress versus the time to crack initiation.

In the initial corrosive attack, the corrosive environment produces a stress riser, such as a pit or crevice. Yielding may occur locally. In the case of aluminum alloys, the soft precipitate free zones (PFZ) will show pronounced grain boundary attack.

If the stress-corrosion-cracking is slow, then the part or component can be inspected at regular intervals to detect cracking before catastrophic failure can occur. This is true whether aircraft, automobile, or steam generators [60].

Typical investigations [61] generally establish the stress-corrosion-cracking susceptibility. The stress intensity factor is varied, and the time to failure under the SCC environment is documented. After failure, the fracture surfaces are examined to determine the extent of cracking prior to failure. In the case of branching cracks, then the resultant critical stress intensity factor, K_{ISCC}, is overly optimistic.

The techniques of measuring the SCC crack growth are similar to fatigue and fracture testing. Electrical potential measurements are used extensively. However, this may affect the anodic dissolution that is occurring at the crack tip. But rates have been observed that compare favorably with other methods. Compliance measurements are also used. Because the size of the stress intensity factor is small, the associated plastic zone at the crack tip is also small. Any observed changes in the compliance can be directly related to the crack opening displacement.

Testing often shows a threshold of K_{ISCC}, which is substantially below K_{Ic}. This threshold value of the critical stress intensity factor for SCC, indicates that the specimen will last indefinitely, assuming no flaws of critical size will occur in service.

If no limit of K_{ISCC} is observed, then loading of the part must be small to ensure that slow crack growth rates occur. This makes certain that the crack can be found in the field before catastrophic failure occurs. Otherwise, a change in the material or environment is required. In an investigation of the crack growth rates during stress-corrosion-cracking of several 7000 series aluminum alloys [62] exposed to intermittent rain and moderate humidity, a range of crack growth rates were found (Fig. 23).

In the first example, very fast cracking was observed in 7079–T6541. In example II, moderate crack growth was observed in 7075–T651 and 7178–T651. Very slow crack growth rates were observed in 7075–T7651 (example III). If a part was manufactured from 7079–T651, and exposed to this corrosive environment, the crack growth rate would be much to fast for routine inspections. The crack growth rates for 7075–T651 and 7178–T651 would allow crack growth rates to be estimated and reasonable inspection cycles instituted. The final example, 7075–T7651, shows a very slow crack growth rate. Routine inspections at short intervals (months) would be unnecessary and expensive.

The stress-corrosion crack growth rates vary markedly with alloy, environment and processing. Both a linear dependence and independence on the stress intensity can be found in some alloys. This has been found [63] for an Al-Zn-Mg alloy that
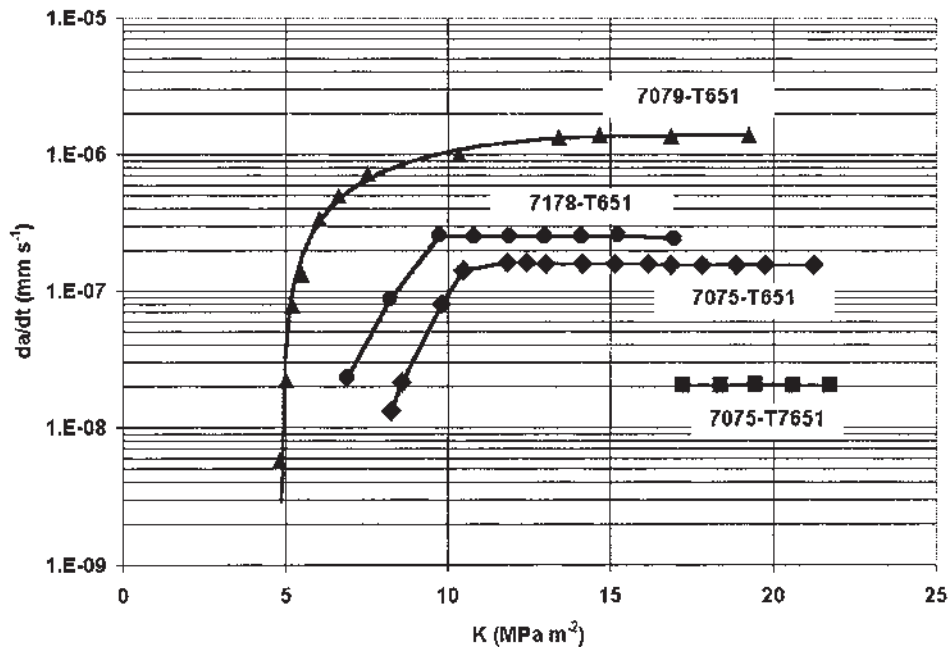
Figure 23 Crack growth rates in several 7000 series aluminum plate, exposed to intermittent rain and moderate humidity.
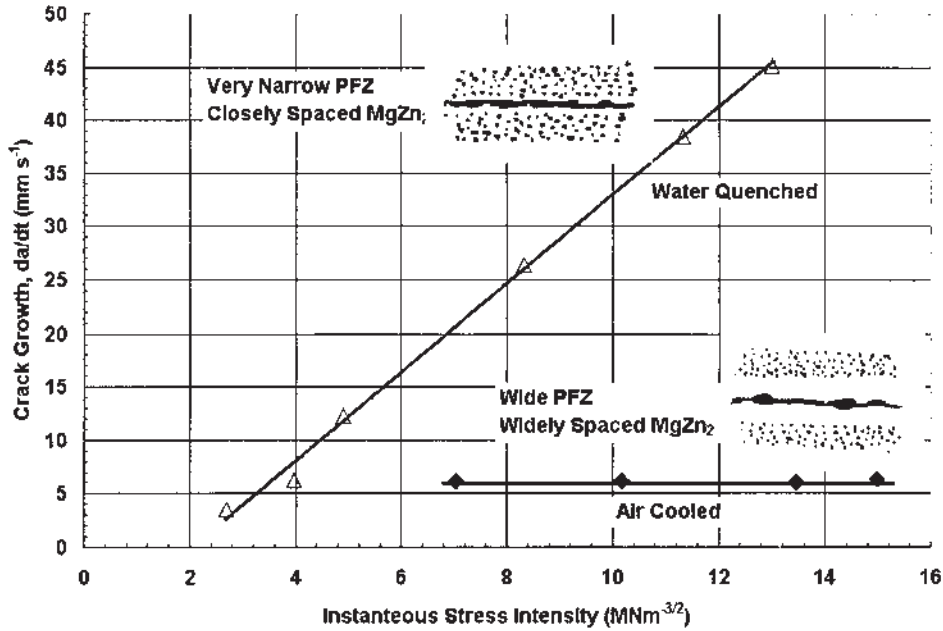
Figure 24 Variation of Stress-Corrosion crack growth rates for a pure Al-Zn-Mg alloy quenched by different methods prior to aging.

had been subjected to different quenching conditions (Fig. 24). This was achieved while maintaining a constant macroscopic microstructure and yield strength.

There is no one mechanism or theory that adequately explains stress-corrosion-cracking failures. Because of this, predictive models are strongly based on empirical evidence. For any combination of material and environment, it is necessary to evaluate the SCC growth rates to insure a safe design.

5 FATIGUE

Parts are subject to varying stresses during service. These stresses are often in the form of repeated or cyclic loading. After enough applications of load or stress, the components fail at stresses significantly less than their yield strength. *Fatigue* is a measure of the decrease in resistance to repeated stresses.

Fatigue failures are brittle appearing, with no gross deformation. The fracture surface is usually normal to the main principal tensile stress. Fatigue failures are recognized by the appearance of a smooth rubbed type of surface, generally in a semicircular pattern. The progress of the fracture (and crack propagation) is generally suggested by "beach marks." This is illustrated in Fig. 25. The initiation site of fatigue failures is generally at some sort of stress concentration site or stress riser. Typical fracture appearance is shown schematically in Fig. 26.

Three factors are necessary for fatigue to occur. First, the stress must be high enough that a crack is initiated. Second, the variation in the stress application must be large enough that the crack can propagate. Third, the number of stress applications must be sufficiently large that the crack can propagate a significant
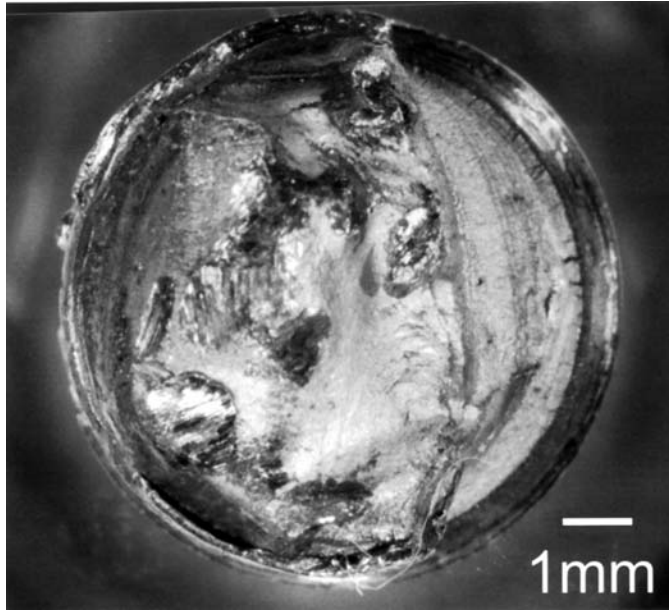
Figure 25   Actual fatigue failure of a bolt, showing characteristic "beachmarks."

distance. The fatigue life of a component is affected by a number of variables, including stress concentration, corrosion, temperature, microstructure, residual stresses, and combined stresses.

Fatigue is caused by a series of loading and unloading, in a variety of waveforms. The load application can be sinusoidal, triangular (sawtooth waveform), or spectral (random) loading.

The range of stress application, σ_r, is suggested by

$$\sigma_r = \sigma_{\max} - \sigma_{\min} \tag{51}$$

and the alternating Stress, σ_a, is represented by

$$\sigma_a = \frac{\sigma_r}{2} \tag{52}$$

The mean stress, σ_m, calculated as:

$$\sigma_m = \frac{\sigma_{\max} - \sigma_{\min}}{2} \tag{53}$$

When representing fatigue data, two quantities, the stress ratio R and the stress amplitude A, completely describe the stress applied. The stress ratio is given by:

$$R = \frac{\sigma_{\max}}{\sigma_{\min}} \tag{54}$$
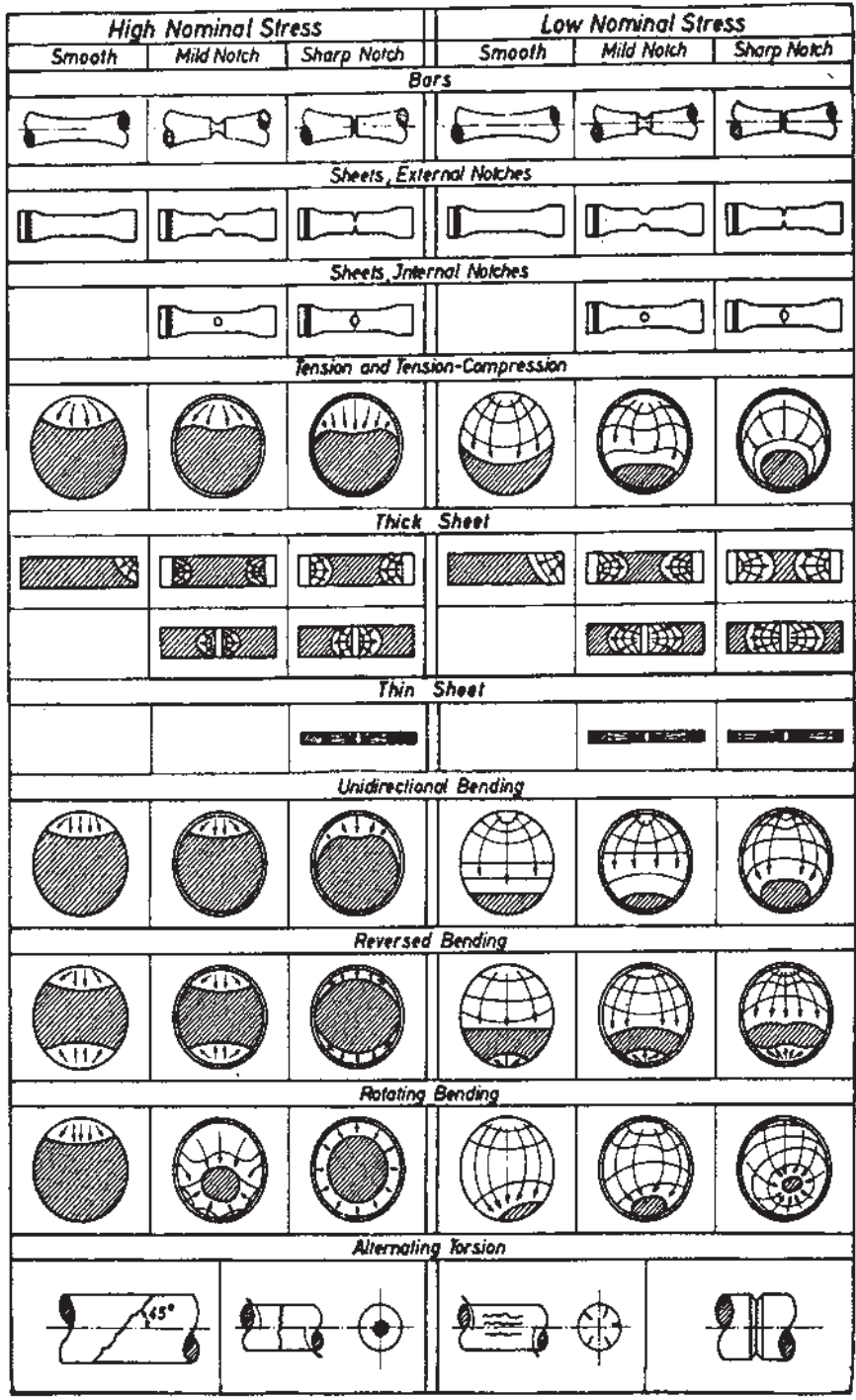
Figure 26 Schematic illustration of simple fatigue failures.
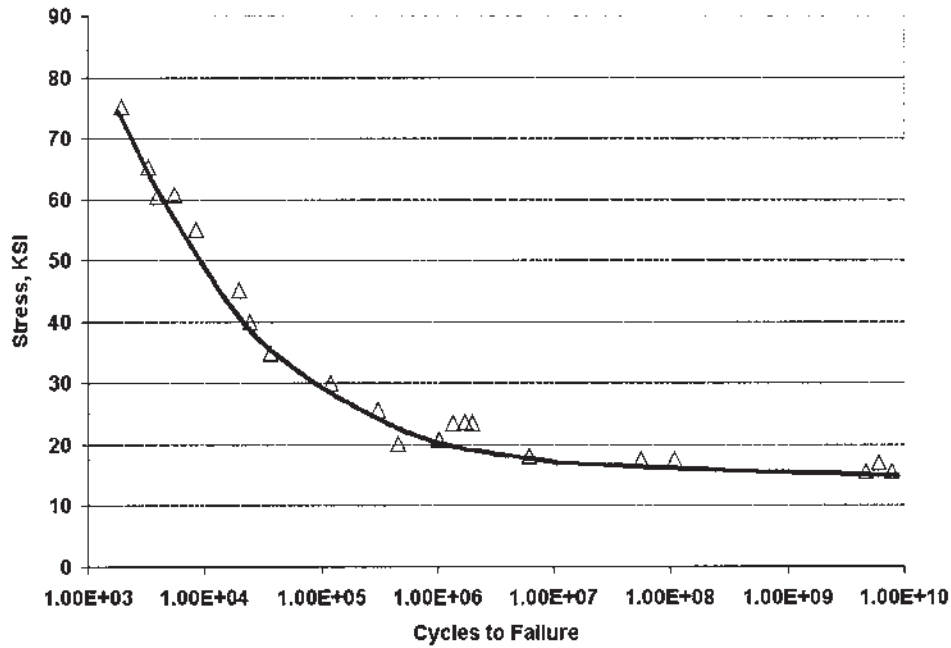
Figure 27 Typical presentation of fatigue data in an S–N curve. Data is shown for 7075–T76.

while the stress amplitude is given by:

$$A = \frac{\sigma_a}{\sigma_m} \tag{55}$$

Engineering fatigue data are usually displayed graphically in an S–N curve (Fig. 27), where the applied stress σ (or S) is plotted against logarithm of the number of cycles, N. As the stress is increased, the number of cycles until failure decreases. For steels, the S–N curve becomes horizontal at some low stress level. This is called the *fatigue limit*, σ_e.

For steels, the fatigue limit σ_e depends on the ultimate tensile strength (σ_{UTS}) [64] in a rotating, reversed fatigue test:

$$\sigma_e = \frac{\sigma_{UTS}}{2} \tag{56}$$

In other loading types, the endurance limit for alternating axial loading (completely reversed) is [65]:

$$\sigma_{e-axial} = 0.035\sigma_{UTS} \tag{57}$$

and for reversed torsional testing, the torsional endurance limit is

$$\tau_e = 0.3\sigma_{UTS} \tag{58}$$

Aluminum alloys do not show a fatigue limit, and failure occurs at some extended number of cycles at low applied stress. An arbitrary fatigue limit for aluminum alloys is determined by specifying the stress at some large number of cycles, typically 10^8 cycles. The endurance limit for aluminum alloys can be estimated at 10^8 cycles by [66]:

$$\sigma_e = \beta\sigma_{\mathrm{UTS}} \tag{59}$$

where β has been suggested to be:

$$\beta = \left(\frac{1 + \left(\dfrac{0.0031n^4}{1 + 0.0065\sigma_{\mathrm{UTS}}}\right)}{1 + 0.0031n^4}\right) \tag{60}$$

where σ_{UTS} is the ultimate tensile strength in MNm^{-2}, n is the logarithm of the desired endurance limit (typically 8). A different approach was taken by Rowe [67]:

$$\sigma_{\mathrm{e_r}} = (1.0075)^{1/n}\sigma_1^{0.911} \tag{61}$$

where σ_{e} is the rotating bending endurance limit at 5×10^8 cycles, σ_1 is the true stress at unit strain, and n is the strain hardening coefficient.

The fatigue limit and fatigue tests show considerable variation [68–70]. Because of this variability, the fatigue life must be examined in terms of probability. This requires testing many samples so that the mean and standard deviation can be determined. In one study [71] 200 steel specimens were tested. It was found that fatigue life followed a gaussian distribution. The fatigue endurance limit was examined by Ransom [72]. In this study 10 S–N curves were developed using 10 specimens for each stress level, for a total of 100 fatigue specimens. The specimens were prepared identically and were from the same bar of material. Ransom found that there was a 20% scatter around the fatigue limit.

5.1 Fatigue Mechanisms

The structural features of fatigue failures are generally divided into four distinct areas [73]:

1. Crack initiation, the early development of fatigue damage.
2. Slip band crack growth, the early stages of crack propagation. This is often called Stage I crack growth.
3. Stable crack growth, which is usually normal to the applied tensile stress. This is called Stage II crack growth.
4. Unstable crack growth with final failure from overload. This is called Stage III crack growth.

Fatigue usually occurs at a free surface, with the initial features of Stage I growth, fatigue cracks, being initiated at slip band extrusions and intrusions [74,75]. Cottrell and Hull [76] proposed a mechanism for the formation of these extrusions and intrusions (shown schematically in Fig. 28) that depends on the presence of slip, with slip systems at 45° angles to each other operating sequentially on loading and unloading. Wood [77] suggested that the formation of the intrusions and
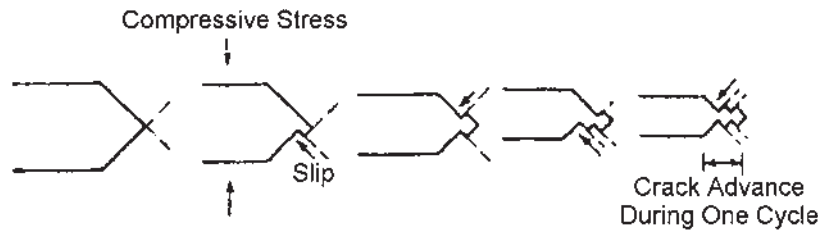
Figure 28   Schematic representation of the mechanism of fatigue intrusions and extrusions.
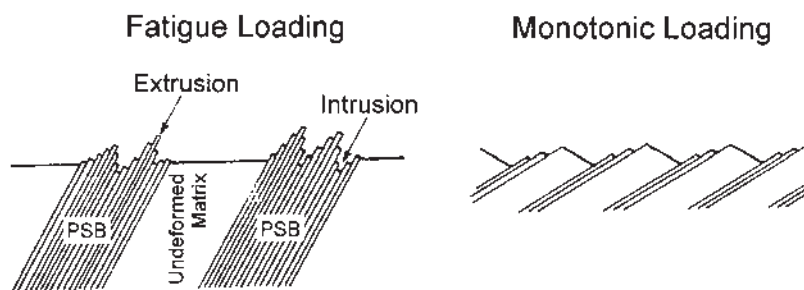
Figure 29   Mechanism of intrusions and extrusions (Courtesy of Wood).

extrusions was the result of fine slip and buildup of notches (Fig. 29). The notch created on a microscopic scale would be the initiation site of stable fatigue crack growth.

In Stage II, stable fatigue crack growth, striations (Fig. 30) often show the successive position of the crack front at each cycle of stress. Fatigue striations are usually detected using electron microscopy and are visual evidence that fatigue occurred. However, the absence of fatigue striations does not preclude the occurrence of fatigue.

Striations are formed by a plastic blunting process [78]. At the end of the Stage I crack tip, there exists sharp notches due to the presence of slip. These sharp notches cause stress to be concentrated at the crack tip. The application of a tensile load opens the crack along slip planes by plastic shearing, eventually blunting the crack tip. When the load is released, the slip direction reverses, and the crack tip is compressed and sharpened. This provides a sharp notch at the new crack tip where propagation can occur. This is shown schematically in Fig. 31.

An alternative hypothesis on striation formation was presented by Forsyth and Ryder [79]. In their model, the triaxial stress state at the crack tip forms a dimple ahead of the crack front. The material between the crack tip and the dimple contracts and eventually ruptures, forming a fatigue striation. This is shown schematically in Fig. 32.

In mild steel, well defined striations are observed, but not as well defined or as spectacular as in aluminum. This was first assumed to be due to the crystal lattice structure, since face-centered cubic austenitic steels show well-defined striations,
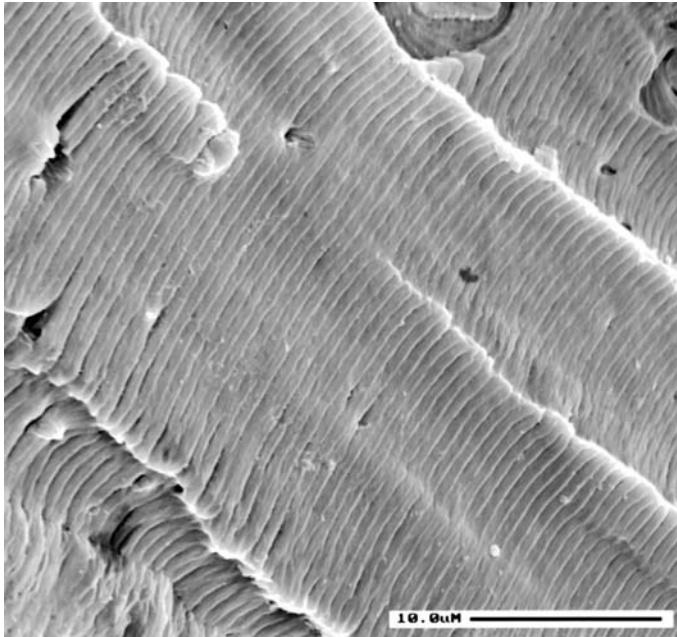
Figure 30   Typical fatigue striations in 7075 aluminum.

Compression Phase

Beginning of
Tension Phase

Plastic Deformation at
Maximum Tensile Stress

Beginning of
Compression Phase
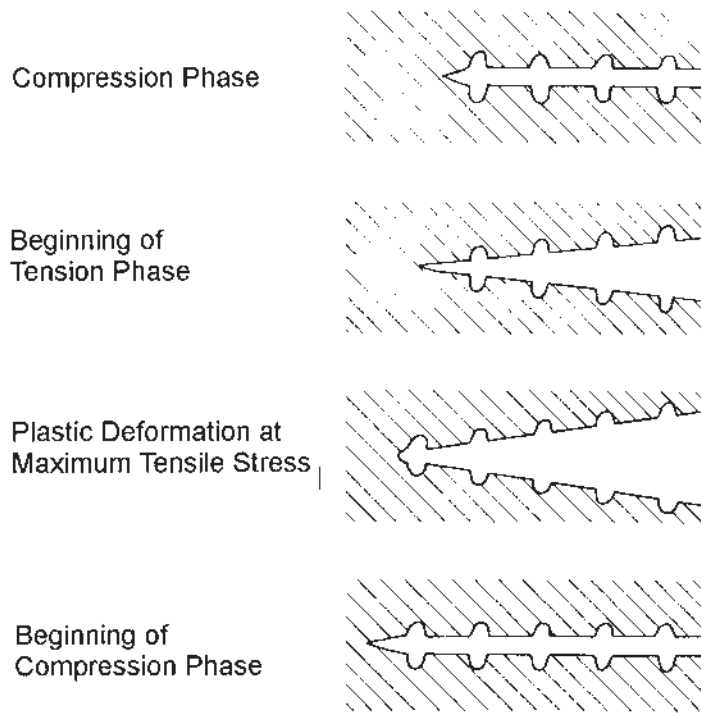
Figure 31   Mechanism for fatigue striation formation.

Ductile   Brittle
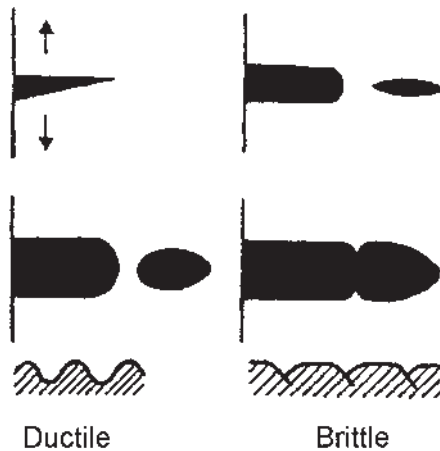
Figure 32   Striation formation from ductile dimple formation ahead of a crack front.

and mild steels (base-structured cubic) do not [80]. Other alloys, such as titanium alloys [81], with a hexagonal close-packed (hcp) crystal structure show very defined striations. However, aluminum alloys (BCC) [82] show strongly defined striations. Therefore, attributing defined striations to crystal lattice alone was discounted as a viable theory.

Deformation and available slip systems were presumed to he more significant [83]. However, this does not follow, because mild carbon steels are more ductile than austenitic steels. It is now generally accepted that fatigue striations form by the plastic blunting process.

5.2   Design for Fatigue

5.2.1   Stress-Life Approach

This approach is the most widely used method of estimating the time until crack initiation. This method is primarily intended for use at high cycles, where stresses are elastic. It does not work well in the low-cycle fatigue regime, or when there is significant plastic deformation at the crack tip. It is usually appropriate for $N > 10^5$ 10^5 cycles.

This method is based on the Wohler or S–N diagram, where the alternating stress is plotted versus the number of cycles to failure. Generally, the data represents reversed rotating bending tests. It treats all strains as elastic.

Steels have a fatigue endurance limit. That is, if the specimen or part is loaded below the endurance limit, then failure will not occur. In general, steels have an endurance limit approximately half the ultimate strength. For steels with an ultimate strength greater than 200 ksi, the endurance limit is generally about 100 ksi. This assumes very polished specimens, with no stress concentrations. Aluminum alloys on the other hand, do not have an endurance limit. A continuously decreasing Stress-Cycle curve is observed. An arbitrary endurance limit for these alloys is established at 10^8 cycles.

Throughout the years, there have been many empirical relationships to predict expected fatigue life. Some examples of these empirical relationships are:

$$\frac{\sigma_a}{\sigma_N} + \frac{\sigma_M}{\sigma_Y} = 1 \qquad\qquad \text{Sonderberg} \qquad\qquad (62)$$

$$\frac{\sigma_A}{\sigma_N} + \frac{\sigma_M}{\sigma_{UTS}} = 1 \qquad\qquad \text{Goodman} \qquad\qquad (63)$$

$$\frac{\sigma_A}{\sigma_N} + \left(\frac{\sigma_M}{\sigma_{UTS}}\right)^2 = 1 \qquad\qquad \text{Gerber} \qquad\qquad (64)$$

$$\frac{\sigma_A}{\sigma_N} + \frac{\sigma_M}{\sigma_f} = 1 \qquad\qquad \text{Morrow} \qquad\qquad (65)$$

In general, the Sonderberg relationship is very conservative, and is rarely used. Most data tends to fall between the Goodman and Gerber curves.

When designing for cyclic loading it is necessary to allow for the factors that influence fatigue. One such method was proposed by Shingley [84]. In this method, correction terms for each of the important deleterious effects are multiplied together to obtain a maximum design stress, $\sigma\prime_e$, for the machine element. This maximum design stress is obtained from data representing a smooth polished rotating beam specimen at the endurance limit σ_e as:

$$\sigma'_e = k_a k_b k_c k_d k_e k_f k_g \sigma_e \qquad\qquad (66)$$

The correction terms K_i, are related to the stress concentration factor by

$$k_i = \frac{1}{K_i} \qquad\qquad (67)$$

where K_f, is the stress concentration in fatigue.

Some of the deleterious effects on fatigue strength or fatigue life are size effects, loading effects, surface finish, surface treatment, residual stresses, temperature, environment and the geometrical stress concentration.

Loading and Size Effects

Aluminum alloys have been found [85] to have a one-to-one correspondence between direct applied stress and rotating bending fatigue strength. The fatigue strength of aluminum is also independent of size effects.

It has also been found [86] that the thicker the test piece, the faster the crack propagation rate. It is likely that the propagation rates for thicker pieces are due to increased plane Strain conditions, with a small plastic zone at the crack tip. Since for a small plastic zone there is a greater stress gradient, a faster crack propagation rate might be expected. Also, in thicker panels there is a higher state of triaxial stress, which would also tend to increase crack growth rates.

Surface Finish

Since fatigue failures usually begin at the surface, the surface condition is very important. Surface roughness is a primary factor influencing fatigue. Highly polished specimens exhibit the longest fatigue life, with increasingly rougher surfaces yielding decreased fatigue life. Fig. 33 shows the correction factor for surface
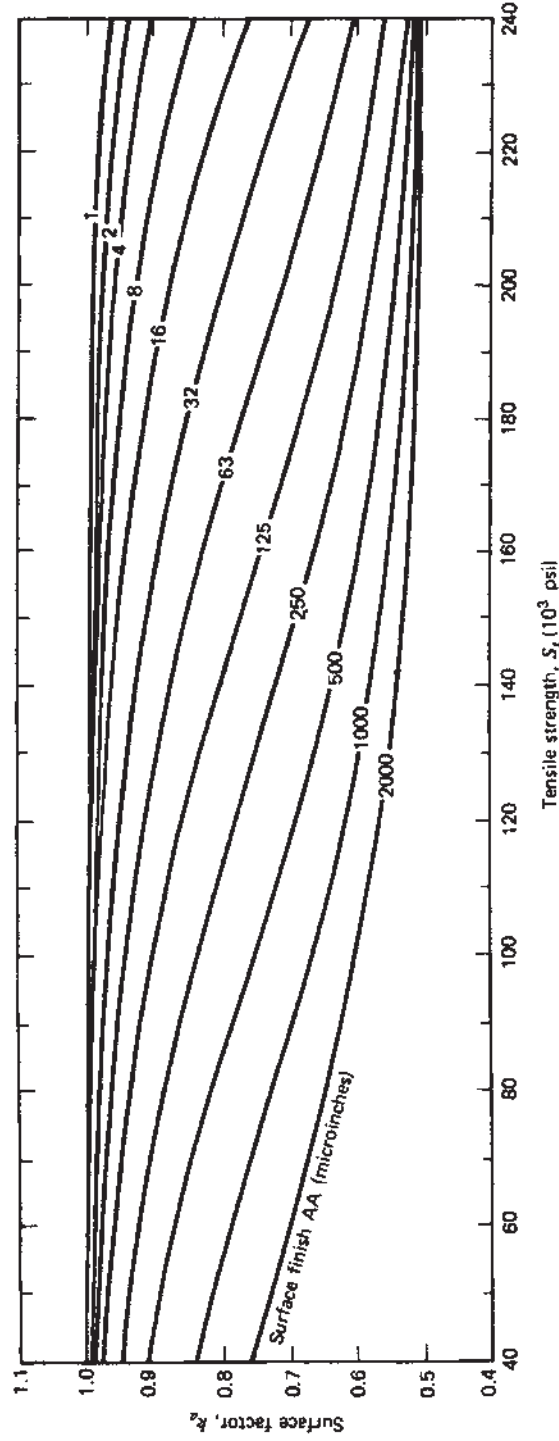
Figure 33 Fatigue stress concentration due to surface finish and surface roughness.

Table 2 Depths of Grooves Commonly Found on Machined Surfaces from Machining

Method	Surface Roughness	Groove Depth (in $\times$ 10^{-3})
Polished	8	0.04
Fine grind	10	0.08
Rough grind	70	0.2–0.4
Fine turn	70–90	0.4–0.8
Rough turn	90–500	0.8–2
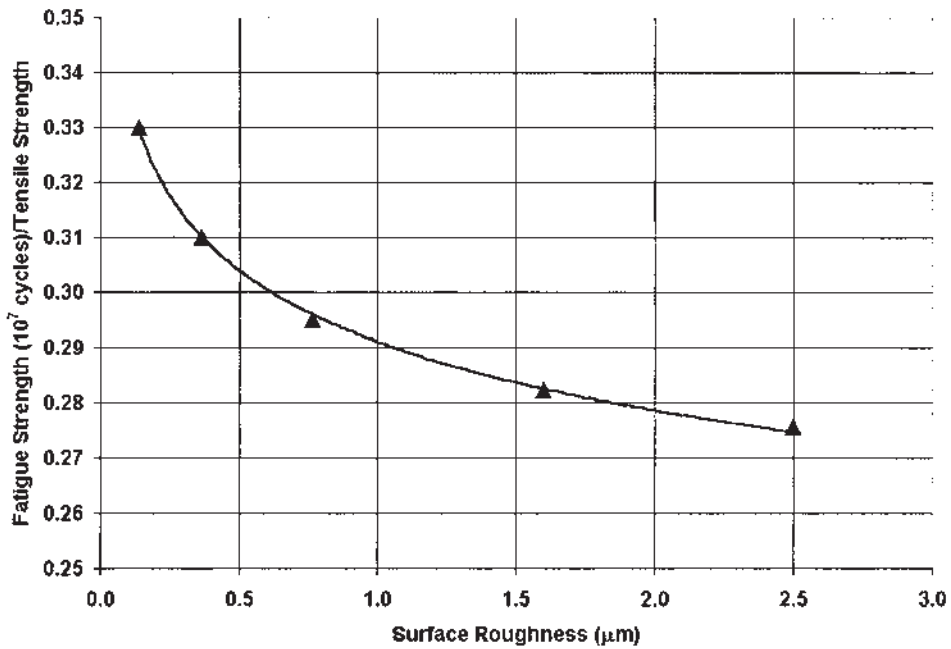
Figure 34   Surface Roughness versus fatigue life for several aluminum alloys.

roughness as a function of strength. While this data is for steel, the general trends are similar. Since the measurement of surface roughness is statistical in nature, extremes of machine grooves can occur. Typical depths of surface grooves from machining [87] are shown in Table 2. These extremes, as well as the stress concentrations they create, must be considered for fatigue.

Rough lathe or coarse grinding reduces the fatigue strength by approximately 20% [88] below polished specimens. A graph of the effect of surface roughness on the fatigue life of aluminum is shown in Fig. 34. Mann [89] evaluated different methods of polishing and tabulated different fatigue limits. These are shown in Table 3.

Table 3   Effect of Polishing Method on Fatigue Strength of 2024–T4
($\sigma_{UTS} = 460$ MN m^{-2})

Surface finish	Fatigue strength ($N = 10^8$)/tensile strength
Circumferentially coarse turned	0.28
Circumferentially diamond turned	0.34
Hand polished 220 grit	0.34
Hand polished 400 grit	0.34
Circumferentially hand polished 220 Grit	0.37

Electropolished specimens have a lower fatigue limits than mechanically polished specimens, by up to 25% [90]. This reduction is due to the removal of surface compressive residual layers induced during mechanical finishing.

Surface Treatments

Unlike steel, there is little information available regarding the effects of surface treatments on the fatigue strength of aluminum. While uncommon, plating of aluminum is done ocassionally. Nickel and Chromium plating has been found to cause substantial reduction in fatigue life [91]. Other studies [92] show no effect of nickel or chromium plating. This study was also a very comprehensive review of the effect of plating on fatigue life.

Anodizing is a electrolytic bath process for producing a hard corrosion resistant film on the surface of aluminum parts. The effects of anodizing on the fatigue strength of aluminum are contradictory. In one study, anodizing was found to have no effect [93], while in another study [94], anodizing was found to reduce the fatigue strength of a Al-5% Cu alloy by 50%. Rotating bending tests [95] on cast and wrought alloys, that were anodized using a 3% chromic acid anodize, had little effect on the fatigue strength when the coating thickness was 0.0025–0.005 mm thick. Subsequent tests [96] on identical specimens, but with a thicker chromic acid anodize layer of 0.1 mm thick caused a 35% reduction in fatigue properties at 10^8 cycles. A 15% sulfuric acid anodize that produced a coating of 0.0125–0.025 mm thick, had no effect at high cycle life (greater than 10^7 cycles), but caused a reduction in the life at higher stress levels. Sulfuric acid anodize coatings 0.025–0.05 mm thick caused a 35% reduction in the fatigue life at 10^8 cycles. In general, the thicker the anodized coating thickness, the larger the decrease in fatigue life [94].

The reductions in fatigue life can be restored by sealing the surface with a boiling dichromate solution. Specimens that were sulfuric acid anodized to a coating thickness of 0.025–0.05 mm thick showed a 40–50% decrease in the fatigue strength (at 10^7 cycles) over uncoated wrought specimens. After sealing the specimens in a boiling dichromate solution, the fatigue strength was improved to within 10% of the uncoated wrought material [96].

Prior to anodizing or other surface treatments, the surfaces of aluminum parts are usually pickled. Acid or caustic acid are used to pickle aluminum alloys. The

Table 4 Effect of Pickling on the Fatigue Strength of 2024 and 7075 after [97]

Specimen Condition	Fatigue Strength at 10^8 cycles (MN m^{-2})	
	2024	7075
As-turned	173	182
As-turned + chromic acid anodizing	179	165
As-turned + acid pickling + anodizing	179	180
As-turned + caustic soda pickling + anodizing	180	200

effect of pickling on the reversed bending fatigue strength of 2024 and 7075 has been studied by [97]. A summary of his results is shown in Table 4. Caustic soda pickling had the greatest beneficial effect. Anodizing without pickling was beneficial in the case of 2024, but was detrimental for 7075. This may be due to the different amounts of residual stresses the different methods produce. Addition information regarding the effects of pickling and anodizing treatments is provided by Harris [98].

Aluminum cladding is done to improve corrosion resistance of sheet products. However, cladding high strength aluminum alloys [99] has been found to reduce the fatigue strength. The fatigue strength of 7075–T6 and 2024–T3 was found [100] to be 50% less in reversed bending and 3% less in reversed direct stress than the unclad condition. Other investigators [101] have also found a 50% reduction in the fatigue strength of 2024 when wrought plate was clad. Sprayed aluminum coatings also have been observed [100] to give similar reductions in the fatigue strength. The reduction in the fatigue strength of pure aluminum cladding on wrought specimens is shown in Fig. 35.

Temperature

In general, there is little information regarding the low temperature fatigue behavior of aluminum. Forest [87] examined the average fatigue strength of aluminum alloys at cryogenic temperatures and found that the fatigue strength increases as a function of temperature (Fig. 36).

The fatigue strength of aluminum is not much effected by temperatures up to about 150°C. But at higher temperatures, the fatigue strength decreases very rapidly. This is due to overaging and coarsening of the precipitates. The best resistance to fatigue at elevated temperatures is shown by alloys that resist overaging. Aluminum alloy 7075 is much more resistant than 2024. Aluminum-silicon alloys also show relatively good fatigue resistance at elevated temperatures (compared to other aluminum alloys). At temperatures above 300°C, the creep and fatigue strength of all aluminum alloys is exceeded by sintered aluminum powder (SAP). This is a powder metallurgy product that generally maintains its strength up to about the melting temperature. Cast alloys in general have poorer fatigue strength than wrought alloys, however the difference in fatigue strength disappears with increasing temperature.
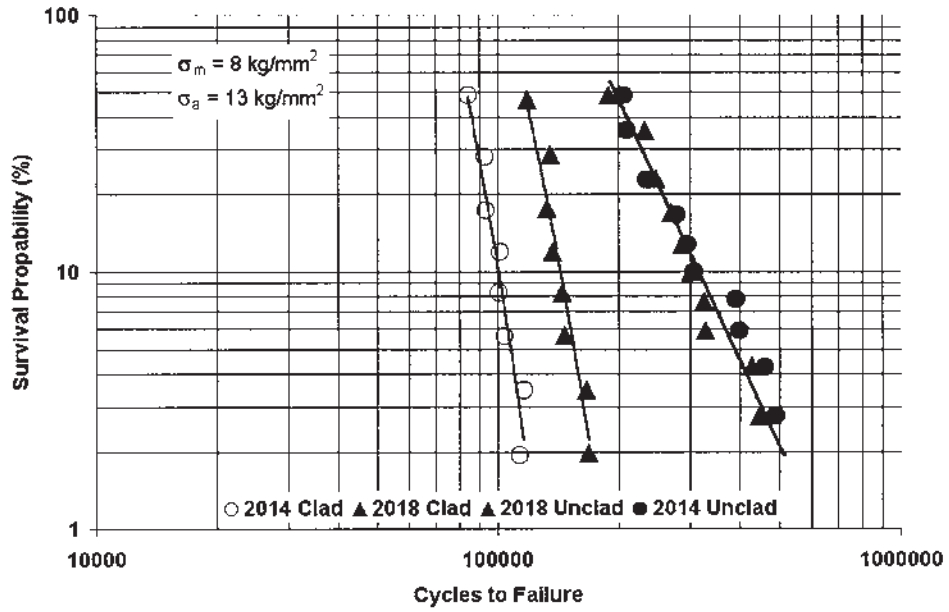
Figure 35 Effect of cladding on the fatigue life of aluminum.
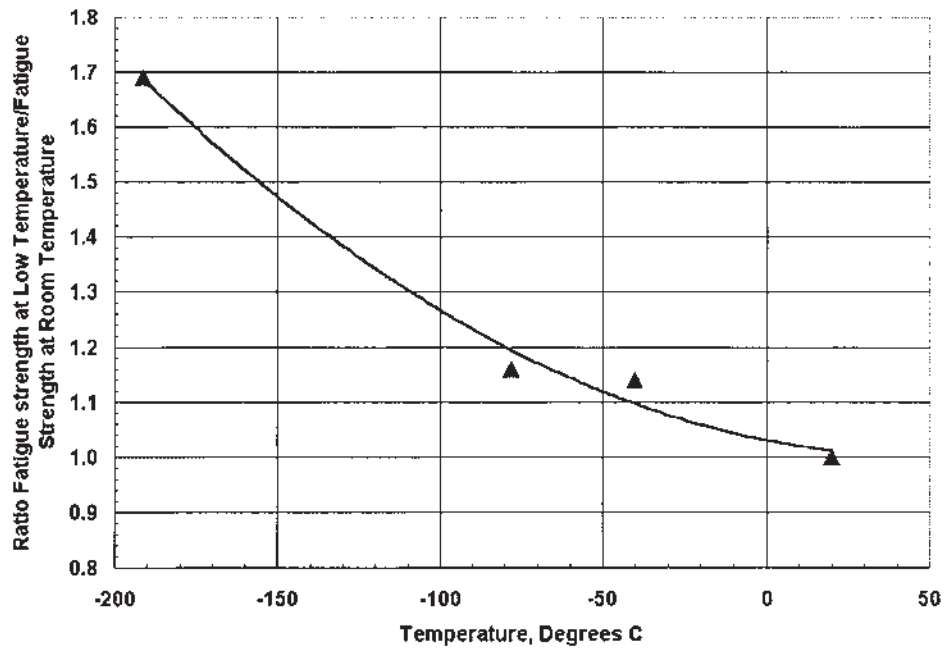
Figure 36 Effect of temperature of the fatigue life of aluminum.

Geometrical Stress Concentration

Test specimens, machined parts or structural members contain geometrical or microstructural notches. These notches are stress concentrations, where the local stress is greater than the average bulk stress. This local stress can be many times greater than the average bulk stress. The ratio of the local stress to the average stress is called the stress concentration factor, K_t:

$$K_t = \frac{\sigma_{\max}}{\sigma_{\mathrm{net}}} \tag{68}$$

This stress concentration is entirely dependent on the geometry and loading. Stress concentrations are determined from geometry effects and are either calculated from elasticity theory [102] *or* with numerical methods or determined experimentally using stress analysis techniques. A very comprehensive set of stress concentration factors has been collected by Peterson [103] and Young [104].

The fatigue life of a component is seriously affected by the presence of stress concentrations. In fatigue, the un-notched fatigue strength is related to the notched fatigue strength by the fatigue stress concentration factor, K_f.

$$K_f = \frac{\sigma_e^{Notched}}{\sigma_e^{Un-notched}} \tag{69}$$

While K_t is dependent on only geometry and loading, K_f is also dependent on the material. To account for this difference, a notch sensitivity factor, q, was developed that relates the stress concentration factor K_t to the fatigue stress concentration factor, K_f:

$$q = \frac{K_f - 1}{K_t - 1} \tag{70}$$

The notch sensitivity, q, of a material has been determined for many materials [105], and some data for aluminum are plotted in Fig. 37. The notch sensitivity varies with the type of notch and notch sensitivity. The notch sensitivity increases with increasing tensile strength.

An alternative approach was taken by Neuber [106], who suggested that the stress concentration for fatigue be expressed as

$$K_f = 1 + \frac{K_t - 1}{1 + \sqrt{\dfrac{\rho}{r}}} \tag{71}$$

where ρ is a material constant. This relationship can be related to the notch sensitivity by:

$$q = \frac{1}{1 + \sqrt{\dfrac{\rho}{r}}} \tag{72}$$

where r is the radius at the root of the notch. The fatigue stress concentration can then be applied to the nominal stress to determine fatigue life using the S–N curve.
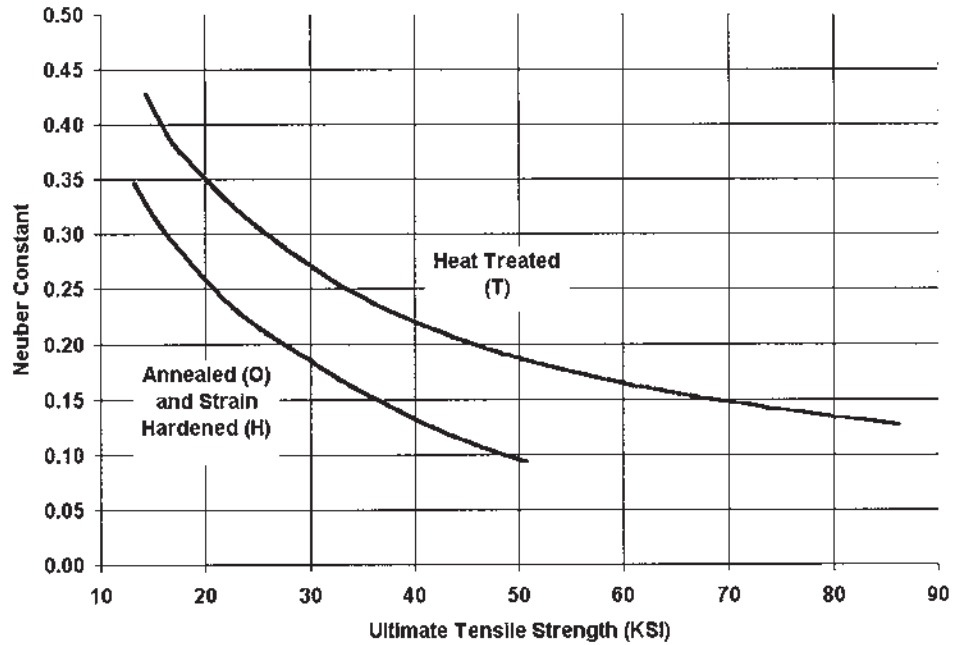
Figure 37 Neuber constants for Aluminum (From Ref. 134.)

5.2.2 Strain-Life Approach

In this approach, it is assumed that in local regions, an adequate stress concentration can occur such that local yielding may occur. This local yielding causes both elastic and plastic strains to occur. The total strain is:

$$\varepsilon_t = \varepsilon_e + \varepsilon_p \tag{73}$$

In most metals, the true stress vs true strain can be represented by the power function

$$\sigma = K(\varepsilon_p)^{1/n} \tag{74}$$

where K is the strength coefficient, and n is the strain hardening exponent. At fracture, the true fracture stress is a function of the load at fracture, P_f, and the instantaneous area at fracture, A_f:

$$\sigma_f = \frac{P_f}{A_f} \tag{75}$$

or

$$\sigma_f = K(\varepsilon_f)^n \tag{76}$$

Rearranging

$$K = \frac{\sigma_f}{\varepsilon_f^n} \tag{77}$$

The elastic strain is

$$\varepsilon_e = \frac{\sigma}{E} \tag{78}$$

So, the total strain, from above is:

$$\varepsilon_t = \varepsilon_e + \varepsilon_p = \frac{\sigma}{E} + \left(\frac{\sigma}{K}\right)^{1/n} \tag{79}$$

During cyclic loading, the strain amplitude can be expressed as:

$$\frac{\Delta\varepsilon}{2} = \frac{\Delta\varepsilon_e}{2} + \frac{\Delta\varepsilon_p}{2} = \frac{\Delta\sigma}{2E} + \frac{\Delta\varepsilon_p}{2} \tag{80}$$

Similar to tension, the power law function for fatigue can be written:

$$\sigma = K'(\varepsilon_f)^{n'} \tag{81}$$

where $K\prime$ is the stable cyclic strength coefficient, and $n\prime$ is the cyclic strain hardening coefficient. Rearranging:

$$\varepsilon_p = \left(\frac{\sigma}{K'}\right)^{1/n'} \tag{82}$$

So the sum of the elastic and plastic components are:

$$\varepsilon = \frac{\sigma}{E} + \left(\frac{\sigma}{K'}\right)^{1/n'} \tag{83}$$

In 1910, Basquin [107] observed that fatigue data, presented in an S–N curve, could be plotted by:

$$\frac{\Delta\sigma}{2}\sigma_f'(2N_f)^b \tag{84}$$

where $\Delta\sigma$ is the true stress amplitude, $2N_f$ are the reversals to failure, b is the fatigue strength exponent (generally between -0.05 and -0.12), and $\sigma_f\prime$ is the fatigue strength coefficient (roughly equivalent to σ_f). Coffin [108] and Manson [109] found that the plastic strain life data could be linearized by the power law function

$$\frac{\Delta\varepsilon}{2} = \varepsilon_f'(2N_f)^c \tag{85}$$

The total strain is equal to the elastic and plastic strains (from above):

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma_f'}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c \tag{86}$$

In general, in a notched test specimen or part, plastic strains will be present at the notch. These strains will effect the prediction and reliability of fatigue life. Knowing the notch strain history and the strain-life history of a test specimen, can allow the designer to predict fatigue life evaluations. This method can also be used to determine the effect of residual stresses.

This method requires that the strains at the notch root be known. This can be determined by experimental stress analysis (photoelasticity or strain gages), finite element analysis (FEA), or by methods that relate the local stresses and strains to the nominal bulk stress.

As was discussed earlier, the stress concentration is used to relate the stress concentration factor to the fatigue stress concentration factor is:

$$K_f = 1 + \frac{K_t - 1}{1 + \sqrt{\dfrac{\rho}{r}}} \tag{87}$$

As the applied stress increases, the stress concentration K_t is constant until yielding occurs. At the onset of local yielding, the local values of stress and strain are no longer related by the stress concentration factor. The stresses and strains after yielding are related to the stress concentration factors:

$$\begin{aligned} K_\sigma &= \frac{\sigma}{S} \\ K_\varepsilon &= \frac{\varepsilon}{e} \end{aligned} \tag{88}$$

where S and e are the engineering stress and strain, and σ and ε are the true stresses and strains. After yielding, K_σ decreases with respect to K_t, while K_ε increases with respect to K_t. Neuber [110] examined a variety of notch geometry and derived the following relation ship:

$$K_t = \sqrt{K_\sigma K_\varepsilon} \tag{89}$$

or

$$K_t^2 = \frac{\sigma}{S} \frac{\varepsilon}{e} \tag{90}$$

This relationship is called Neuber's rule. A modification of Neuber's rule was developed by Seeger and Heuber [111]:

$$K_p^2 S^* e^* = \sigma \varepsilon \tag{91}$$

where

$$\begin{aligned} K_p &= \frac{S_t}{\sigma_y K_t} \\ S^* &= \frac{K_t}{K_p} S \end{aligned} \tag{92}$$

S_L is the limit load.

Topper, et al [112], proposed replacing K_t in Neuber's rule by the fatigue stress concentration factor K_f:

$$\sigma \varepsilon = \frac{K_f S^2}{E} \tag{93}$$

From earlier discussion:

$$\varepsilon = \frac{\sigma}{E} + \left(\frac{\sigma}{K'}\right)^{1/n'} \tag{94}$$

and

$$\sigma\varepsilon = \frac{K_f S^2}{E} \tag{95}$$

Then

$$\left[\frac{\sigma}{E} + \left(\frac{\sigma}{K'}\right)^{1/n'}\right]\sigma = \frac{(K_f S)^2}{E} \tag{96}$$

This equation, is commonly used to estimate the initial loading at the notch, and can be solved iteratively.

Since the notch experiences cyclic stresses, and Neuber's rule must be satisfied, then:

$$\Delta\sigma\Delta\varepsilon = \frac{(K_f S)^2}{E} \tag{97}$$

Using the equation for the strain-life hysteresis curve:

$$\frac{\Delta\varepsilon}{2} = \frac{\Delta\sigma}{2E} + \left(\frac{\Delta\sigma}{2K'}\right)^{1/n'} \tag{98}$$

and combining

$$\frac{(K_f \Delta S)^2}{2E} = \frac{\Delta\sigma^2}{2E} + \Delta\sigma\left(\frac{\Delta\sigma}{2K'}\right)^{1/n''} \tag{99}$$

The values of $\Delta\sigma$ are determined iteratively. Once $\Delta\sigma$ is known, then $\Delta\varepsilon$ can be determined:

$$\Delta\varepsilon = \frac{\Delta\sigma}{E} + 2\left(\frac{\Delta\sigma}{2K'}\right)^{1/n'} \tag{100}$$

From this, the life of the specimen or the part, in the presence of a notch, is determined iterately by:

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma_f}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c \tag{101}$$

5.2.3 Comparison of Methods

Using the Stress-Life approach, the life of a specimen or part can be estimated quickly and simply. Reasonable estimates can be made. This method is good when the stresses are low, constant amplitude, and long life is desired. There are many sources of information and S–N curves, tested under a large variety of circumstances, in a variety of configurations.

Unfortunately, this method is entirely empirical in nature, and does not provide any clue as to the mechanisms involved. This method ignores plastic strains, and assumes that the strains are completely elastic. This limits the usefulness of the method to long life, low stress applications. Also, because all strains are assumed to be elastic, there is no way to allow for the development of residual stresses and strains at the notch root due to variations in the stress cycle. It is very difficult to predict spectrum loading with this method.

The Strain-Life method accounts for the actual response of the test piece or component. Because of this, the plastic strain response at notches, or during high stress, low cycle events, is modeled accurately. Because the plastic strain is modeled accurately, the life of a component can be predicted from spectrum loading, or by momentary overload situations. Complex geometries are more easily modeled, and the method is applicable to elevated temperature testing, because of the decrease in strength and the increase in ductility.

The Strain-Life method is difficult to use and difficult to compute. Using complex load histories mandates the use of at least a programmable calculator. For spectrum loading, a computer is necessary. The entire method is based on a prior knowledge of strains at notches. This must be done using some method like photoelasticity, strain gages, or finite element analysis. Because of this, there are still some very empirical elements of the Strain-Life method. Unfortunately, the effects of fabrication and surface treatment must be evaluated empirically. There is no systematic approach to estimate these effects.

6   CREEP AND STRESS RUPTURE

The effects of temperature on mechanical properties and material behavior are commonplace in everyday living. Example include pipes bursting in the middle of winter, the expansion of a bridge in the middle of summer, and the sagging of a fireplace grate. Each of these examples is an indication that properties change with temperature. In addition, the discussion above indicated that steels become more brittle as the temperature is decreased. There are many other effects of temperature that have been cited [113]. Even the concept of elevated temperature is relative [114]. What is considered hot for one material may be considered cold for another; for instance, gallium has a melting point of 30°C while Tungsten has a melting point of about 3400°C.

Creep is the continuous deformation of a material as a function of time and temperature. This topic is treated very thoroughly by Finnie and Heller [115]. The creep of a material is shown in Fig. 38. It can be seen from the figure that creep in a material occurs in three stages:

1. Stage I, where a rapid creep rate is seen at the onset of load application, then gradually decreasing;
2. Stage II, where creep remains at a steady-state rate
3. Stage III, where the creep rate shows an increasing rate until failure occurs

The behavior and creep rate, are sensitive to the temperature to which the material is exposed, the surrounding atmosphere, and the prior strain history. Andrade and Chalmers [116] were pioneers in the study of creep and proposed that
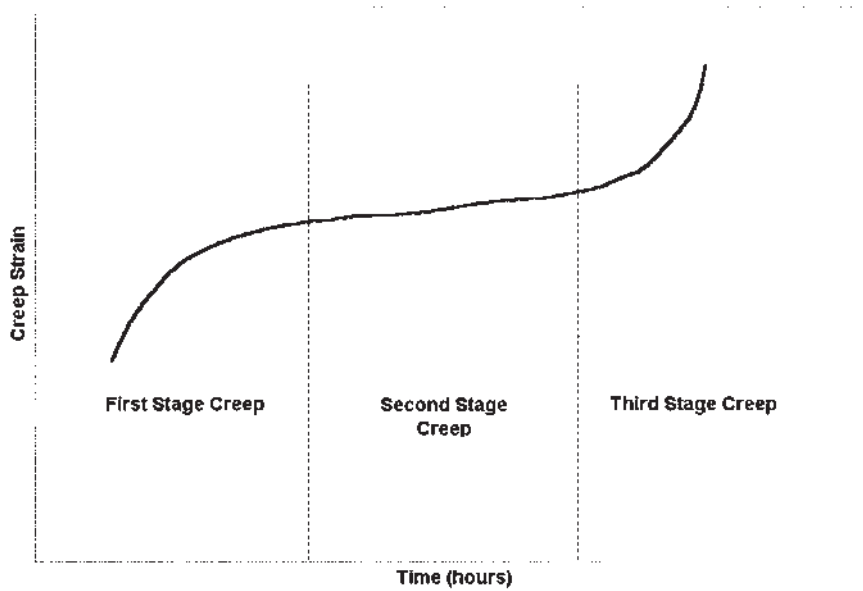
Figure 38 Schematic representation of creep.

creep followed the equation:

$$\varepsilon = \varepsilon_0\left(1 + \beta t^{\frac{1}{3}}\right)e^{kt} \tag{102}$$

where β and K are material constants that can be evaluated by several different methods [117]. A better fit for the creep of materials was proposed by Garofalo [118]. He indicated that

$$\varepsilon = \varepsilon_0 + \varepsilon_t(1 - e^{-n}) + \frac{d\varepsilon}{dt}t \tag{103}$$

where $d\varepsilon/dt$ is the steady-state creep rate, ε_0 is the straln on loading, r is the ratio of the transient creep rate to the transient creep strain, and ε_t is the transient creep strain.

Very early it was recognized that fractures at elevated temperatures occurred along grain facets [119]. In Stage III creep, intergranular wedge cracks and cavities form. Wedge-shaped cracks and creep cavities usually initiate at or near grain boundary triple points and propagate along grain boundaries normal to the applied tensile stress. Creep cavities form at higher temperatures and lower working stresses. These structural features are shown in Fig. 39.

Creep testing is usually performed for 1000–10,000 h with strains of up to 0.5%. Stress rupture testing, or testing to failure, uses much higher loads and temperatures, and the test is usually terminated alter 1000 h. In stress rupture testing, the time to failure is measured at a constant stress and constant temperature. This test has gained acceptance for elevated temperature testing of turbine blade materials in jet engines.
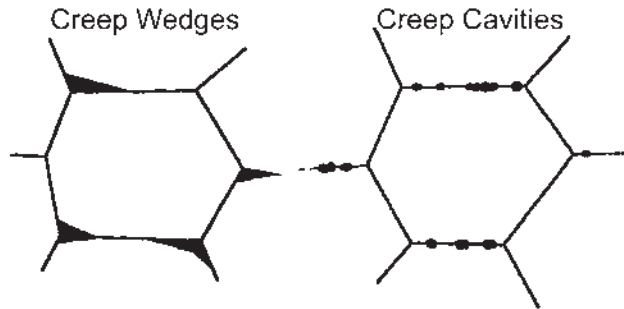
Figure 39   Creep cavities and creep wedges forming at grain boundaries.

Using a tensile machine and high temperature furnace (Fig. 40), the strain is measured in creep testing by special extensometers suited for elevated temperatures. In stress rupture testing, simple apparatus such as dial calipers are used, since only the overall strain at constant time and temperature are needed.

7   TENSION TESTING

Tensile testing is the most generally useful of all mechanical tests. Both the strength and a measure of ductility are obtained. In this test, a specimen is loaded axially and the load is increased continuously until failure occurs. The load and elongation are continuously plotted during testing and then converted to engineering stress a.

$$\sigma = \frac{P}{A_0} \tag{104}$$

and engineering strain e,

$$e = \frac{\delta}{L_0} \tag{105}$$

where P is the load applied, A_o is the original cross-sectional area, δ is the measured elongation, and L_o is the original gauge length. The resulting data are plotted as engineering stress and strain, schematically shown in Fig. 41. For a rigorous mathematical treatment of the tensile test and the changes that occur, the reader is referred to Nadai [120].

At the beginning of the test, there is a linear region where Hooke's law is followed. The slope of this linear region is the elastic modulus E (for aluminum, the elastic modulus is about 10×10^6 psi). The yield stress σ_{YS}, which is the limit of elastic behavior, is defined as the point at which a small amount of permanent deformation occurs. This deformation is defined in the United States as a strain of 0.002 in/in or 0.2% Strain. Loading past this point causes plastic deformation. As the plastic deformation increases, strain hardening occurs, making the material
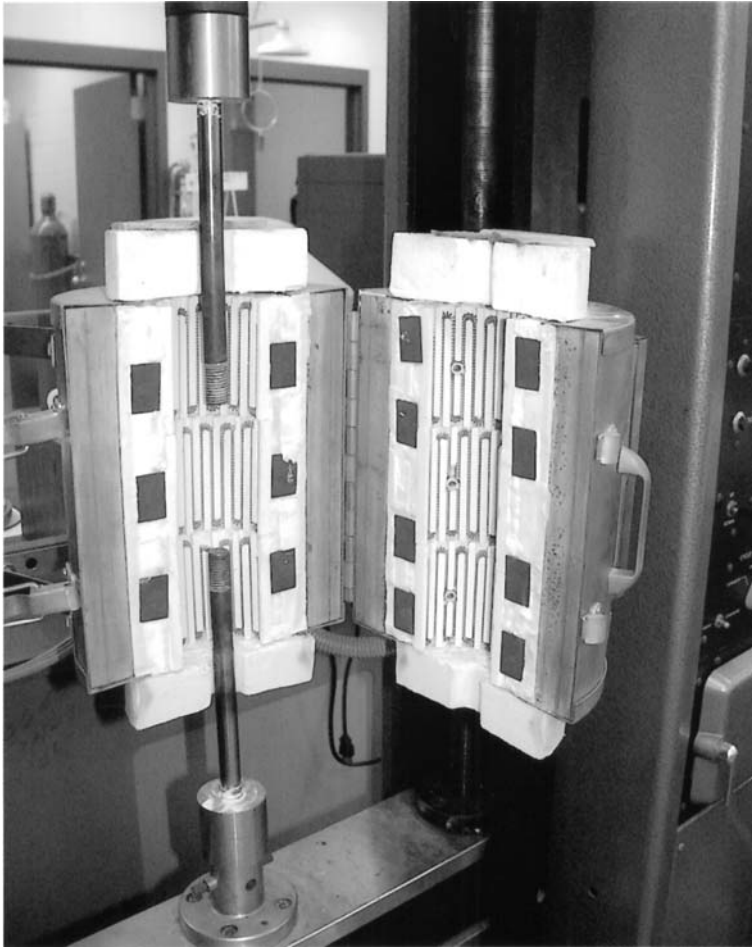
Figure 40 Illustration of creep testing apparatus.

stronger. Eventually, the load reaches a maximum value and failure occurs. The ultimate tensile strength, σ_{UTS}, is determined by taking the maximum load experienced and dividing it by the original area:

$$\sigma_{UTS} = \frac{P_{max}}{A_0} \tag{106}$$

In ductile materials, the strain may increase after the maximum load is reached and the applied load is decreased. At this point is the onset of necking.

The ultimate strength is the most quoted property but is the least useful of all the properties determined from the tensile test. In ductile materials, ultimate strength is a measure of the maximum load that the material can experience in uniaxial loading. However, in brittle materials, the ultimate tensile strength is valid design information. The ultimate tensile strength used to be the basis for many designs
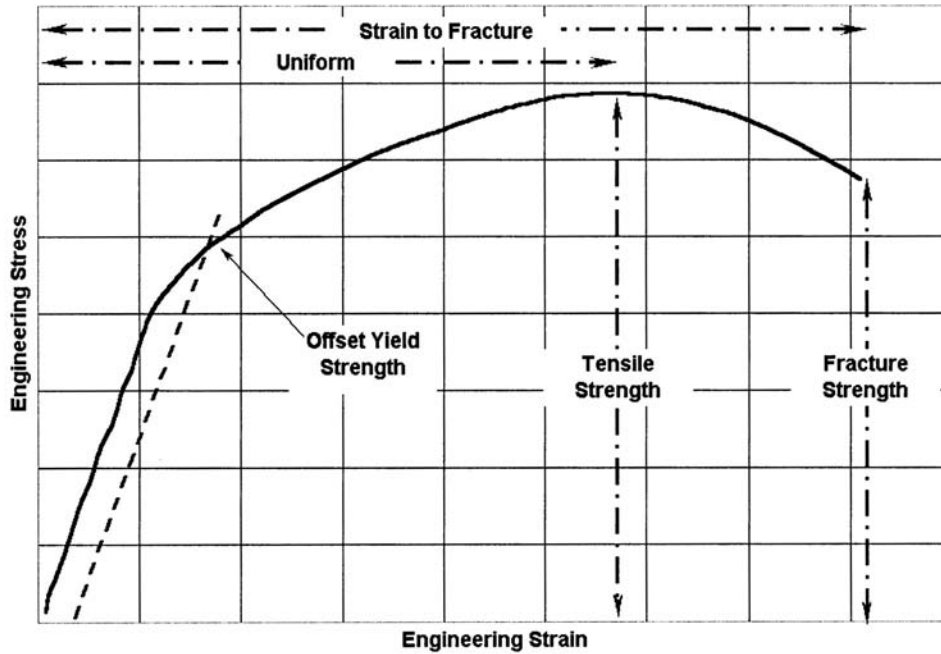
Figure 41   Schematic representation of tensile test data.

and design codes (with a margin of safety), but now design codes rely on the yield strength instead. Because of its reproducibility, the ultimate tensile strength is often used for procurement, specifications, and quality documents.

The yield strength σ_{ys} is the stress required to obtain a small (0.2%) permanent strain. In other words, if the specimen were unloaded at the yield stress, it would be 0.2% longer than originally. The yield strength is now the basis for many design codes and is used extensively in the determination of many other properties such as fracture toughness and fatigue strength.

Ductility in the tensile specimen is measured and reported two different ways: In terms of elongation,

$$\%E = (100)\frac{L_f - L_0}{L_0} \tag{107}$$

and in terms of the reduction in area,

$$\%RA = (100)\frac{A_f - A_0}{A_0} \tag{108}$$

where L_f is the final gage length after fracture, L_o is the original gauge length, A_o is the original cross-sectional area, and A_f is the final cross-sectional area after fracture. Since the final strain will be concentrated in the necked region, the
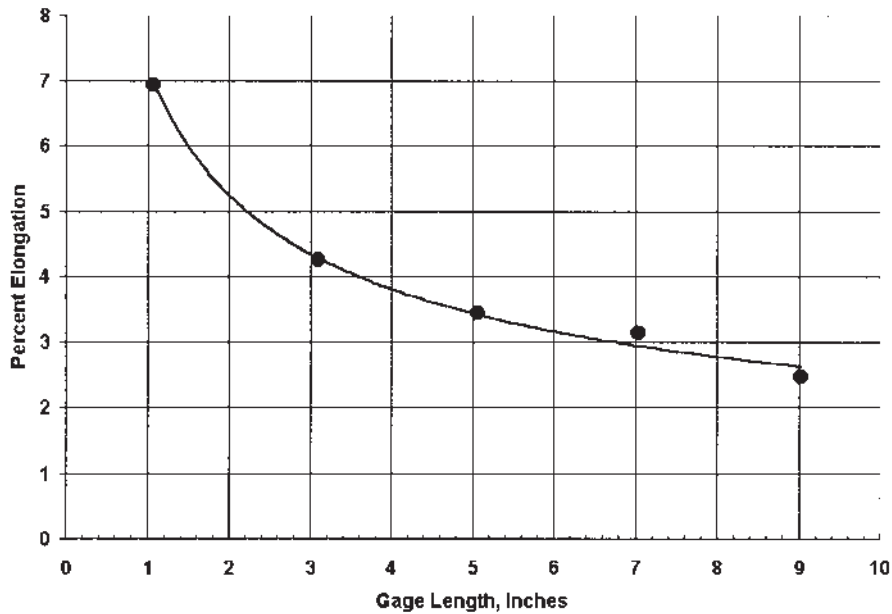
Figure 42 The relation between gage length and elongation.

elongation is dependent on the gauge length L_o. For that reason, in reporting elongation, the gauge length is always provided. The change in elongation as a function of gauge length is provided in Fig. 42.

Testing machines used for tensile testing are generally simple. Either screw-type (Fig. 43) or hydraulic (Fig. 44) machines are used. The load is measured by a load cell composed of strain gauges or an LVDT (linear velocity displacement transducer). The load is recorded by a chart recorder as a plot of load versus strain.

The strain is measured by extensometers, which measure in situ the strain experienced by the tensile specimen. Modern extensometers use strain gauges to provide strain measurements. This is an advantage because it sends an electric signal proportional to the experienced strain to the chart recorder.

The shapes of tensile specimens are standardized by ASTM. Their configurations are shown in Fig. 45, and their dimensions are given in Table 5. Care in fabricating the tensile specimen is necessary to preclude faulty or inaccurate results. Heating and cold working of the specimens must be minimized during fabrication or inaccurate yield and ultimate stress values may result. The specimens must be straight and flat; otherwise a distortion of the elastic region and inaccurate yield strength may result. During machining, the tensile specimens must be symmetrical about the load axis. If not, bending may occur, leading to erroneous results because *of* the combined stresses of tension and bending.

The tensile specimen is held during testing by either hydraulic or mechanical grips. Both methods grip the specimen by wedges with serrated surfaces. In mechanical grips, the load applied to the specimen forces the wedges tighter against the specimen by inclined surfaces inside the grip. With hydraulic wedges, hydraulic pressure is used to force the wedges against the specimen and hold it in place.

Figure 43   An example of a screw-type tensile testing machine.

For light strength steels or steels that have hard or highly polished surfaces, hydraulic grips are preferred to prevent the specimen from slipping in the grip.

8   HARDNESS TESTING

Hardness testing is probably the most common type of mechanical test performed in the United States, and perhaps the most common worldwide. The term "hardness" is poorly defined and is relative to the measuring device. There are three basic types of hardness tests: the scratch test, the indenter, and the dynamic rebound test. The scratch test is familiar to mineralogists, who use the Mohs scale.

The Mohs scale is the relative hardness of 10 minerals arranged in order. Talc is the softest (Mohs 1), and diamond is the hardest (Mohs 10). Most hard metals fall in the range of Mohs 8. There is inadequate differentiation along the scale to be of much use to a metallurgist. In dynamic tests, an indenter is dropped on the material, and

Figure 44 An example of a hydraulic tensile testing machine.
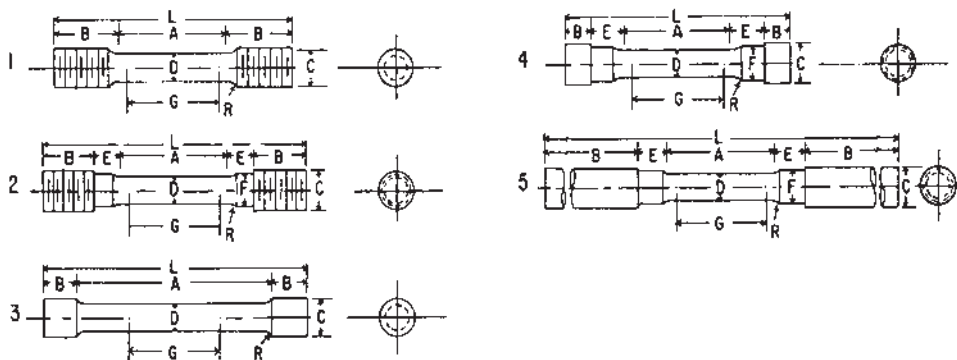
Figure 45 Tensile specimen configurations allowed by ASTM E8.

Table 5   Dimensions of Tensile Specimens per ASTM E8 and Fig. 45

	Specimen dimensions (mm)				
	1	2	3	4	5
G. Gauge length	50	50	50	50	50
D. Diameter	12.5	12.5	12.5	12.5	12.5
R. Fillet radius	10	10	2	10	10
A. Length of reduced section	60	60	100	60	60
L. Overall length	125	140	140	120	240
B. Length of grip section	35	25	20	13	75
C. Diameter of grip section	20	20	18	22	20
E. Length of shoulder	–	16	–	20	16
F. Diameter of shoulder	–	16	–	16	15

hardness is defined as the energy of impact. The dynamic hardness test is commonly used for rubbers and polymers. One exception is the Shore sceleroscope which is used for metals. The indenter type of hardness test is the most widely accepted for metals.

In the indenter type of hardness test, an indenter is pressed into the material and released, and either the diameter or the depth of the impression is measured. The load and the impression measurement determine the hardness. Since a hardness impression is made, there is a plastic zone around the hardness indentation that is surrounded by undisturbed elastic material. This elastic zone hinders plastic flow. Since the plastic region is constrained by the elastic region, the compressive strength of the material in the area of the hardness impression is higher than the value of simple compression. This is a classic problem in plasticity and should be able to be explained by slip line theory. The load required to indent a specific distance 3 by a punch is given by Hill [121] as

$$P = \frac{4a\sigma_{YS}}{\sqrt{3}}\left[1 + \frac{\pi}{2} - \left(\frac{\delta}{2R}\right)^{\frac{1}{2}}\right] \tag{109}$$

where R is the radius of the punch, P is the applied load, and r is the resulting radius after punching. Nadai [120] determined the pressure to indent, using slip line theory, and found

$$P = \frac{2\pi}{\sqrt{3}}\left[\frac{1}{2} + \frac{\pi}{4} + \alpha + \frac{\cot\alpha}{2}\right] \tag{110}$$

where α is the included angle of the conical indenter.

In hardness testing. the impression is asymmetric, so that the slip line theory is not applicable [122] [123], but an elastic/plastic boundary problem, best explained from the Hertz theory of contact stresses [124]. This model accounts for the material displaced by the indenter by the decrease (by compression) in volume of the elastic

underlying material. No upward flow around the indenter is predicted. This agrees with observation. This explanation is the basis for all indention hardness tests used for metallic materials.

8.1 Brinnell Test

The Brinnell test was first proposed in 1900 and has since become widely accepted throughout the world. It is accomplished by indenting the surface with a 10 mm steel ball with a 3000 k load. For soft metals a 500 kg load is usually used because otherwise the impression is too deep. For hard metals, a tungsten carbide ball is used to prevent distortion of the indenter. The diameter of the round indentation is measured with a low power microscope after loading. At least two measurements of the diameter are made, and the results are averaged. To ensure that accurate measurements are made, the surface must be free of dirt and scale. The hardness expressed as a Brinnell hardness number (BHN) is determined from the equation

$$\text{BHN}\left(\frac{\text{kg}}{\text{mm}^2}\right) = \frac{2P}{\pi D}\left[D - (D^2 - d^2)^{\frac{1}{2}}\right] \tag{111}$$

where P is the applied load (kg), D is the diameter of the indenter (10 mm), and d is the measured diameter of the impression. The Brinnell number could also be calculated by measuring the depth of the impression, t

$$\text{BHN} = \frac{P}{\pi Dt} \tag{112}$$

or

$$\text{BHN} = \frac{2P}{\pi D^2(1 - \cos\phi)} \tag{113}$$

since $d = D \sin \varphi$, where φ is the included angle of the chord of the impression.

Because of the large size of the impression made in the Brinnell test, it averages out any local inhomogeneities. It also precludes the testing of small objects or objects in which the Brinnell impression can be a site of crack initiation. Therefore, it is commonly used for castings, forgings, or raw stock. The size of the piece tested for Brinnell hardness should be at least 10 times the depth of the impression. Because of the plastic zone surrounding the impression and the elastic constraint, additional Brinnell hardnesses should not be measured any closer to the impression than 4 times the impression diameter. The distance from an edge when taking a hardness reading should also be at least 4 times the impression diameter.

8.2 Vickers Hardness or Diamond Pyramid Hardness

In the Vickers hardness test, a square-based pyramid shaped diamond penetrator is used. The included angle of the pyramid is 136 $\sim$ between opposite faces. The test

is taken by indenting, under load, with the penetrator, and after release of the load, measuring the width of the diagonals. The Vickers hardness number (VPH) or diamond pyramid hardness is calculated as

$$\text{DPH} = \frac{2P \sin\left(\frac{\theta}{2}\right)}{L^2} \tag{114}$$

where P is the load (kg), L is the average length of the diagonals, and θ is the angle between opposite faces (136°).

The Vickers hardness test has gained wide acceptance in the world for research because the loads can be varied between 1 and 120 kg. It is a continuous scale of hardness that is internally consistent; i.e. hardness numbers determined from one test with one load can be compared to those determined in another test using a different load (except for very light loads) because the impressions are geometrically similar. Other tests require changing loads and indenters, and the hardnesses obtained cannot be strictly compared with others.

This test has not been widely accepted in the United States because it is slow and requires careful surface preparation. In addition, the opportunity exists for operator error when measuring the diagonals.

Sometimes, impressions are obtained that are not perfectly square. These fall into two categories: pincushion-shaped impressions and barrel-shaped impressions (Fig. 46). The pincushion-shaped impression is caused by the tested metal sinking in around the flat faces of the pyramid. This often occurs when testing very soft or annealed metals, with the result that inaccurate low hardness numbers are obtained by overestimating the diagonal lengths. Barrell-shaped impressions are usually found when testing highly cold worked materials and are caused by the piling up of material around the indenter. This produces low diagonal values, with the hardnesses erring on the high side. Asymmetrical indentations are caused by improper alignment of the test piece or that the indenter is not normal to the test piece surface.

8.3   Rockwell Hardness Test

The Rockwell hardness test is very widely used in the United States because of its speed and its freedom from errors by operating technicians. The impression is small, making it possible to test a wide variety of parts.
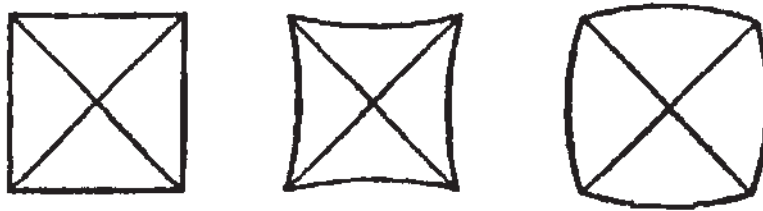
Figure 46   Typical Vickers hardness indenter defects.

In this test, a minor load of 2 kg is applied to the pair to seat the indenter and part; then the major load is applied. After 30.5 the depth of the impression is measured and exhibited on a rotary dial or, in newer machines, on a digital readout. Some machines are equipped with printer ports or RS-232C interfaces for communicating with a computer.

The scale used in Rockwell testers is based on 100 divisions, with each division equal to a depth of 0.00008 in. The scale is reversed, so that a small impression results in a high hardness number. The number from the Rockwell test is an arbitrary number that is only consistent within the same scale. This is unlike the Brinnell and Vickers hardness tests, which provide numbers that are based on mass per unit area (kg/mm^2).

More than one indenter can be used with the Rockwell test. The most common is the Brale indenter, which is a 120° diamond cone and is used for the Rockwell C test. Other indenters used are 1/16 and 1/8 in diameter steel balls. Major loads can be varied from 60 to 150 kg. Table 6, Description of Standard Rockwell Hardness Scales, lists the application, indenter, and major load used for the three types of Rockwell hardness tests.

Because the scale on a Rockwell hardness machine is arbitrary, it is necessary to ensure that one machine records the same hardness as another machine. This is accomplished by the use of standardized test blocks calibrated by the manufacturer. Generally, three test blocks of values throughout the range are adequate to maintain the hardness machine in calibration. The hardness machine calibration should be verified every time the hardness machine is used. Results outside the hardness block calibration is cause to require a new calibration to be performed.

8.4 Rockwell Superficial Hardness Test

The superficial tester operates in an identical fashion to the Rockwell hardness tester. In fact, they lock very similar. Two types of indenters are used: a 1/16 in diameter steel ball (used for surface hardnesses of brasses, aluminum sheet, etc.) and the Brale penetrator. This Brale penetrator is similar to the one used in the Rockwell hardness tests except that its spherical end is shaped to higher tolerances. This penetrator is designated the N-Brale indenter.

Table 6 Description of Standard Rockwell Hardness Scales

	Hardness scale		
Scale	Rockwell A	Rockwell B	Rockwell C
Indenter	Brale	1/16" diameter ball	Brale
Major load (kg)	60	100	150
Application	Very hard materials; tungsten carbide, hard thin materials, carburized or nitrided surfaces	Rolled steel sheet, brass, aluminum annealed steels, etc.	Fully hardened steels, carburized steels, etc.

Table 7   Indenters and Major Loads used for the Rockwell Superficial Hardness Test Scales

Scale	Indentor	Major Load (kg)
15-N	N Brale	15
30-N	N Brale	30
45-N	N Brale	45
15-T	1/16″ diameter steel ball	15
30-T	1/16″ diameter steel ball	30
45-T	1/16″ diameter steel ball	45

The minor load used to seat the indenter and specimen is 3 kg. The major loads used are 15.30, and 45 kg. The major load is applied for 30.5, and the depth of the penetration is measured. The scale of the superficial Rockwell test is arbitrary like that of the Rockwell test, except that each division represents 0.001 in in depth. The scale is reversed, so deeper impressions mean lower hardness numbers. Since the loads used on the superficial Rockwell tester are different from those used for the Rockwell tests. Different scales were established defined by the load and indenter as shown in Table 7.

Since the impression left by the superficial test is not very deep, it is important that the test surface be smoother than is necessary for the standard Rockwell test. The superficial hardness test is more accurate when the part has a good surface finish. the test is sensitive to thin and hard particles under the test piece or the indenter, giving unusually high readings. Soft particles underlying the test piece would provide lower readings.

8.5   Tukon Microhardness Test

The Tukon or Knoop microhardness test was developed in 1939 [125] by the National Bureau of Standards as a method of measuring the hardness of very small constituent phases, segregation effects, and hard and brittle materials. With proper selection of the load, the depth of the impression will not exceed 1 μm.

The Knoop penetrator is a pyramidal diamond, cut to have an included transverse angle of 130° and an included longitudinal transverse angle of 170° 30′ [126]. The resulting impression is rhombic, with the long diagonal approximately seven times the length of the transverse diagonal, as shown in Fig. 47.

The test is conducted by placing the prepared metallographic specimen on a microscope stage, and the desired location for the impression is determined with the aid of a metallographic microscope. Once located, the indenter stage is located over the desired area, and the tester is actuated. The specimen is moved upward automatically by means of an elevating screw until it makes contact with the indenter. The preload, is applied for approximately 20 sec, then the selected major load is applied gradually, reaching maximum load in approximately 20 sec. The load is removed, and the specimen is lowered. The operator moves the microscope to view the impression, and the impression's long diagonal is measured.
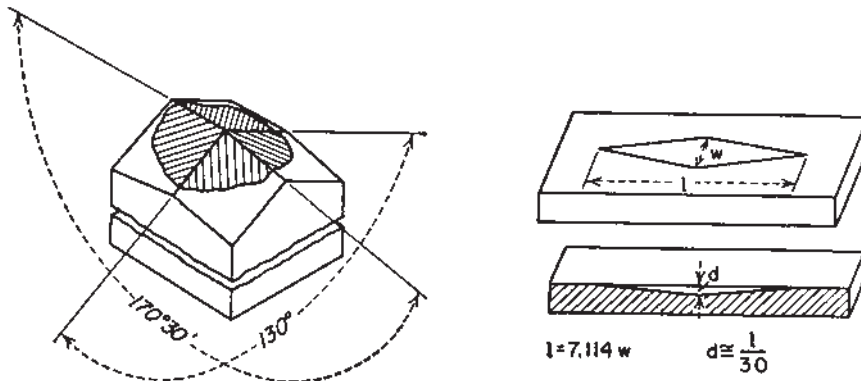
Figure 47   Schematic of the Knoop or Tukon indentor (Courtesy of Kehl).

The longitudinal diagonal of the Knoop impression is measured with a filar eyepiece, which is part of the microhardness tester. This eyepiece has a fixed micrometer scale with a movable vertical hairline. The micrometer screw is divided into 100 divisions, with each division corresponding to a lateral movement of 0.01 mm. Movement is maintained in one direction to eliminate any errors due to backlash in the micrometer gear train.

Using the filar eyepiece, the long diagonal is measured, and the Knoop hardness number (KHN) is calculated:

$$\text{KHN} = \frac{P}{0.7028L^2} \tag{115}$$

where P is the applied load and L is the length of the long diagonal (mm). Since the impression left by the microhardness tester is very small, it is necessary that the surface to be measured be prepared metallographically and be free of any surface scratches or other detrimental defects. The surface may be either etched or unetched. In general, better accuracy can be achieved if the specimen is unetched. It is very important that the specimen be perpendicular to the penetrator. If the impression is lopsided, the measurements of the long diagonal will be inaccurate because of asymmetric elastic recovery. The specimen must be removed and re-leveled and oriented property so that accurate measurements can be taken.

9   TOUGHNESS TESTING

9.1   Dynamic Tear Test

The dynamic tear test was developed to measure the transition temperature of the steels used in the pipe industry. It is an ASTM standard [127] and is essentially a large scale Charpy V notch test. ASTM E604 has established the width of the specimen to 0.625 in, but thicknesses as great as 12 in have been tested. Typical specimen dimensions are shown in Fig. 48.
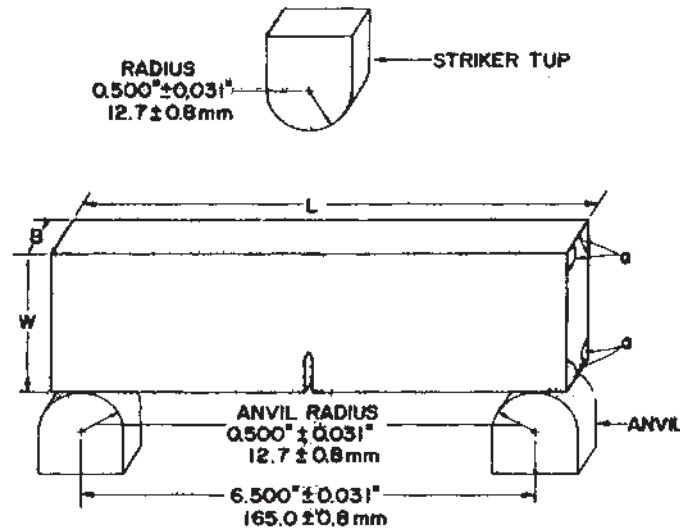
Figure 48   Dynamic tear test specimen.

The test involves the dynamic tear specimen being impact loaded in three-point bending and the energy absorbed is measured. Other measurements performed include the percentage of brittle or ductile failure to determine the FATT temperature. The testing machine can either be as pendulum type like a Charpy V notch machine or a drop-weight type with the capacity to break the specimen in a single blow. For most materials, with a 0.625 in (16 mm) thick specimen, the capacity necessary to conduct the dynamic tear test is approximately 2700 J (2000 ft/Lb$_f$).

The notch on the specimen is prepared by machining a notch 0.475 in deep with a radius of 0.0625 in. The angular root section of the notch is usually made by a precisely ground saw or an electric discharge machine. The notch if further sharpened by pressing a hardened high-speed tool steel (HRC 60 mm) knife blade into the machined notch to a depth of 0.130 in. Notches have also been made with an electron beam welder. The high energy density of the beam creates a very sharp, brittle, and well-defined heat-affected zone. The notch can also embrittled by the use of alloying elements.

The strain rate of the dynamic tear test is similar to that of the Charpy V notch test at approximately 10^{-3} s^{-1}. Because of the greater section thickness, a higher state of triaxial stress exists in the specimen during testing than in the Charpy V notch test. A plane strain condition will be reached earlier than in the charpy test as the testing temperature is decreased.

9.2 R-Curve Analysis

When a crack extends by a small amount, the surface energy of the crack increases. This can be represented by the equation:

$$G = -\frac{1}{B}\left(\frac{\partial U_a}{\partial a}\right)_v \tag{116}$$

where the increase in surface energy is represented by the partial differential. When this above equation is reformulated according to Irwin [128], this results in:

$$G = \frac{\pi\sigma^2 a}{E} = G_c = 2\gamma_e = R \tag{117}$$

where R is the crack extension resistance and is a constant. If this is applied to both elastic and plastic materials that exhibit plastic deformation, then there is a component of elastic crack surface energy and plastic surface energy increase during crack extension:

$$G = \frac{\pi\sigma^2 a}{E} = G_c = 2(\gamma_e + \gamma_p) = R \tag{118}$$

When the tested materials are ductile, then γ_p is much greater than γ_e. Therefore the crack resistance, R, is primarily plastic surface energy and can therefore be neglected. Because of the large amount of plastic energy, R is not necessarily a constant any longer. This is because R and G_c are only a constant under conditions of plane strain. Therefore, at instability under conditions of plane strain:

$$G_{Ic} = (1 - v)\frac{\pi\sigma^2 a}{E} \tag{119}$$

or

$$\sigma_c = \sqrt{\frac{EG_{Ic}}{(1 - v)\pi a}} \tag{120}$$

From these equations, it can be seen that higher stresses cause instability at shorter crack lengths. Now imagine that a cracked specimen is held in a tensile machine. If the load is held constant, then the crack resistance increases linearly with increasing crack length. But if the grips are fixed, and the load is allowed to decrease as the crack extends, then the specimen stiffness will decrease. The resulting crack resistance will also increase less strongly. This is illustrated in Fig. 49. Instability, or final fast fracture, is preceded by slow stable crack growth when the specimens are predominately under plane stress.

Because slow stable crack growth can not be predicted well using the plane strain fracture toughness specimen (K_{Ic}), the R-curve test is well suited to characterize slow stable crack growth. Essential an R-Curve is a plot of the crack resistance as a function of crack extension. This resistance is typically expressed in terms of the stress intensity factor.

$$K_R = \sqrt{E'G} \tag{121}$$

R-curves are determined using load or displacement control. Using load control, the load is increased, and the crack grows, then blunts. This is repeated until the stress intensity at the crack tip is equal to the critical stress intensity. Final fast fracture or instability then occurs.
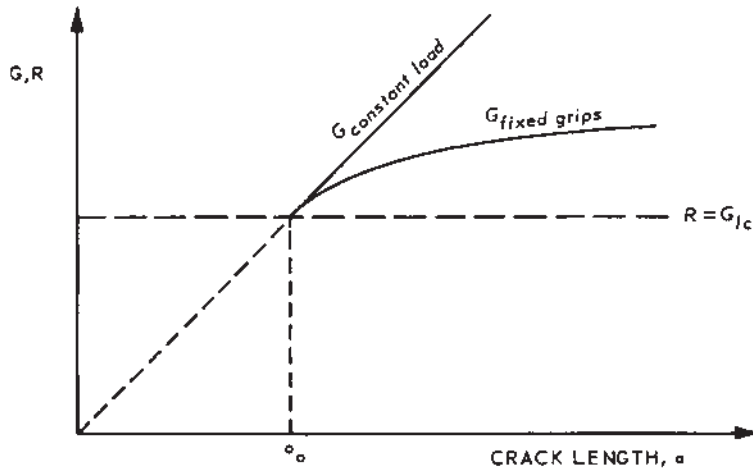
Figure 49   Difference in the crack resistance of a cracked panel under constant load, and a decreasing stiffness.

Using displacement control, the crack is forced open using a wedge. The wedge is progressively inserted to gain larger and larger displacements. Because of the geometry of the test, the entire R-curve can be obtained.

R-Curve testing has been accepted as a standard test [129]. In general there are three types of specimens: the center-cracked tension (CCT) specimen; the compact (CT) specimen; and the crack line wedge loaded (CLWL) specimen. The CCT and CT specimen are tested in load control, while the CLWL specimen is loaded by displacement control. The specimen configurations are shown in Fig. 50.

In general, the specimens are fatigue cracked. A starter notch developed by electric discharge machining (EDM) is often adequate. The initial crack length is generally no more than 40% of the specimen width.

The specimen size must be selected so that the uncracked ligament is loaded elastically. The specimen is loaded incrementally, and allow enough time between loading to allow the crack front to stabilize. The length of the crack is measured using either electrical resistance, or by an optical microscope. The load versus crack extension is used to calculate the stress intensity factor, or K_R. For the CCT specimen, the stress intensity is:

$$K_I = \frac{P}{BW}\sqrt{\pi a \sec\left(\frac{\pi a}{W}\right)} \tag{122}$$

or

$$K_I = \frac{P}{BW}\left[1.77 - 1.77\left(\frac{2a}{W}\right) + 1.77\left(\frac{2a}{W}\right)^2\right] \tag{123}$$

B is the material thickness, W is the specimen width, and a is the crack length.
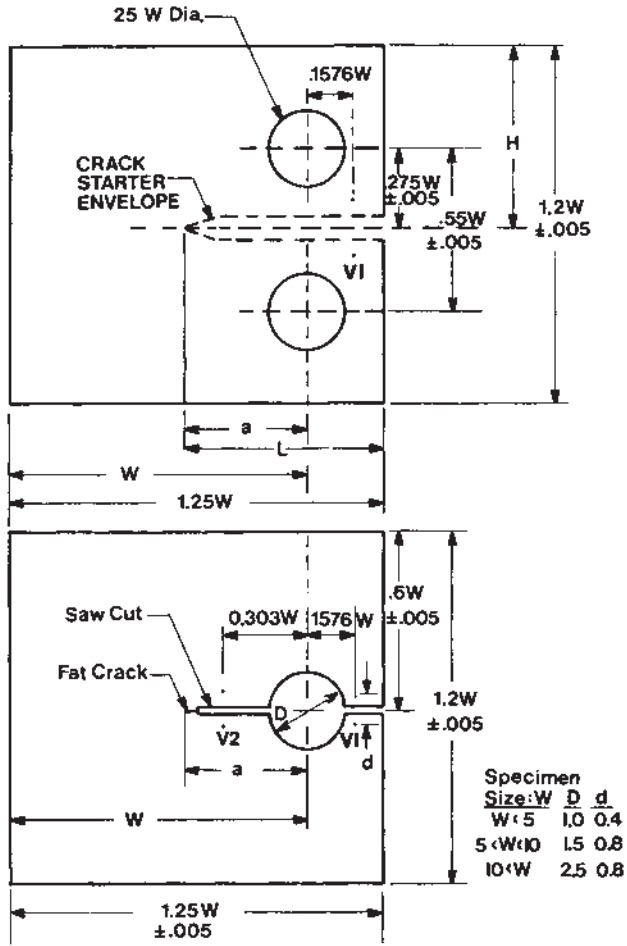
Figure 50 Typical configuration of specimens for *R*-Curve testing.

9.3 Fracture Toughness (K$_{Ic}$) Testing

As indicated above, the use of fracture mechanics is important in determining the maximum flaw Size that a material can withstand before failing catastrophically. As has been noted, cracking in a thick plate is worse than in a thin plate. This is because of plane strain conditions. At the crack tip, the plastic zone is small, with a high Stress gradient across the plastic zone. A schematic of the plastic zone is shown in Fig. 51. In addition, very high triaxial Stresses are present. Because of this, the fracture appearance changes with specimen thickness. This is shown in Fig. 52.

In thin plates, the fracture is characterized by a mixed mode ductile and brittle fracture, with the presence of shear lips. Under plane strain conditions, when the plate is thick enough the fracture is flat, and the fracture stress is a constant with increasing thickness. The minimum thickness for plane strain conditions to occur
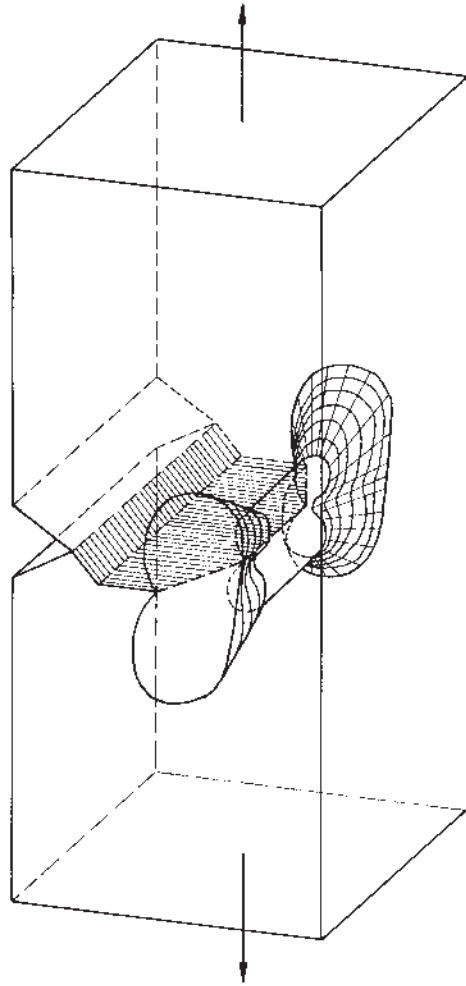
Figure 51 Plastic zone at crack tip.

is given by

$$B = 2.5\left(\frac{K_{Ic}}{\sigma_{YS}}\right)^2 \tag{124}$$

Different configurations are used to determine the plane strain fracture toughness, K_{Ic} [130]. These are shown in Fig. 53. Other specimens include the center-cracked plate [131]. The center-cracked plate is also used in determining the fatigue crack growth rate.

The notch is machined in the specimen and made sharper by fatiguing at low cycle, high strain until the crack is about the width of the test specimen. The initial crack length is measured by including the length of the fatigue crack and the notch.
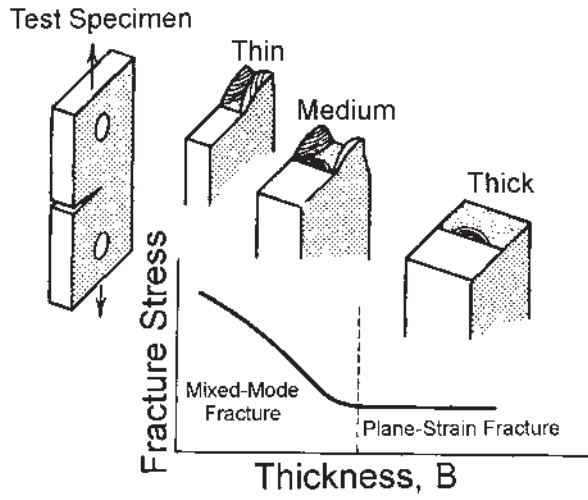
Figure 52 Thickness of specimen and resulting fracture surface.
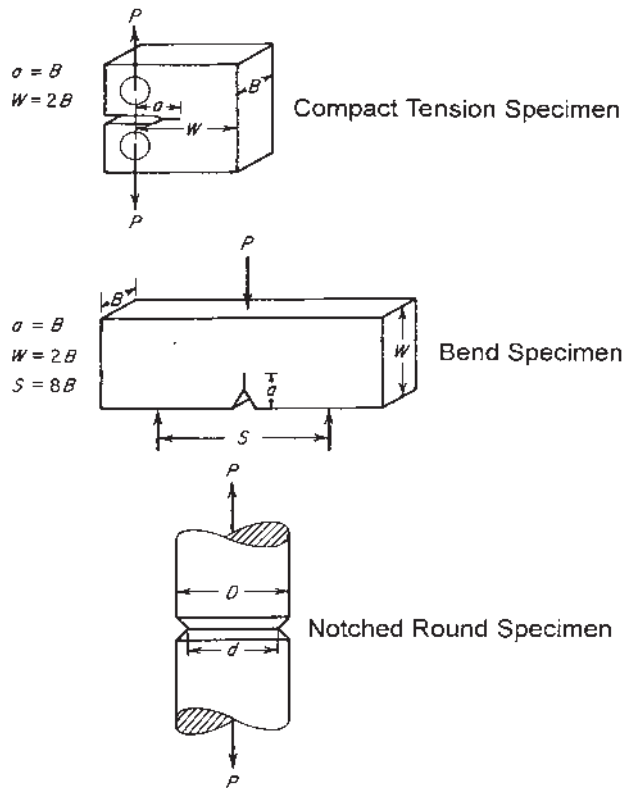
Figure 53 Typical K_{Ic} specimens (a) Compact tension specimen; (b) Bend specimen; (c) Notched round specimen (Courtesy of Dieter).
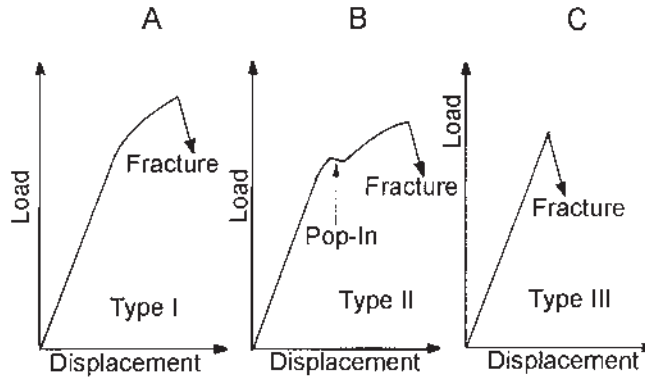
Figure 54   Typical K_{Ic} load responses to determine fracture toughness.

Testing of the specimen is accomplished by loading the specimen in tension, with the load and crack opening displacement continuously recorded until failure. In general, there are three types of load responses to the testing, shown in Fig. 54. Type I loading is characteristic of ductile materials, with no onset of unstable brittle fracture. For this type of response, the K_{Ic} value is difficult to obtain, and very thick specimens must be used before plane strain conditions occur. Type II response shows a sharp drop in the load with some load recovery. In this mode, the crack becomes unstable and propagates partially through the material until plastic tearing blunts the crack front. For this response, PQ is considered the maximum load for K_{Ic} determination. In type III loading, a maximum load is reached, with brittle crack propagation occurring rapidly. In this response, plane strain conditions exist.

A conditional value of fracture toughness, K_o is calculated using (for the compact specimen)

$$K_Q = \frac{P_Q S}{Bb}\left[2.9\left(\frac{a}{b}\right)^{1/2} - 4.6\left(\frac{a}{b}\right)^{31/2} + 21.8\left(\frac{a}{b}\right)^{5/2} - 37.6\left(\frac{a}{b}\right)^{7/2} + 38.7\left(\frac{a}{b}\right)^{9/2}\right]$$

(125)

For the three-point bending specimen,

$$K_Q \frac{P_Q}{Bb^{1/2}}\left[2.96\left(\frac{a}{b}\right)^{1/2} - 185.5\left(\frac{a}{b}\right)^{31/2} + 655.7\left(\frac{a}{b}\right)^{5/2}\right.$$
$$\left. -1017\left(\frac{a}{b}\right)^{7/2} + 638.9\left(\frac{a}{b}\right)^{9/2}\right]$$

(126)

For the notched round specimen,

$$K_Q = \frac{P}{D^{3/2}}\left[1.72\frac{D}{d} - 1.27\right]$$

(127)

Table 8 Typical K_{Ic} Values for Aluminum Plate Alloys

Alloy and Temper	K_{Ic}, ksi in
2014–T651	22
2024–T351	31
2024–T851	23
2219–T851	33
7075–T651	26
7075-T7351	30
7050–T7451	32
7475–T651	38
7475–T7351	47

To determine the fracture toughness, $K\sim$, the crack length is measured and B is calculated:

$$B = 2.5\left(\frac{K_Q}{\sigma_{YS}}\right)^{\frac{1}{2}} \tag{128}$$

If both B and a are less than the width b of the Specimen, then $K_Q = K_{Ic}$. If not, then a thicker specimen is required, and K_Q is used to determine the new thickness using Eq. (128). Typical K_{Ic} values for aluminum alloys are shown in Table 8.

10 FATIGUE TESTING

To determine the fatigue strength of a material, it is necessary to conduct fatigue testing. Even if the fatigue strength of a material is known, and a component has been designed and fabricated, it is recommended that the design be verified, if it is exposed to a fatigue environment.

There are three parts to the fatigue test. Each is a significant portion of the test. First the machine used must be appropriate to the test. Second, the specimens must be representative of the results being sought. Lastly, the results must be presented in concise fashion.

10.1 Fatigue Machines

There are many types of machines used for fatigue testing. The most common is the rotating bending machine, followed by the servo-hydraulic machine.

The rotating bending machine allows a simple means to apply a bending moment to a rotating specimen. There are a variety of different methods of loading: either one point loading, as in a cantilever; or by two or four point loading. In general, specimens used for rotating bending are tapered across the cross-section to allow for a constant stress. The stresses at the surface of the specimen vary sinusoidally between the maximum compressive and tensile stresses:

$$\sigma = \pm\frac{32M}{\pi d^3} \tag{129}$$
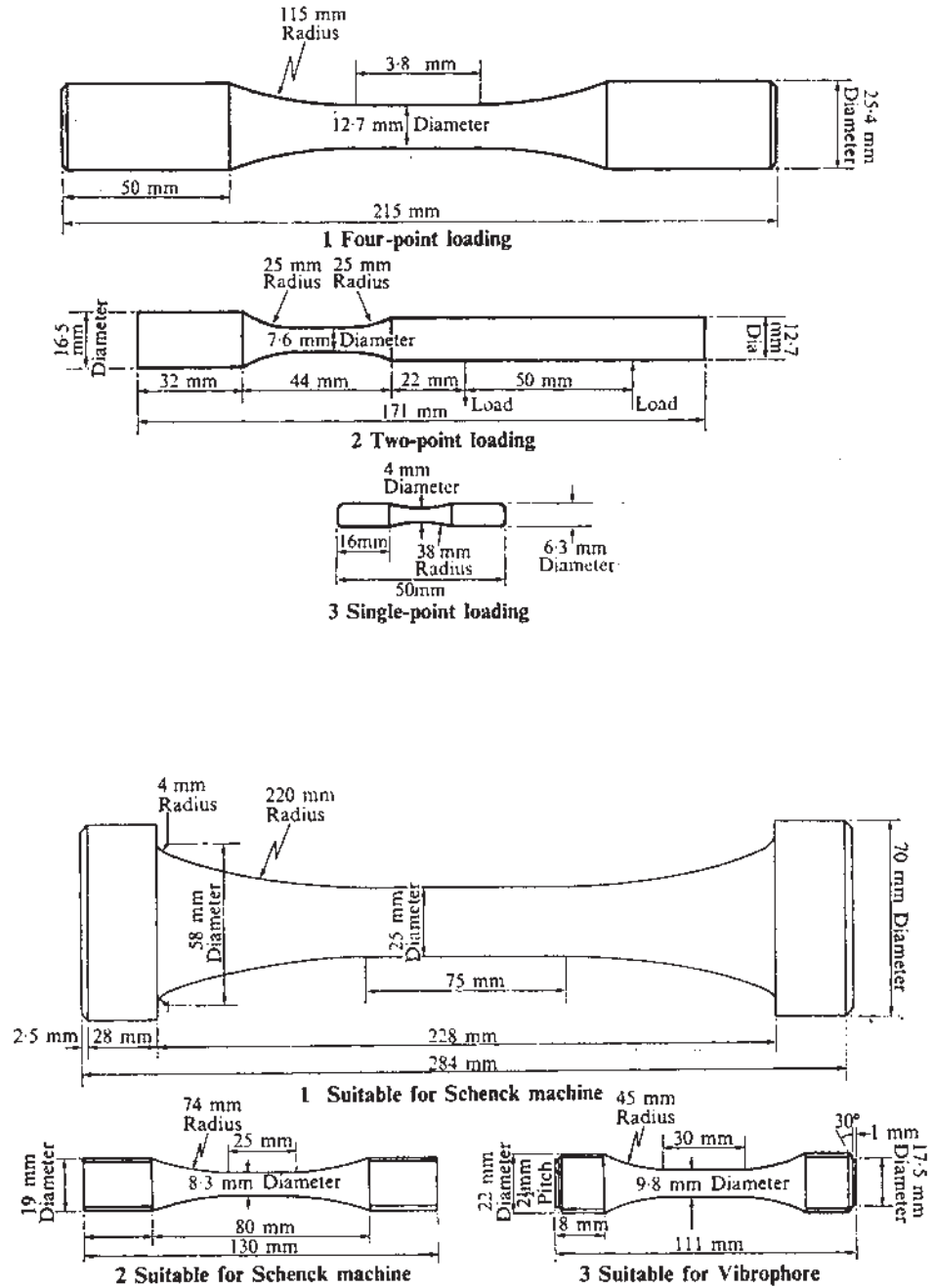
Figure 55   Typical fatigue specimens for rotating bending and axial fatigue testing.

This is equivalent to an $R = 1$. Rotating bending fatigue machines have the advantage of a simple test. It is easy to adjust the speed of the test, and it is easy to determine the applied stresses. No extensive calibration is required.

Servo-hydraulic machines (Fig. 44) have a load generated by a hydraulic cylinder, which is powered by a hydraulic pump. A load cell measures the load, and the signal is compared with the hydraulic cylinder input by using a differential amplifier. Using close loop control, very accurate load control can be achieved. A variety of inputs can be utilized: computer control, function generator, or via network control.

This type of machine is very flexible, but very complex and expensive. It is capable of large displacements, and random waveforms. It is possible to test a part to a more realistic service history than can be accomplished using a conventional rotating bending fatigue test machine.

10.2 Specimens

In general, specimens for fatigue testing have a circular cross-section. Since fatigue is generally always initiated at a free surface, it desired to insure that the surface is easily reproducible, and free from any detrimental scratches or machining marks. Typically specimens are mechanically polished, making sure that the polishing scratches are longitudinal to the axis of the part. Progressively finer grits are used. After mechanical polishing, the parts may be electropolished to minimize the residual stress effects of mechanical working. Typical fatigue specimens are shown in Fig. 55.

In certain cases, when the compressive loads are significant, it is necessary to ensure that little bending of the specimen occurs. Very gradual radii must be used to make sure that failure occurs at the reduced section, not at some geometrical stress riser. The grip sections must grasp the specimen firmly. This is to prevent fretting, with possible premature specimen failure. In addition, the grips must be axially concentric to prevent offset loading of the specimen.

In real-life, parts are not polished, and operate with undamaged surfaces, with very prescribed fatigue waveforms. Often specimens with shapes other than circular cross-sections are tested, with surfaces near the typical operation. Specimen design will often depend on whether it is supposed to be reflective of an actual part or the fatigue properties of that material.

REFERENCES

1. M. Gensamer, *Fatigue and Fracture in Metals*, Wiley, New York, 1952.
2. W. M. Garrison and N. R. Moody, J. Phy. Chem. Sol., 1987, 48, p. 1035.
3. J. F. Knott, ''Advances in Fracture Research,'' Proc. 7th International Conference on Fracture, Oxford, UK, 1989, p. 125.
4. J. F. Knott, Fracture 1977, Vol. 1, Proc. 4th International Conference on Fracture, Waterloo Canada, June, 1977, p. 61.
5. H. G. Wilsforf, Mat. Sci. Eng., 1983, 59, p. 1.
6. T. L. Anderson, *Fracture Mechanics*, CRC Press, New York, 1995.
7. R. von Mises, Z. Ver. Deut. Ing., 1914, 58, p. 750.
8. M. L. Williams, ASTM STP, 1954, 158, pp. 11–44.
9. M. E. Shank, ASTM STP, 1954, 158, pp. 45–110.

10.   J. McCall and P. French, *Metallography in Failure Analysis*, Plenum New York, 1978, p. 6.
11.   L. F. Coffin, J. Appl. Mech., 1950, 17, p. 233.
12.   R. C. Grassi, and I. Cornet, J. Appl. Mech., 1949, 16, p. 178.
13.   P. Burton, J. Appl. Mech., 1961, 28, p. 259.
14.   E. Orowan, Welding J., 1955, 34, p. 157.
15.   A. Griffith, Phil. Trans. Roy. Soc. London, 1920, 221A, p. 163.
16.   G. Dieter, *Mechanical Metallurgy*, McGraw Hill, New York, 1976, p. 253.
17.   C. E. Inglis, Trans. Inst. Nav. Arch., 1913, 55, p. 219.
18.   E. Klier, Trans. Am. Soc. Met., 1951, 43, p. 935.
19.   L. C. Chang, J. Mech. Phys. Solids, 1955, 3, p. 212.
20.   D. K. Felbeck and E. Orowan, Welding J., 1955, 34, p. 570s.
21.   E. Orowan, ed., *Fatigue and Fracture of Metals*, MIT, Wiley, New York, 1952.
22.   G. R. Irwin, "Fracture," *Encyclopedia of Physics*, Vol. VI, Springer, Heidelberg, 1958, p. 561.
23.   P. C. Paris and G. C. Sih, In: Fracture Toughness Testing, ASTM STP 381, J. E. Shawley, and W. F. Brown, eds., American Society for Testing and Materials, Philadelphia, 1965.
24.   J. W. Faupel and F. E. Fisher, *Engineering Design*, Wiley, New York, 1981.
25.   R. W. Hertzberg, *Deformation and Fracture Mechanics of Engineering Materials*, Wiley, New York, 1976.
26.   S. T. Rolfe and J. M. Barsom, *Fracture and Fatigue Control in Structures*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
27.   Anon., Standard Test Methods for Notched Bar Impact Testing of Metallic Methods, E23–88, American Society for Testing and Materials, Philadelphia, PA.
28.   Anon., Standard Test Method for Dynamic Tear Testing of Metallic Materials, E604–83, American Society for Testing and Materials, Philadelphia, PA.
29.   Anon., Standard Test Method for Plane Strain Fracture Toughness, K_{ic}, of Metallic Materials, ASTM E399–74, American Society for Testing and Materials, Philadelphia, PA.
30.   G. R. Irwin, J. A. Kies, and H. L. Smith, Proc. Am. Soc. Test Mat., 1958, 58, p. 640.
31.   J. M. Kraft, A. M. Sullivan, and R. W. Boyle, Proc. Symp. Crack Propagation, Cranfield UK, 8, 1961.
32.   W. F. Brown and J. E. Strawley, Am. Soc. Test. Mat. STP, 1965, 381, p. 133.
33.   A. S. Tetelman, and A. J. McEvily, Fracture of Structural Materials, Wiley, New York, 1967.
34.   A. H. Cottrell, Proc. Royal Soc., 1963, A276, p. 1.
35.   A. H. Knott, Mat. Sci. Eng., 1971, 7, p. 1.
36.   A. A. Wells, Proc. Crack Propagation Symp. Vol. 1, Paper 84, Cranfield, UK, 1961.
37.   G. R. Irwin, Sagamore Reseach Proc., Vol. 4, 1961.
38.   F. M. Burdekin and D. E. Stone, J. Strain Anal., 1966, 1, p. 145.
39.   BS5762: "Methods for Crack Opening Displacement (COD) Testing," British Standards Institute, 1979.
40.   E1290–93 "Standard Test Method for Crack Tip Opening Displacement Testing," ASTM, Philadelphia, 1989.
41.   J. R. Rice, J. Applied Mech., 1968, 35, p. 379.
42.   T. L. Anderson, *Fracture Mechanics*, CRC Press, Boca Raton, 1995.
43.   J. R. Rice,and G. F. Rosengren, J. Mech. Phy. Sol., 1968, 16, p. 1.
44.   J. W. Hutchinson, J. Mech. Phy. Sol., 1968, 16, p. 13.
45.   C. F. Shih, J. Mech. Phys. Sol., 1981, 29, p. 305.
46.   D. M. Tracy, Trans, ASME, J. Eng. Mat. Tech., 1976, 98, p. 91.
47.   J. R. Rice and E. P. Sorensen, J. Mech. Phy. Sol., 1978, 26, p. 163.

48. H. Kobayashi, J. Jap. Soc. Safety Eng., 1987, 26, p. 363.
49. D. Walton and E. G. Ellison, Inst. Metal Rev., 1972, 17, p. 100.
50. T. J. Crooker and E. A. Lange, Inst. Metal. Rev., 1972, 17, p. 94.
51. R. I. Stephens, "Linar elastic fracture mechanics and its application to fatigue," SAE Paper 740220, Automotive Engineering Congress, Detroit, MI, 1974.
52. D. Broek, *Elementary Engineering Fracture Mechanics*, Noordhoff, Leyden, The Netherlands, 1974.
53. W. F. Brown and J. E. Shawley, *Plane Strain Fracture Toughness Testing of High Strength Metallic Materials*, ASTM STP 410, ASTM, Philadelphia, PA, 1966.
54. A. S. Kobayashi, ed., *Experimental Techniques in Fracture Mechanics*, Vols. I and II, Soc. Experimental Stress Analysis, Westport, CT, 1975.
55. J. E. Campbell, W. E. Berry, and C. E. Feddergen, eds., *Damage Tolerance Handbook-A Compilation of Fracture and Crack Growth Data for High Strength Alloys*, MCIC-HB-01, MCIC, Battelle Memorial Institute, Columbus, OH, January, 1972.
56. T. P. Rich and D. [??] Cartwright, eds., *Case Studies in Fracture Mechanics*, AMMRC MS 77–5, Army Materials and Mechanics Research Center, Watertown, MA, June, 1977.
57. H. Tada, P. C. Paris, and G. R. Irwin, *The Stress Analysis of Cracks Handbook*, Del Research Corp., Hellertown, PA, 1973.
58. H. Liebowitz, ed., *Fracture Mechanics of Aircraft Structures*, AGARD-AG-176, NTIS, Springfield, VA, 1974.
59. J. Schijve and D. Broek, Aircraft Eng., 1962, 34, p. 314.
60. D. Kalderon, Proc. Ins. Mech. Eng., 1972, 186, p. 341.
61. B. F. Brown, Met. Rev., 1968, 129, p. 171.
62. M. V. Hyatt and M. O. Speidel, Report D6–24840, Boeing Commercial Airplane Group 1970 Seattle Washington.
63. R. Shipp, Thesis, Cambridge University, 1971.
64. H. J. Grover, S. A. Gordon, and L. R. Jackson, Fatigue of Metals and Structures, NAVWEPS Rep. 00–25–534, Bureau of Naval Weapons, Dept. of the Navy, Washington, DC, 1960.
65. K. E. Thelning, *Steel and Its Heat Treatment*, Butterworths, London, UK, 1984.
66. R. B. Heywood, *Designing against Fatigue*, Chapman and Hall, London, 1962.
67. G. H. Rowe, J. Mater., 1966, 1, p. 689.
68. J. T. Ransom and R. F. Mehl, Trans. AIME, 1949, 185, p. 364.
69. P. H. Armitage, Metal. Rev., 1964, 6, p. 335.
70. R. E. Little and E. H. Jebe, *Statistical Design of Fatigue Experiments*. Wiley, New York, 1975.
71. H. Muller-Stock, Mitt. Kohle Eisenforsch GmbH, 1938, 8, p. 83.
72. J. T. Ransom, ASTM STP, 1952, 121, p. 59.
73. W. J. Plumbridge and D. A. Ryder, Metall. Rev., 1969, 14, p. 136.
74. P. J. Forsyth and C. A. Stubbington, J. Inst. Met., 1955, 83, p. 395.
75. W. A. Wood, *Some Basic Studies of Fatigue in Metals*, Wiley, New York, 1959.
76. A. H. Cottrel and D. Hull, Proc. Roy. Soc. Lon., 1953, 242A, p. 211.
77. W. A. Wood, Bull, Inst. Met., 1955, 3, p. 5.
78. C. Laird, Fatigue crack propagation, in ASTM STP 415, American Society of Testing and Materials, Philadelphia, 1967, p. 136.
79. P. J. Forsyth and D. A. Ryder, Metallurgica, 1961, 63, p. 117.
80. G. Jacoby, *Current Aeronautical Fatigue Problems*, J. Schijve, ed., Pergamon, New York, 1965, p. 78.
81. W. R. Warke and J. M. McCall, *Fractography Using the Electron Microscope*, ASM Tech. Rep. We-2-65, American Society of Metals, Metals Park, OH, 1965.
82. G. Jacoby, Fractographic methods, Exp. Mech., 1965, p. 65.

83. P. J. Forsyth, A two stage process of fatigue crack growth, Symp. Crack Propagation, Cranfield, UK, 1961, Vol. II, p. 76.
84. J. E. Shigley, *Mechanical Engineering Design*, McGraw-Hill, New York, 1977.
85. A. R. Woodward, K. W. Gunn, and G. Forest, Int. Conf. Fatigue, Inst. Mech. Eng., 1956, p. 158.
86. D. Broek and J. Schijve, *The Influence of Sheet Thickness in the Fatigue Crack Propagation in* 2024-T3 *Alcad Sheet Material*, NLR Tech, Rep. M2129, Amsterdam, 1963.
87. P. G. Forest, *Fatigue in Metals*, Addison-Wesley, Reading, PA, 1962.
88. N. E. Frost, K. J. Marsh, and L. D. Pook, *Metal Fatigue*, Oxford University Press, London, 1974.
89. J. Y. Mann, Dept. of Supply, Australia, Aero. Res. Lab., Report SM147, 1950.
90. T. T. Oberg, and E. J. Wad, Wright Air Development Department, Tech. Note Report, 1956, pp. 56–289.
91. E. R. Gadd, J. Royal Aero. Soc., 1954, 57, p. 565.
92. R. A. Hammond, and C. Williams, Metal. Rev., 1960, 5, p. 165.
93. J. M. Finney, Metallurgica, 1958, 60, p. 273.
94. E. G. Savage, E. F. Sampson, and J. K. Curran, RAE Technical Note, Met., 1954, p. 200.
95. G. W. Stickley and F. M. Howell, Proc. ASTM, 1950, 50, p. 735.
96. G. W. Stickley and F. M. Howell, Proc. ASTM, 1960, 60, p. 577.
97. J. M. Finney, Metallurgica, 1959, 60, p. 93.
98. W. J. Harris, *Metallic Fatigue*, Pergamon Press, Oxford, 1961.
99. R. M. Brick and A. Phillips, Trans, ASM, 1941, 29, p. 435.
100. G. E. Nordmark, Spacecraft Rockets, 1964, 1, p. 125.
101. W. O. Dinsdale and R. P. Newman, Br. Welding J., 1961, 8, p. 402.
102. I. S. Sokolnikoff, *Mathematical Theory of Elasticity*, McGraw-Hill, New York, 1956.
103. R. E. Peterson, *Stress Concentration Design Factors*, Wiley, New York, 1974.
104. W. C. Young, *Roark's Formulas for Stress and Strain*, McGraw-Hill, New York, 1989.
105. R. E. Peterson, in *Metal Fatigue* (G. Sines and J. L. Waisman, eds.), McGraw-Hill, New York, 1959.
106. H. Neuber, *Theory of Notch Stresses*, J. W. Edwards, Ann Arbor, MI, 1946.
107. O. H. Basquin, ASTM Proc., 1910, 10, p. 625.
108. L. F. Coffin, Trans. ASME, 1954, 76, p. 931.
109. S. S. Manson, NACA TN 2933, 1953.
110. H. Neuber, J. Appl. Mech., Trans. ASME, E28, 1961, p. 544.
111. T. Seeger and P. Heuber. J. Test Evaluation, 1980, 8, p. 199.
112. T. H. Topper, R. M. Mwetzel, and J. Morrow, J. Mater., 1969, 4, p. 200.
113. J. E. Dorn, ed., *Mechanical Behavior of Materials at Elevated Temperature*, McGraw-Hill, New York, 1961.
114. R. W. Guard, Prod. Eng., 1956, 27(10), pp. 160–174.
115. I. Finnie and W. R. Heller, *Creep of Engineering Materials*, McGraw-Hill, New York, 1959.
116. E. N. da C. Andrade, and B. Chalmers, Proc. Roy. Soc Lond., 1932, 138A, p. 348.
117. J. B. Conway, Trans, Metall. Soc., AIME, 1965, 223, p. 2018.
118. F. Garofalo, *Properties of Crystalline Solids*, ASTM STP 283, ASTM, Philadelphia, 1965.
119. W. Rosenhahn and D. Ewen, J. Inst. Met., 1913, 10, p. 119.
120. A. Nadai, *Theory of Flow and Fracture of Solids*, Vol. I, McGraw-Hill, New York, 1950.
121. R. Hill, *The Mathematical Theory of Plasticity*, Clarendon Press, Oxford, UK, 1950.
122. M. C. Shaw and G. J. DeSalvo, J. Eng. Ind., 1970, 92, p. 469.
123. M. C. Shaw and G. J. DeSalvo, Met. Eng. Q., 1972, 121(5), p. 1.

124. S. Timoshenko and J. N. Goodier, *Theory of Elasticity*, 2nd Edn, McGraw-Hill, New York, 1972, p. 372.
125. F. C. Knoop, C. G. Peters, and W. B. Emerson, ''A sensitive pyramidal-diamond tool for indentation measurements,'' J. Res. Natl. Bur. Stand., 1959, 23, p. 49.
126. G. L. Kehl, *The Principles of Metallographic Laboratory Practice*, McGraw-Hill, New York, 1949.
127. Anon., Standard Test Method for Dynamic Tear Testing of Metallic Materials E604–83, American Society for Testing and Materials, Philadelphia, PA, 1983.
128. G. R. Irwin, ''Fracture,'' in *Encyclopedia of Physics* (S. Fluge, ed.), Springer-Verlag, Berlin, 1958, p. 551.
129. Standard Practice for R-Curve Determination, ASTM E561, ASTM Philidelphia, PA.
130. G. E. Dieter, *Mechanical Metallurgy*, McGraw-Hill, New York, 1976, p. 496.
131. K. M. Kraft, *Techniques of Metal Research*, Vol. V, Wiley, New York, 1971, Chap. p. 7.
132. J. F. Knott, *Fundamentals of Fracture Mechanics*, Wiley, New York, 1973.
133. J. H. Faupel, F. E. Fisher, *Engineering Design*, Wiley Interscience, New York, 1981, p. 873.
134. R. C. Juvinall, *Engineering Considerations of Stress, Strain and Strength*, McGraw Hill, New York, 1967.

11

Corrosion of Aluminum and Its Alloys

T. DAVID BURLEIGH

New Mexico Tech, Socorro, New Mexico, U.S.A.

1 INTRODUCTION

This chapter examines the issues concerning what is corrosion and why does aluminum corrode. Corrosion is the chemical reaction of a metal, in this case aluminum, with its environment, which leads to the deterioration of the properties of the aluminum. Aluminum is a very reactive metal, but it is also a passive metal. This contradictory nature is explainable because nascent aluminum reacts with oxygen or water and forms a coherent surface oxide which impedes further reaction of aluminum with the environment (Fig. 1). Aluminum is chemically very reactive. For example, powdered aluminum is used as rocket propellant for propulsion of the space shuttle's solid fuel rockets. Additionally, the reaction of aluminum with water releases a tremendous amount of energy:

$$Al + 3H_2O \rightarrow Al(OH)_3 + 3H_2 \uparrow \quad \Delta G = -1.1 \frac{MJ}{mol} \quad (1)$$
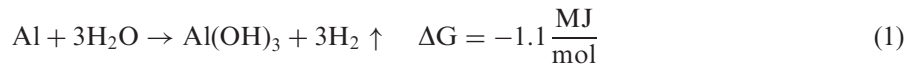
In the above equation, one pound of aluminum reacting with water can release the energy equivalent to nine pounds of dynamite. Figure 2 illustrates that aluminum metal is converted from the bauxite ore only by the input of a large amount of electrical energy and heat. In turn, aluminum metal has a similar driving force to return to its hydroxide state. Corrosion is the reaction of aluminum with water and the subsequent deterioration of its properties. Corrosion, by definition, is a slow process, requiring days or years to occur to a noticeable extent, as opposed to similar electrochemical reactions such as etching, brightening, or anodizing which occur in minutes or less.
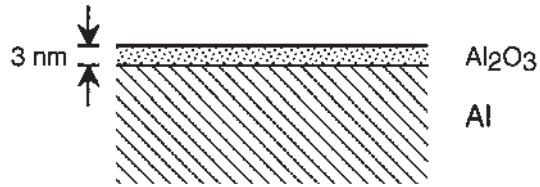
Figure 1 Aluminum is passivated in air by a thin film of aluminum oxide.
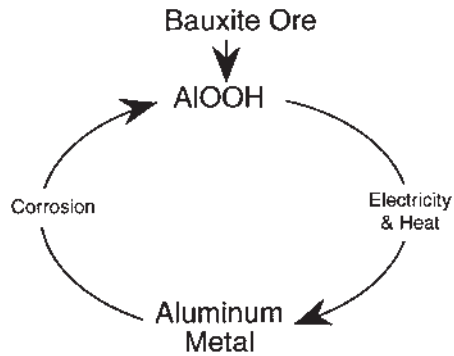
Figure 2 Refining produces aluminum metal, but corrosion returns the metal back to aluminum hydroxide.

2 DISCUSSION

2.1 The Pourbaix Diagram for Aluminum

The corrosion of aluminum requires the presence of water. The chemical equilibrium of aluminum and water is best illustrated by the Pourbaix Diagram shown in Fig. 3 [1]. The Pourbaix diagram, also known as the potential-pH diagram, shows the equilibrium phases for the aluminum-water system at different pHs and potentials. The pH (x-axis) varies from acidic at low pH to caustic at the high pH. The electrochemical potential (y-axis) is measured against the hydrogen electrode, with the negative potentials being chemically reducing, while the positive regions are chemically oxidizing. The region of water stability is bounded by the dashed lines labeled (a) and (b). Line (a) is the hydrogen line, below which water is no longer stable and decomposes into hydrogen and OH^- (alkalization). The (b) line is the oxygen line above which water decomposes into oxygen and H^+ (acidification). The world's natural environments generally exist within the regions where water is stable, between lines (a) and (b). In acidic conditions (low pH), aluminum will dissolve as Al^{+3}, and in caustic conditions (high pH) aluminum will dissolve as AlO_2^-. In neutral pHs, from 4 to 8, the hydroxide of aluminum is insoluble and will form a passive film on the aluminum's surface and protect the aluminum, as shown in Fig. 4. Between these regions of passivation and corrosion in Fig. 3 are four parallel lines labeled -6, -4, -2, and 0, which denote the chemical activity (on the log scale) of the ionic species. Chemical activity is roughly proportional to the con-
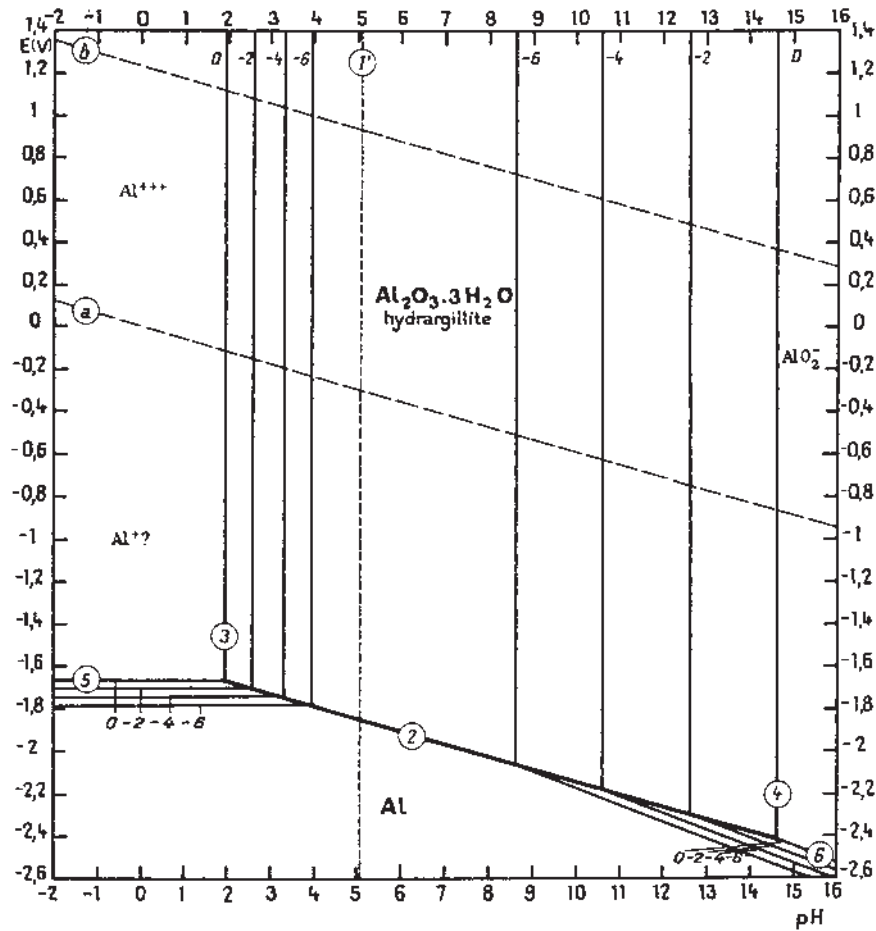
Figure 3 The Pourbaix Diagram for aluminum and water illustrates the stable phases for the different potentials and pHs. (From Ref. 1, with permission from CEBELCOR.)

centration. The 0 line denotes a unit chemical activity (10^0) which would be a saturated ionic solution in contact with the solid phase. The -6 denotes a very dilute solution (10^{-6}), roughly 1 ppm of the ionic species in contact with the solid phase.

By using the Pourbaix diagram, one can explain why aluminum corrodes (dissolves) in liquid concrete, but is stable in solid concrete (assuming no salt is present). Concrete has a pH $= 13$, which on the Pourbaix diagram corresponds to the region where AlO_2^- with an activity of 0.01 (10^{-2}) is stable with solid gibbsite (hydragillite). In liquid flowing concrete, it is difficult to maintain an activity of 0.01 of AlO_2^-, so the aluminum continuously dissolves. However, in solid concrete, solid state diffusion is very slow, and the aluminum/concrete interface can become saturated in ionic species, which are immobile, and prevent further dissolution.

The Pourbaix diagrams illustrate that in fresh water aluminum will be passive, assuming there are no halide ions such as chlorides present. (The effect of chlorides is described in later section.) The Pourbaix diagrams in Fig. 3 and 4 illustrate that
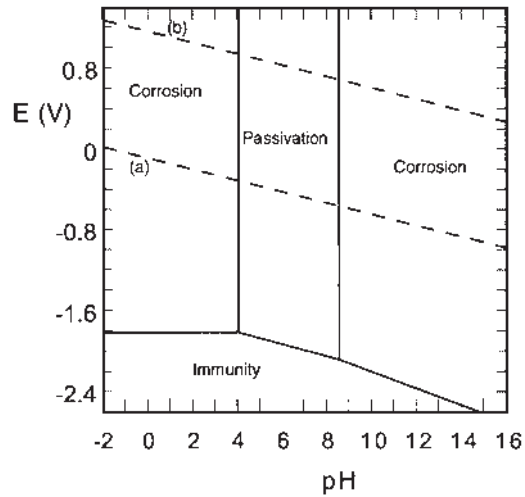
Figure 4 The Pourbaix diagram is redrawn showing the regions of corrosion, passivation and immunity, by assuming passivation by a film of gibbsite, γ-Al(OH)$_3$.
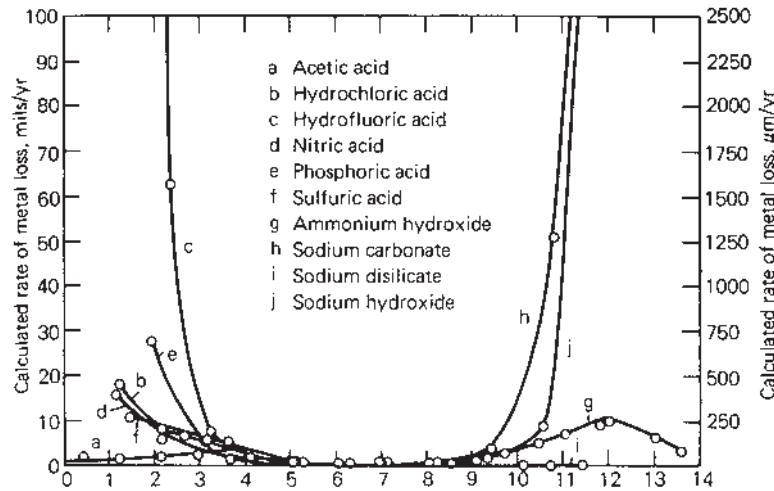
Figure 5 The effect of pH on corrosion of 1100-H14 alloy by various chemical solutions. (From Ref. 2, with permission from ASM International.)

aluminum will dissolve in both acids and bases. Figure 5 shows experimental data from Hatch [2] that aluminum dissolves rapidly both in high pH solutions (alkaline or caustic) and in low pH solutions (acids) due to the instability of the aluminum oxide and hydroxide films. (There are exceptions such as in sodium disilicate (i), where the aluminum passivates with insoluble silicate compounds. In concentrated nitric acid (a), the dissolution rate is also very slow, and the passivation mechanism is not understood.) A word of caution about using these Pourbaix diagrams: first,
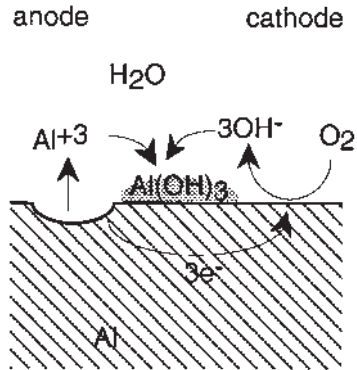
Figure 6 The dissolution-precipitation mechanism of aluminum corrosion. The aluminum dissolves as a cation at the anode. The electrons travel through the metal to the cathode where they are consumed by the cathodic reaction. The aluminum cation and the hydroxyl anion combine in the liquid and precipitate as a solid.

these diagrams show the equilibrium state, but predict nothing about the rates of reaction (fast versus slow). Second, even if the bulk solution has a neutral pH, aluminum can still corrode via localized corrosion since pits or crevices can have a different pH and chloride concentration than the bulk solution.

2.2 Electrochemistry of Corrosion

Aluminum corrosion is an electrochemical reaction, not simply a chemical reaction. Electrochemical means that the reaction depends on a transfer of electrons from one site, the anode, where aluminum is dissolving and releasing electrons, to a second site, the cathode, where the electrons are consumed (Fig. 6). The reaction for the anodic dissolution of aluminum is shown in Eq. (2)

$$Al \rightarrow Al^{+3} + 3e^- \qquad \text{(anodic reaction)} \qquad (2)$$
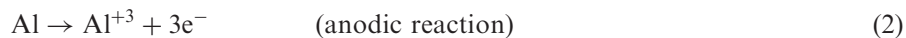
The cathodic reaction on the other hand consumes the electrons. In water there are two possible cathodic reactions. If oxygen is present, the cathodic reaction is the reduction of oxygen in which oxygen is reduced to hydroxyl anions, as shown in Eq. (3). This cathodic reaction will cause an increase in the pH at the surface:

$$\frac{1}{2}O_2 + H_2O + 2e^- \rightarrow 2OH^- \qquad \text{(cathodic reaction)} \qquad (3)$$
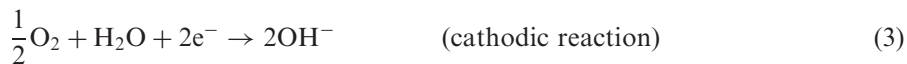
The reduction of oxygen is a fairly rapid reaction when it occurs on impurities such as Fe or Cu precipitates in the aluminum matrix. For this reason, single phase aluminum alloys will have greater corrosion resistance than aluminum alloys which contain second-phase intermetallic particles.

If there is no oxygen present, the second possible cathodic reaction in water is hydrogen evolution:

$$2H^+ + 2e^- \rightarrow H_2 \uparrow \qquad \text{(cathodic reaction)} \qquad (4)$$
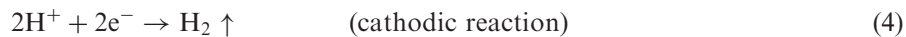
This reaction is generally slow on aluminum except in acidic solutions or unless a large negative potential is applied to the aluminum. This second cathodic reaction also increases the pH at the surface of the aluminum. When bare aluminum is immersed in water, its surface will initially become alkaline due to these cathodic reactions.

Once the aluminum cation has dissolved into the aqueous solution, it may diffuse to another area where it precipitates from the solution as a gelatinous aluminum hydroxide (Eqs. (5) or (6). Upon drying and dehydrating, the hydroxide gel forms the white crystalline corrosion product powder found on dry corroded aluminum.

$$Al^{+3} + 3OH^- \rightarrow Al(OH)_3 \downarrow \qquad \text{precipitation} \qquad (5)$$

$$Al^{+3} + 3H_2O \rightarrow 3H^+ + Al(OH)_3 \downarrow \qquad \text{precipitation} \qquad (6)$$

In both of the above reactions, the precipitation of aluminum hydroxide leads to acidification of the region. The corrosion mechanism of aluminum is known as Dissolution-Precipitation Mechanism (Fig. 6). Figure 7 from Wefers [2], illustrates the surface of aluminum after exposure to water at 62°C. The aluminum oxide is covered with a gelatinous hydroxide. On the top surface, crystals of bayerite begin to grow. Above 70°, boehmite replaces the bayerite as the stable phase [3]. Aluminum corroded in saltwater can develop a thick hydroxide gel which will form a mud crack pattern when dried, as shown in Fig. 8 [4].

2.3 Effect of Chlorides on Corrosion Rates

Chlorides (e.g. NaCl) dissolved in water, can accelerate the corrosion of aluminum. The chlorides increase the conductivity of the water, and they assist in the dissolution process. As shown above in Eq. (6), when aluminum cations precipitate, the aqueous solution becomes acidic with positively charged hydrogen cations (H^+). These positive cations attract negative anions such as hydroxyls (OH^-) and chlorides (Cl^-). The chloride anions are smaller and more mobile and can diffuse faster than the hydroxyls, thus the anodic site becomes acidic and concentrated in chlorides, that is, dilute hydrochloric acid. (Not to be confused with the cathodic site which is becoming more alkaline, Eqs. (3) and (4)). The passive film on aluminum is unstable in dilute hydrochloric acid and can not regenerate. Thus wherever the passive film is defective, the underlying aluminum base metal begins to dissolve. This reaction is auto-catalytic (self-accelerating) because the dissolution-precipitation of aluminum causes more chloride anions to be attracted to the site. The aluminum can combine with the chlorides to form $AlCl_3$ (Eq. (7)) which is very soluble (449 gm can dissolve in one liter of water at 25°C [5]), in contrast to $Al(OH)_3$ which is insoluble in neutral water (10^{-8} g dissolve in one liter of water at 25°C [5]).

$$Al + 3Cl^- \rightarrow AlCl_3 + 3e^- \qquad (7)$$

As the concentration of chlorides is increased, the corrosion mechanism can change from passivity to pitting or to anodic brightening or active dissolution, as shown in Fig. 9 from T. P. Hoar [6]. In a sense, pitting is anodic brightening on a localized scale (inside the pit). T. P. Hoar [6] proposed that chloride clusters form about the aluminum cations, and these can diffuse a short distance and then

D. R. Micholas   **10,000X**   ⊢——⊢   **A**
                                              1 μm

A: Al Immersed in Water, 1 day at 335K
B: Film Texture, Schematic
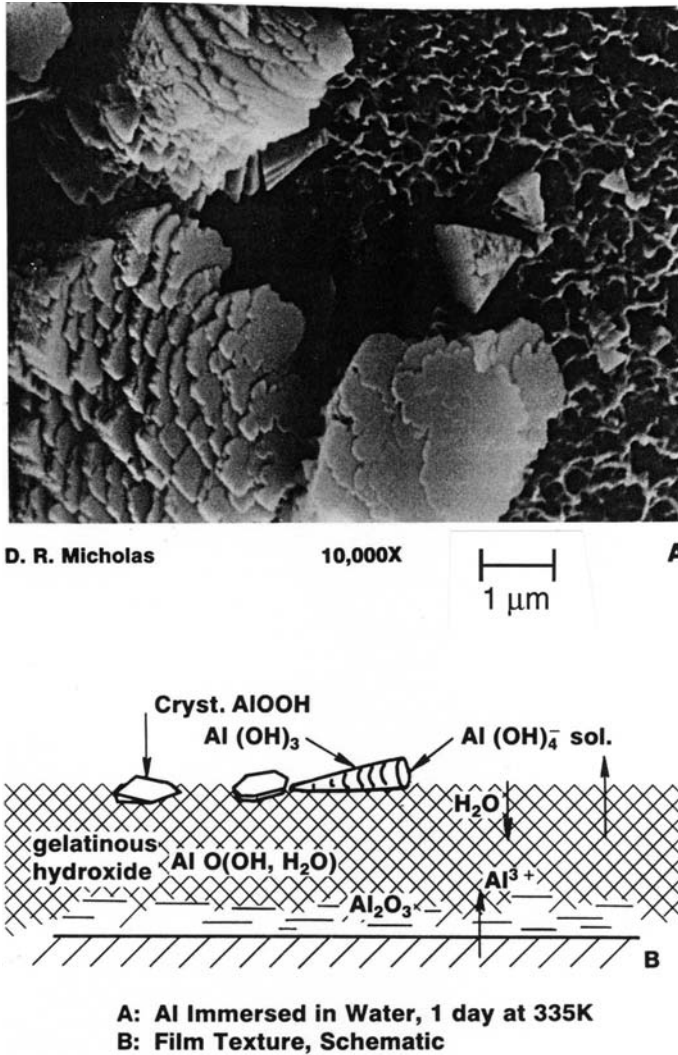
Figure 7   (A) The surface of aluminum after one day in hot water (62°C.) (B) Schematic of the multilayered hyroxides film on aluminum. (From Ref. 3, with permission from Alcoa.)

react with hydroxyls or water, and precipitate out as aluminum hydroxide, as shown in Fig. 10 and Eq. (8).

$$AlCl_3 + 3OH^- \rightarrow Al(OH)_3 + 3CL^- \tag{8}$$

Chloride anions are released from the cluster and return to the pit to continue the dissolution-precipitation reaction. As verification of this model, Wong and Alkire [7] measured the solution in the the aluminum pits and demonstrated that they contained aluminum hydroxyl-chlorides, $AlCl_x(OH)_{3-x}$. Thus an entire range of aluminum hydroxyl-chlorides are possible.

Figure 8 The surface of aluminum corroded in saltwater at room temperature has a mud crack pattern where the thick hydroxide gel dried and cracked. (From Ref. 4, with permission from ASTM.)
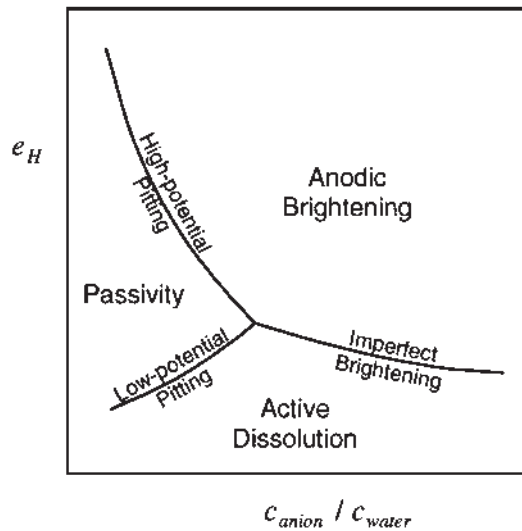
Figure 9 The concentration of anion (e.g. Cl^-) can change the corrosion mode from passivity, to pitting, to active dissolution, to anodic brightening. (From Ref. 6, with permission from The Electrochemical Society.)

3   FORMS OF ALUMINUM CORROSION AND THEIR REMEDIES

Aluminum alloys may corrode via several different pathways. Recognizing the pathway or the forms of aluminum corrosion is an important step to determine the appropriate remedy for each problem.
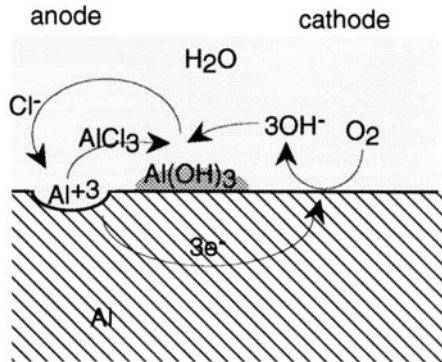
Figure 10   The chloride anions accelerate the dissolution of aluminum by chloride clusters which diffuse from the anodic site, and then react with hydroxyls and precipitate out of solution.
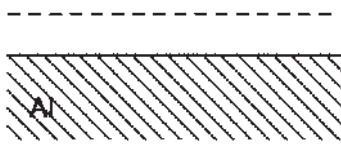
Figure 11   Uniform corrosion of aluminum.

3.1   Uniform Corrosion

General corrosion, or uniform corrosion (Fig. 11), occurs in the solutions where pH is either very high or very low, or at high potentials in electrolytes with high chloride concentrations. In acidic (low pH) or alkaline (high pH) solutions, the aluminum oxide is unstable and thus non-protective. Figure 5 demonstrated that aluminum dissolves rapidly in both high pH (e.g. caustic etching in sodium hydroxide) or low pH (e.g. dissolution in hydrochloric acid). Notably there are a few exceptions, as seen in Fig. 5. The second case of uniform dissolution, can occur in a high chloride concentration with a high applied potential (voltage). This is shown in Fig. 9 as anodic brightening.

The remedy for uniform corrosion is to change the electrolyte, use cathodic protection, add inhibitors, or replace the aluminum with a more corrosion resistant alloy. These options are discussed in detail in Section 4.

3.2   Galvanic Corrosion

Economically, galvanic corrosion creates the largest number of corrosion problems for aluminum alloys. Galvanic corrosion, also known as dissimilar metal corrosion, occurs when aluminum is electrically connected to a more noble metal, and both are in contact with the same electrolyte (Fig. 12). The Galvanic Series in Flowing Seawater from the ASM Handbook Volume 13 [8], is shown in Fig. 13. The aluminum alloys exhibit corrosion potentials between -0.6 V and -0.8 V versus the Saturated Calomel Electrode (SCE). The exact corrosion potential of an aluminum
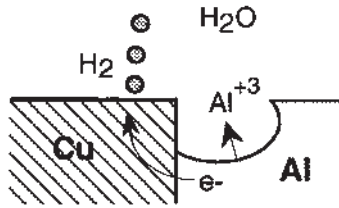
Figure 12   Galvanic corrosion of aluminum in contact with a more noble metal, e.g. copper. The aluminum is the anode and dissolves preferentially, while the copper is the preferential cathodic site.

alloy depends on its composition and heat treatment. Table 1 from Burleigh et al. [9] shows a listing of the corrosion potentials of many aluminum alloys with different heat treatments. When an aluminum alloy is connected to copper, which has a corrosion potential of -0.1 V, then there will be a 0.5 V difference between the two metals. Aluminum will dissolve preferentially, and the released electrons will flow to the copper where they shall participate in either of the two cathodic reactions, oxygen reduction, or hydrogen evolution (Fig. 12). In galvanic corrosion, the surface area of the cathode (the copper in this example) is a limiting factor. If the cathode surface area is reduced (by coating or by design), then the maximum cathodic current is correspondingly reduced.

A catastrophic form of galvanic corrosion can occur if mercury (Hg) contacts aluminum under aqueous chloride conditions. Mercury amalgamates with aluminum and in the presence of water and chlorides can produce voluminous gelatinous aluminum hydroxide. An unpublished story tells when mercury entered into a sewage treatment plant which had aluminum gates [10]. The mercury corroded the gates with a combination of amalgamation, galvanic corrosion and liquid metal embrittlement (grain boundary attack), and the ensuing cathodic reaction (hydrogen evolution) was observed by the fizzing of the aluminum gates.

The 2000 series aluminum alloys contain $> 1\%$ Cu for precipitation hardening, and these copper precipitates (Al_2Cu) can act as cathodic sites which reduce the corrosion resistance of the alloy (the rate is limited by the low surface area of the precipitates). Constituent particles such as Al-Fe can also act as the cathodic sites. Alcladding, which will be described in detail in the next section, uses galvanic corrosion to protect the copper containing alloys.

The remedy for galvanic corrosion begins on the drawing board. Good design can eliminate many galvanic corrosion problems. Ideally one should not put galvanically dissimilar metals in direct contact (see Table 1). If this can not be avoided, then an electrical insulator would prevent electrical transfer. If direct contact is required (e.g. rivets and fasteners) then water should be excluded (with sealants), the cathodic area should be coated, and the product should be designed to facilitate the water run-off.

3.3   Crevice Corrosion

Crevice corrosion requires the presence of a crevice, a salt water environment, and oxygen (Fig. 14). The crevice can result from the overlap of two parts, or the

Volts versus saturated calomel reference electrode

Figure 13   The galvanic series for metals immersed in flowing seawater. (From Ref. 8, with permission from ASM International.)

Table 1   The Corrosion Potentials of Aluminum Alloys and Metals According to the ASTM G69 Test Method (NaCl and Hydrogen Peroxide Solution) from Burleigh [9].

Lower Limit	E_{corr} (V_{SCE})	Upper Limit	Metal or Alloy
	+ 0.23	+ 0.32	Cr (99.9%)
	+ 0.12		Ni 270 (1 sample)
− 0.70		+ 0.27	Stainless Steel 316
− 0.11	+ 0.00	+ 0.08	Cu (99.999%)
− 0.08		− 0.02	Bronze (Cu94-Sn6)
− 0.05		+ 0.27	Ti-6Al-4V
− 0.30		− 0.08	Brass (Cu63-Zn37)
− 0.40		− 0.19	Sn (99.99%)
− 0.20		+ 0.02	Zr (99.9%)
− 0.80		− 0.40	Mn
− 0.43		− 0.25	Monel 400 (Ni65-Cu33-Fe2)
− 0.31	− 0.30	− 0.25	Stainless Steel 321
− 0.60		− 0.49	Fe
− 0.56	− 0.55	− 0.54	*2219-T3, T4**
− 0.62	− 0.60	− 0.59	*2014-T4, 2017-T4, 2024-T3, T4*
	− 0.62		2324-T39
	− 0.64	− 0.63	2036-T3, T4
	− 0.65		2036-T6, 2090-T3, T4
	− 0.66	+ 0.14	Ti (99.7%)
− 0.76	− 0.67	− 0.66	2091-T3, T8
	− 0.69		*2014-T6*, 2008-T4
	− 0.70		8090-T3, 6010-T4
	− 0.71	− 0.45	Pb
− 0.74	− 0.71	− 0.70	2008-T6
− 0.73	− 0.71	− 0.70	*2219-T6, T8*
− 0.73	− 0.71		*2024-T8*
	− 0.71		*6061-T4*, 6009-T4
− 0.74	− 0.73		6013-T6, T8
− 0.76	− 0.74	− 0.72	*7075-T6, 7178-T6*
	− 0.74		*3003, 1100, 6061-T6, 6053, 6063*
	− 0.74		5030-T4, 2090-T8
− 0.76	− 0.75	− 0.73	Al (99.999%)
− 0.77	− 0.75	− 0.73	*7075-T7, 8090-T7, 7049-T7, 7050-T7, 7475-T7*
	− 0.75	− 0.74	7055-T77
	− 0.75		*3004, 1060, 5050*
	− 0.76		*5052, 5086, Alclad 2024, Alclad 2014*
	− 0.77		*5154, 5454, 5254, 5042*
	− 0.78		*5056, 7079-T6, 5456, 5083*, 5182
− 0.87	− 0.84		*7039-T6, T63, 7005*
	− 0.87		*7002, Alclad 3003, Alclad 6061, Alclad 7075*
	− 0.94		7003
− 1.01	− 0.98		Zn
	− 1.64		*Mg*

* Potential for alloys or metals in italics were calculated from Binger, et al.,[2] by adding + 92 mV[5.]
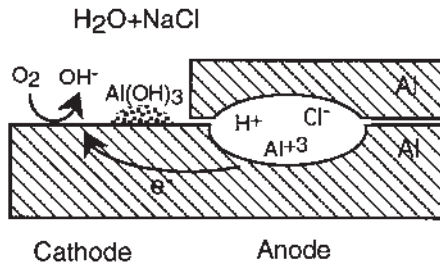
Figure 14 Crevice corrosion can occur in a saltwater environment if the crevice becomes deaerated, and the oxygen reduction reaction occurs outside of the crevice mouth. Under these conditions, the crevice becomes more acidic, and corrosion occurs at an increasing rate.

gap between a bolt and a structure. When aluminum is wetted with the saltwater and water enters the crevice, little happens initially. Over time, inside the crevice oxygen is consumed due to the dissolution and precipitation of aluminum (Eqs. (2), (3), and (5)). If the crevice is narrow and the inward diffusion of oxygen is restricted, oxygen becomes depleted, and the crevice becomes acidic via the precipitation reaction in Eq. (6). This acidic environment contains H^+ cations which attract anions in order to maintain the electrical neutrality of the crevice environment. The hydroxyl anions, OH^-, diffuse slower and loose the crevice race to the faster Cl^- anions. The crevice, which was originally neutral gradually becomes acidic due to the formation of a dilute hydrochloric acid environment. The anodic dissolution of aluminum accelerates inside this acidic crevice, and the cathodic reaction, oxygen reduction, occurs outside of the crevice mouth. Crevice corrosion is often called "autocatalytic" since it is self-accelerating.

The crevice in the above description is acidic. Crevices can be instead slightly alkaline if the outside of the crevice has a good coating or is always dry. If the oxygen reduction reaction (Eq. (3)) can not occur external to the crevice mouth, then the cathodic reaction must occur inside the crevice. In the absence of oxygen, the cathodic reaction would be hydrogen evolution, Eq. (4) in an acid, or Eq. (9) in a neutral solution. This reaction would push the crevice alkaline, and corrosion would occur at a slower rate than the acidic crevice with the external cathode.

$$2H_2O + 2e^- \rightarrow H_2 + 2OH^- \uparrow \quad \text{(cathodic reaction)} \tag{9}$$

The best way to prevent crevice corrosion is also on the drawing board. The designer should eliminate crevices and gaps by continuous welding, or sealing the crevices with sealant, or allowing drainage. The outside should be coated to minimize the area available for the cathodic reaction. Maintenance by waxing or repainting minimizes the external cathodes, and seals the crevices which would prevent the occluded cells from initiating and growing.

3.4 Pitting Corrosion

Pitting corrosion is very similar to crevice corrosion. Pitting of aluminum alloys occurs if the electrolyte contains a low level of chloride anions, and if the alloy is at a potential above the "pitting potential." Much research has been conducted
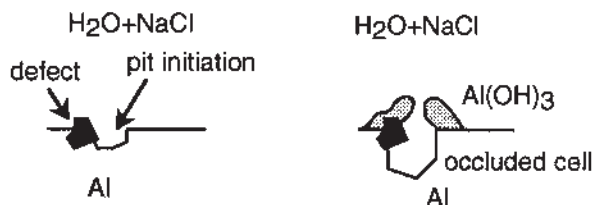
Figure 15   The pit initiates at defects, and growth occurs if an occluded cell forms.

on the pitting corrosion of aluminum (e.g. Schmutz and Frankel [11]). The pitting potential of unalloyed aluminum in a NaCl-H_2O_2 solution is approximately -0.76 versus SCE [9]. Figure 9 from Hoar [6] (although highly generalized for all metals), illustrates that pitting occurs when the chloride anion concentration is too low for anodic brightening, but too high for complete passivity. Pitting initiates at defects on the surface of the aluminum, such as at second phase particles or on grain boundaries. (If a high potential is applied to aluminum, then the pitting is reported to occur randomly across the surface.) The potential can be affected by having oxygen present for oxygen reduction (Eq. (3)) or galvanic contact with a more noble metal, or via an applied voltage.

Embryo pits, which form and repassivate, occur rapidly over the surface of aluminum. This breakdown and repassivation produces an electrical noise which can be measured to give an estimate of the corrosion rate of the aluminum. Some of these embryo pits reach a critical size in which repassivation does not occur, and the pit continues to grow. This growth occurs because the pit becomes an occluded cell by the precipitation of its own corrosion product outside the mouth of the pit (shown schematically in Fig. 15). This occluded cell is similar to the crevices mentioned previously and is also acidic and high in chlorides. The dilute hydrochloric nature of the pit leads to continued dissolution. As the pit acidifies, hydrogen evolution becomes favorable as the cathodic reaction inside the pit, and the escaping hydrogen bubbles effectively pump out the $AlCl_3$, which reacts outside the pit with water and precipitates the $Al(OH)_3$. This action forms white corrosion chimneys which can be found on the surface of pitted aluminum as shown in Fig. 16 from [12]. For a detail description of the pitting process, a ten-step mechanism for the pitting of aluminum has been described by Reboul et al. [13].

Although an individual pit may stop growing for a variety of reasons, some pits continue growing deeper over time. Figure 17 (data from Aylor [14]) shows the pitting depth for samples of alloy 6061-T6 immersed in seawater at three different worldwide locations. The pit depth continually increases, although at a decreasing rate over the ten year exposure period.

The best way to prevent pitting is to prevent the occluded cells from initiating and growing. This prevention can be achieved through regular maintenance of the surface, such as regular washing and polishing to remove dust particles, and waxing or repainting to seal occluded cells or defects. Bare aluminum which is regularly maintained by washing and buffing appears to last indefinitely in a semi-industrial exposure [15].
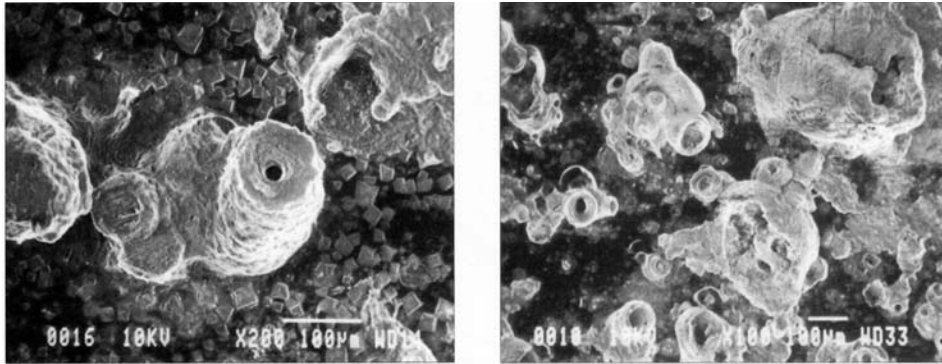
Figure 16 These SEM micrographs show the corrosion chimneys which form over corroding pits during immersion in a NaCl-H_2O_2 solution. Cubic salt crystals are present on the surface. SEM photographs by Bob Petri. (From Ref. 12, with permission from NACE International.
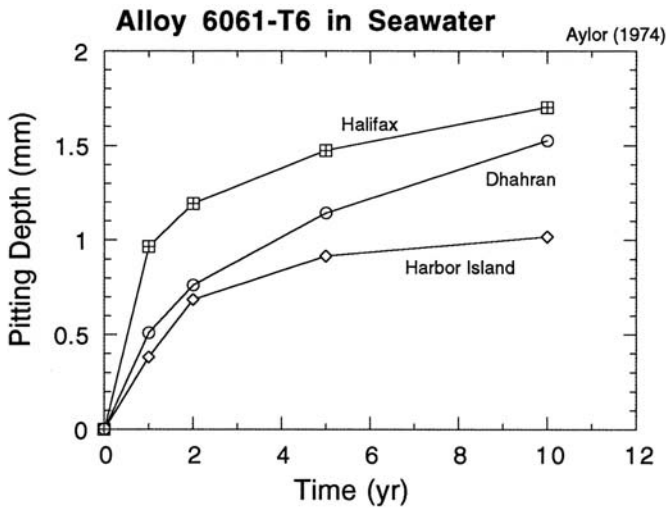
Figure 17 The pits continue growing in depth over the ten-year period of immersion in seawater. (From Ref. 14.)

3.5 Intergranular Corrosion

Intergranular corrosion is a special form of corrosion characterized by the preferential attack of the grain boundaries. Intergranular (IG) corrosion is also referred to as intergranular attack (IGA). IG corrosion only occurs if the grain boundary regions are compositionally different from the bulk of the alloy. This compositional difference occurs during heat treating, aging, or welding by diffusion of atoms and precipitation of second phase particles.

Intergranular corrosion initiates often at the sites of crevice corrosion or pitting corrosion. The acidic pit environment leads to the preferential attack of the grain boundary region, as illustrated in the schematic in Fig. 18. The two different mechanisms by which IG corrosion may occur are illustrated in Fig. 19. The first mechanism is the precipitation of a more active phase, for example Al-Mg at the grain boundaries in 5000 series alloys. On exposure to saltwater, this phase dissolves preferentially (Fig. 19(a)), leaving the bulk grains separated from each other. The second mechanism of IG corrosion occurs if a more noble phase precipitates at the grain boundaries, for example $CuAl_2$ in the 2000 or the 6000 alloys which contain copper. In this case, the areas adjacent to the grain boundaries are depleted of the copper, which makes them more reactive, and allows their dissolution in the acidic pit environment, leaving behind the grain boundary precipitates and the bulk matrix grains (Fig. 19(b)). This second mechanism is seen in the TEM
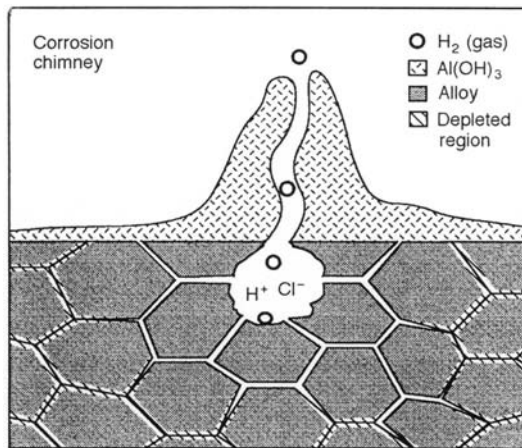
Figure 18   The pitting corrosion forms an acidic environment which leads to preferential dissolution of the grain boundary regions. (From Ref. 12b, with permission from NACE International.)
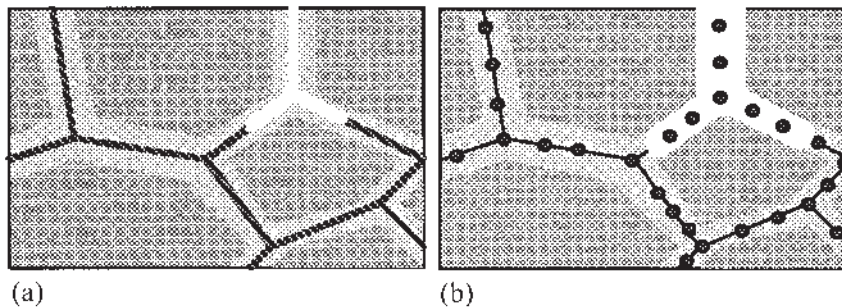
Figure 19   Intergranular corrosion can occur by either (a) the dissolution of a more active grain boundary precipitate; or (b) the dissolution of a depleted matrix around a more noble grain boundary precipitate.
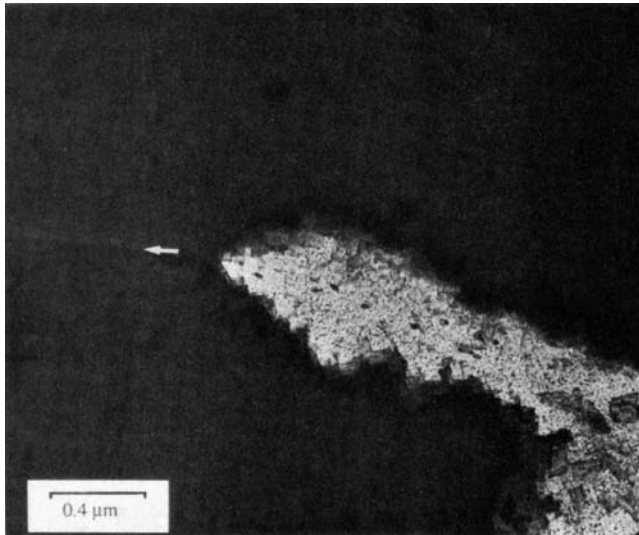
Figure 20 TEM micrograph showing the preferential dissolution of the grain boundary region, leaving the Al-Cu precipitates suspended in the corrosion product. TEM photographs by Ray Kilmer. (From Ref. 12, with permission from NACE International.)

micrograph in Fig. 20 where the TEM foil was etched in a dilute HCl solution [12]. The grain boundary (white arrow) contains Al-Cu precipitates, which remain after the depleted grain boundary region is dissolved. Figure 21 shows different types of intergranular corrosion for different aluminum alloys [2].

Intergranular corrosion can be prevented by using a single phase alloy, however, this is not always possible. Proper heat treatments can minimize the differences in potential between the bulk alloy and its grain boundary. For example in the 6000 alloys, the T6 temper often shows better IG resistance than the T4 temper. If a 2000 alloy is solutionized and not quenched rapidly, then precipitates may form on the grain boundaries and lead to IGA. Adhesive bonding is preferable over welding since the welding causes uncontrolled heat treatments and precipitation of second phases in the heat affected zone (HAZ). Eliminating the sites of pitting or crevice corrosion will also stop the initiation of IG corrosion.

3.6 Exfoliation Corrosion

Exfoliation corrosion is a special form of intergranular corrosion which occurs when the grains are flattened by heavy deformation during hot or cold rolling, and where no recrystallization has occurred. Exfoliation corrosion has the appearance of leaves of a book. Figure 22 (courtesy of Alcoa [16]) shows a plate of alloy 7075-T6 after six years of seacoast exposure. The 0.75 in thick plate was machined to the mid-plane leaving a plate of 0.375 in thickness. This plate was exposed at the seacoast in the bare condition, and the left side swelled to approximately four times the original thickness. Exfoliation is characteristic for the 2000 (Al-Cu), 5000 (Al-Mg), and 7000 (Al-Zn) series alloys which have grain boundary precipitation or depleted grain boundary regions.
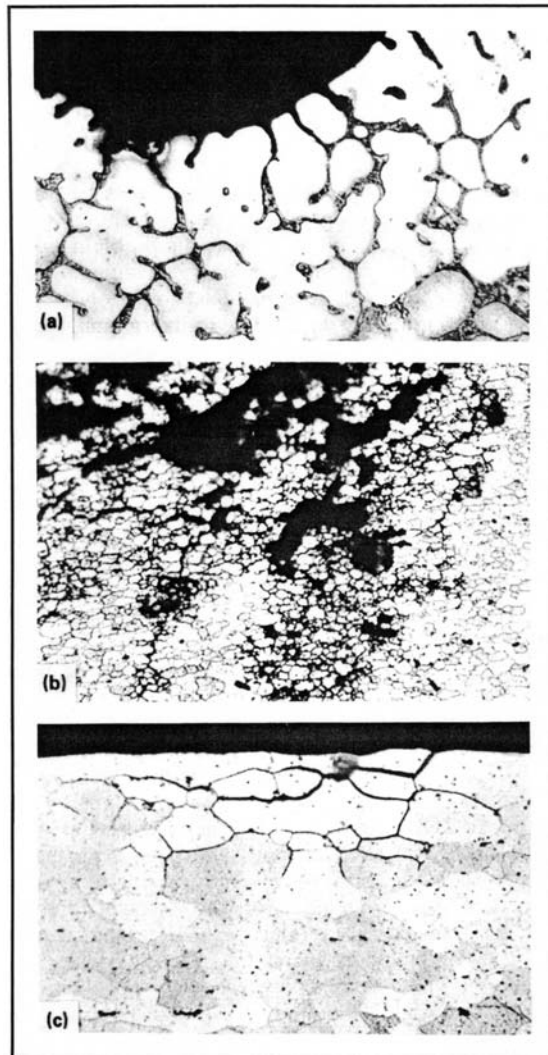
Figure 21   Various types of intergranular corrosion: (a) cast structure; (b) wrought unrecrystallized structure; and (c) recrystallized structure. (From Ref. 2, with permission from ASM International.)

The remedy for exfoliation is similar to above for IG corrosion. To prevent the exfoliation of alloy 7075-T6, the newer alloy 7150-T77 can be substituted wherever 7075-T6 is used.

3.7   Erosion-Corrosion

Erosion-corrosion of aluminum occurs in high velocity water and is similar to jet-impingement corrosion. Figure 23 shows a water-cooled aluminum heat sink used for cooling electronic equipment. In this case, the closed-loop cooling water became
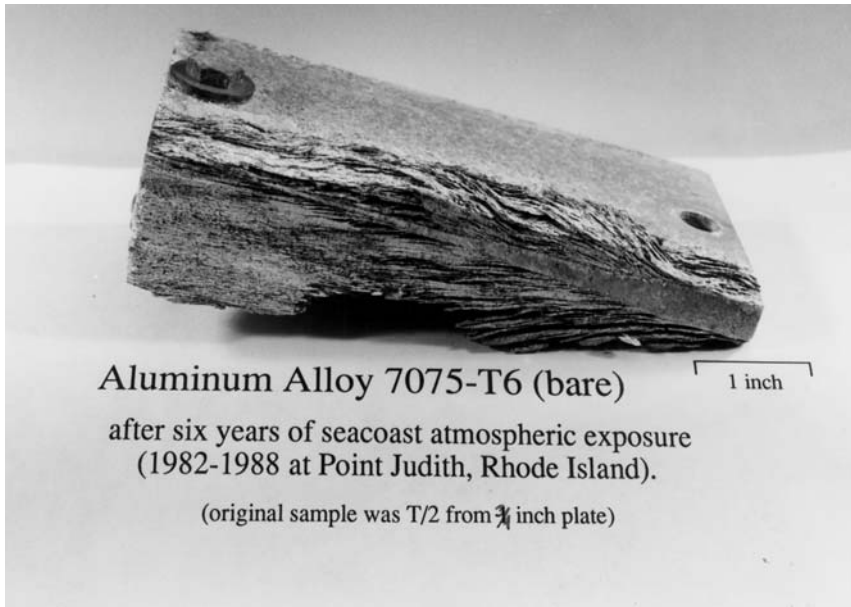
Figure 22 Exfoliation of alloy 7075-T6 after six years seacoast exposure in the bare condition. (From Ref. 16, used by the permission of Alcoa.)
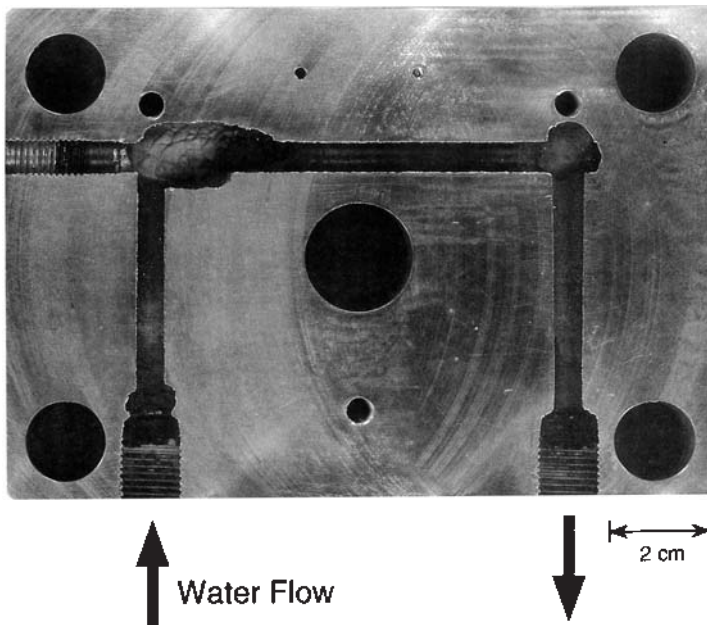
Figure 23 Erosion-corrosion of a water-cooled, alloy 6061-T6 heat sink used for cooling electronic equipment. The pH of the cooling water was > 9 which led to the dissolution of the aluminum.

alkaline with a pH > 9, and severe erosion-corrosion occured at the right-angle bends in the heat sink. Erosion-corrosion of aluminum is very slow in pure water, but is accelerated at pH > 9, especially with high carbonate and high silica content of the water [16b,16c].

Aluminum is very stable is neutral water, however it will corrode in either acidic or alkaline waters as shown in the Pourbaix [1] diagram in Fig. 3 and 4. Notice that at pH = 8.6, aluminum has an activity of 10^{-6}, or approximately a solubility of 1 ppm in water as AlO_2^-. With high velocity water which contains no AlO_2^-, the aluminum will be continually dissolving, trying to reach the activity of 10^{-6}. In addition, if the surface is abraded, the passive film will not be able to reform at pH > 9.

Jones [16b] examined the erosion-corrosion of aluminum alloy 8001 in high temperature, high pressure water. In distilled water (pH = 8.3), the erosion-corrosion was negligible (5 μm depth) after 24 hr. However, in distilled water with 50 ppm $CaCO_3$, 10 ppm SiO_2, and pH = 8.3, the erosion corrosion depth was 689 μm after 24 hrs. The carbonate and the silica interfered with the stable passive film growth on aluminum. Li [16c] studied the effect of different solutions on the jet-impingement corrosion of aluminum alloy 1100. Jet impingement of tap water at pH = 7.1 showed no detectable corrosion after an hour at 12 ft/sec impact velocity. However, a 0.1 M Na_2CO_3 solution at pH = 11.7 showed a weight loss of 2.2 mg/hr under the same jet impingement conditions.

To prevent erosion-corrosion, one may change the water chemistry or reduce the velocity of the water, or both. For the water chemistry, the pH must be below 9, and the carbonate and the silica levels must be reduced. If the pH can not be changed, then another possibility is to saturate the closed-loop water with AlO_2^- to reduce the corrosion rate.

3.8   Stress Corrosion Cracking (SCC)

Stress corrosion cracking (SCC) is the bane of aluminum alloys. Sometimes very little corrosion is evident on the outer surface, but it is possible for a critical length stress corrosion crack to be extending deep into the structure, causing a rapid and unexpected failure. Figure 24 [16] shows stress corrosion cracks in sheet 2091-T3 after 8 days exposure at 28 ksi stress in the ASTM G44 alternate immersion test [17]. In SCC, the external corrosion is minimal, but the cracks are significant. SCC requires three simultaneous conditions, first a susceptible alloy, second a humid or water environment, and third a tensile stress which will open the crack and enable crack propagation. (A compressive stress will close the crack and prevent SCC.) SCC can occur in two modes, intergranular stress corrosion cracking (IGSCC) which is the more common form, or transgranular SCC (TGSCC). In IGSCC, the crack follows the grain boundaries (Fig. 22 and 24). In transgranular stress corrosion cracking (TGSCC), the cracks cut through the grains and are oblivious to the grain boundaries (see Fig. 25 from Burleigh [18]). Much research has been conducted on the subject of SCC of aluminum alloys. Review articles on the SCC of aluminum alloys have been written by Sprowls [19], Holroyd [20] and Burleigh [21].

Figure 26 [22] shows the evolution of the alloys used for the upper wing skin plate for different aircraft over the years. A higher yield strength alloy provides more strength for less weight. The general trend to use higher strength alloys
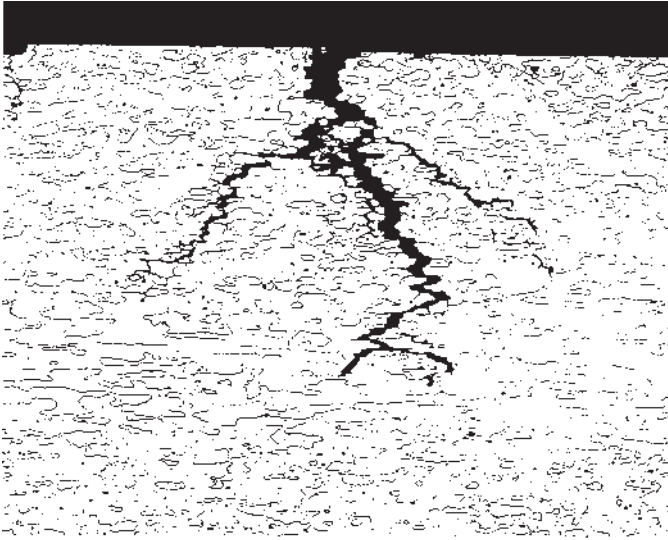
Figure 24 Stress corrosion cracks in a 2091-T3 sheet which was exposed for 8 days to the alternate immersion test (ASTM G44) at 28 ksi. (From Ref. 16, with permission from Alcoa.)
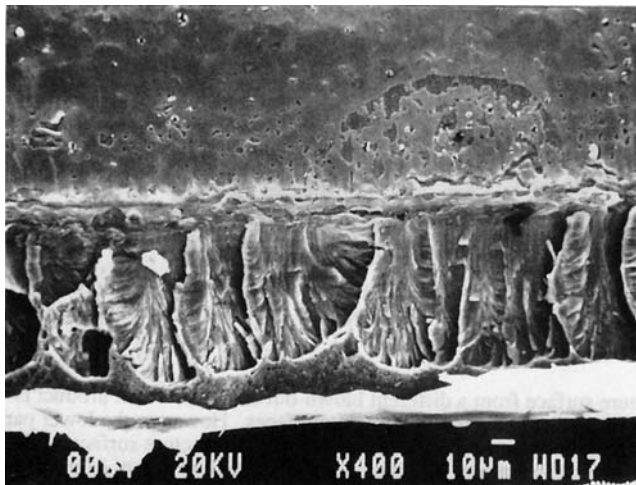
Figure 25 Transgranular SCC of a 5182 soda can end. The blow-out occurred under hot humid storage conditions. (From Ref. 18, with permission from TMS.)

peaked in 1950 with alloy 7178-T651 used on the Boeing 707, then the industry changed to using lower strength alloys. The yield strength of the upper wing skin did not exceed the 1950 level until the Boeing 777 in the 1990s. The reason lower strength alloys were selected for the Boeing 747 and the L-1011 was that the aircraft designers chose an alloy with better SCC resistance rather than the higher yield
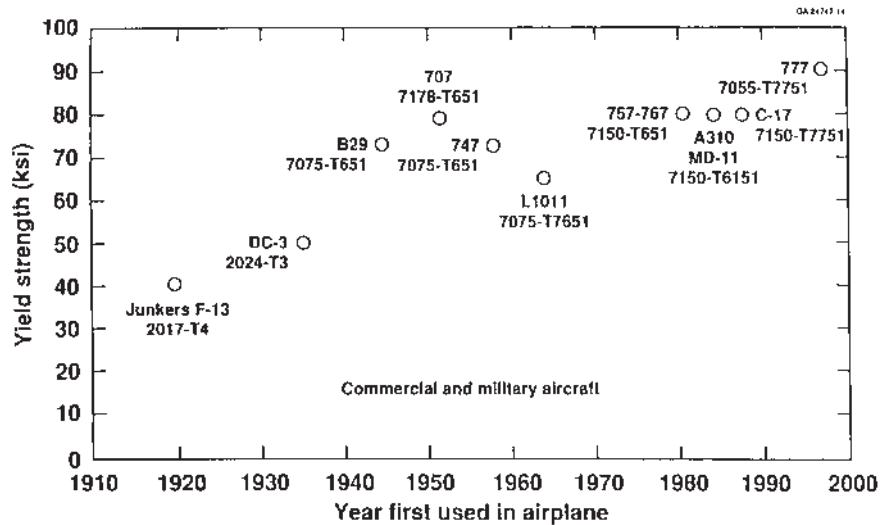
Figure 26 The aerospace industry initially chose stronger alloys for the upper wing skin plate until the 1950s when they reversed and went to a lower strength alloy with better SCC resistance. The 1950 strength levels were not exceeded until the 1990s. (From Ref. 22, with permission from SAWE.)

strength. The pre-1950 alloys exhibited high yield strength but possessed poor SCC resistance.

The aerospace alloy 7079-T651 was designed in the 1950s to be SCC resistant forging alloy, but it proved instead to have exceptionally poor SCC resistance and has since been abandoned. The aerospace industry was initially misled by alloy 7079 because the development corrosion tests were conducted on smooth compact tension specimens tested for weeks or months in an alternate saltwater immersion test (ASTM G44 [18]). The stress corrosion crack took a long time to initiate on the smooth surface on alloy 7079-T6 compared to alloy 7075-T6 in the same test environment. Once a crack did initiate in 7079-T6, it grew very rapidly (Fig. 27 from [23]) as compared to other aluminum alloys. When the 7079 forgings were put into service, they served for several years, and then suddenly cracked. As a consequence, new alloy development today relies on a multitude of corrosion tests rather than one test technique.

Notable in Fig. 28 [20] is that the SCC cracks will propagate in distilled water, although a hundred times slower than in saturated NaCl solution. Figure 29 illustrates that as the relative humidity increases from 1% to 100%, the crack velocity for 7075-T6 also increases 100 times. SCC of the 7000 series alloys is considered by many researchers to be a hydrogen-induced cracking [20,21]. For most aluminum alloys, the crack propagates along the grain boundaries, thus it is named intergranular stress corrosion cracking (IGSCC). For alloy 2024-T3, the cracking mechanism is considered to be due to anodic dissolution of the grain boundary regions as opposed to the hydrogen-induced cracking [21]. The IGSCC crack growth occurs perpendicular to direction of the applied stress because the applied stress opens the properly oriented grain boundaries. SCC cracks predominantly grow
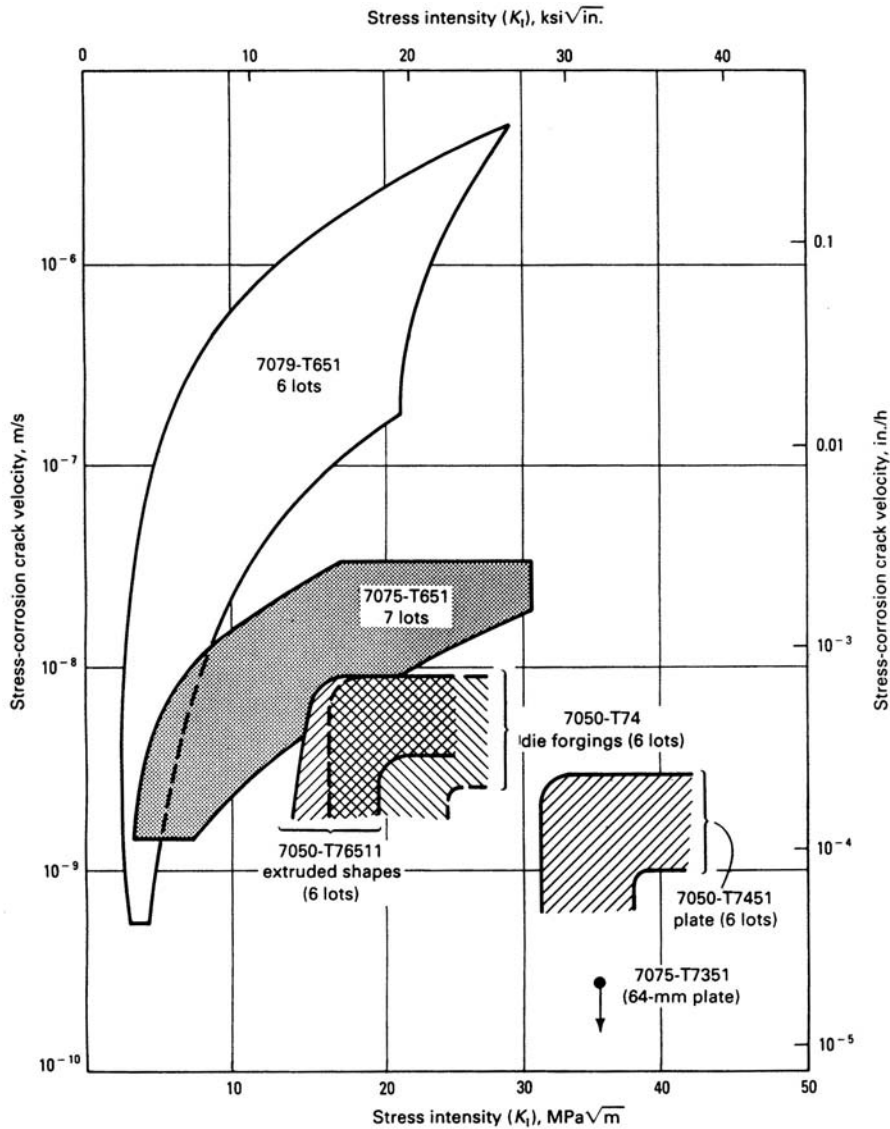
Figure 27   The SCC crack velocity is very high in alloy 7079-T651 in 3.5% NaCl solutions. (From Ref. 20b, used by the permission of ASM International.)

along high angle grain boundaries. The stresses in the short transverse direction (Fig. 30) have much lower SCC threshold levels than stresses in the other directions (Fig. 31).

Transgranular stress corrosion cracking (TGSCC) has been observed occasionally on aluminum alloys. It was first reported on alloy 7050-T73651 (Lifka [24]) and has since been documented on alloy 5182 soda can ends which blew out during storage in the high humidity summer months (Burleigh [18]). TGSCC is character-
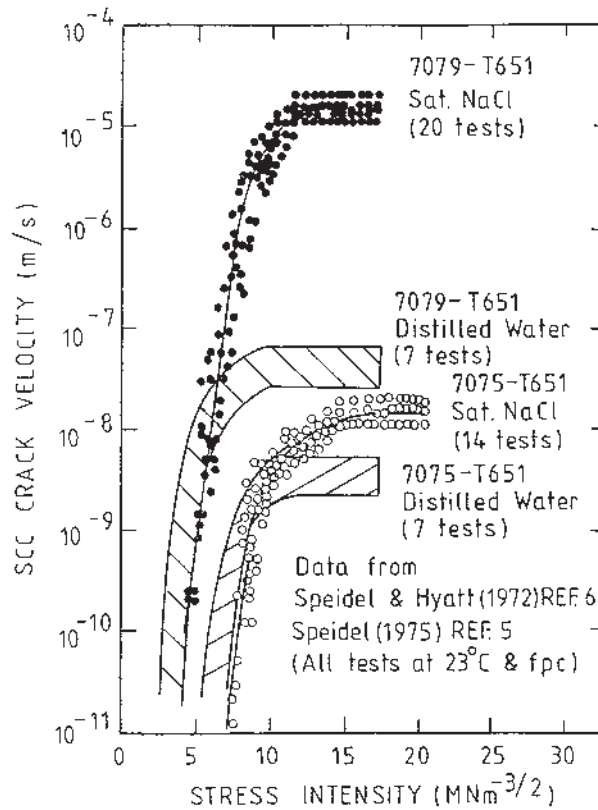
Figure 28   The stress corrosion crack will propagate in distilled water as well as in NaCl solutions. (From Ref. 20, with permission from NACE International.)

istic by having a fan-shaped fracture pattern (Fig. 25) sometimes with fatigue-like striations. TGSCC is believed to be related to hydrogen embrittlement.

For prevention of SCC, one can change the alloys or the temper (use 7150-T77 instead of 7075-T6), reduce the stress level, remove the aqueous environment, redesign the part, or use better coatings.

3.9 Corrosion Fatigue

Corrosion fatigue can occur when an aluminum structure is repeatedly stressed at low stress levels in a corrosive environment. A fatigue crack can initiate and propagate under the influence of the crack-opening stress and the environment. The characteristic fatigue striations which can be found on the fracture surfaces are shown in Fig. 32. This 2024-T3 sample was fatigued in dry air [25]. Similar striations may sometimes be found on corrosion fatigued samples, but often the subsequent crevice corrosion in the narrow fatigue crack dissolves them. Figure 33 illustrates that the crack growth per cycle depends on the stress concentration, delta K, and the environment. The cracks in alloy 7075 and 2024 grow fastest in NaCl solution, and slowest in a vacuum [26]. Figure 34 demonstrates that for a million
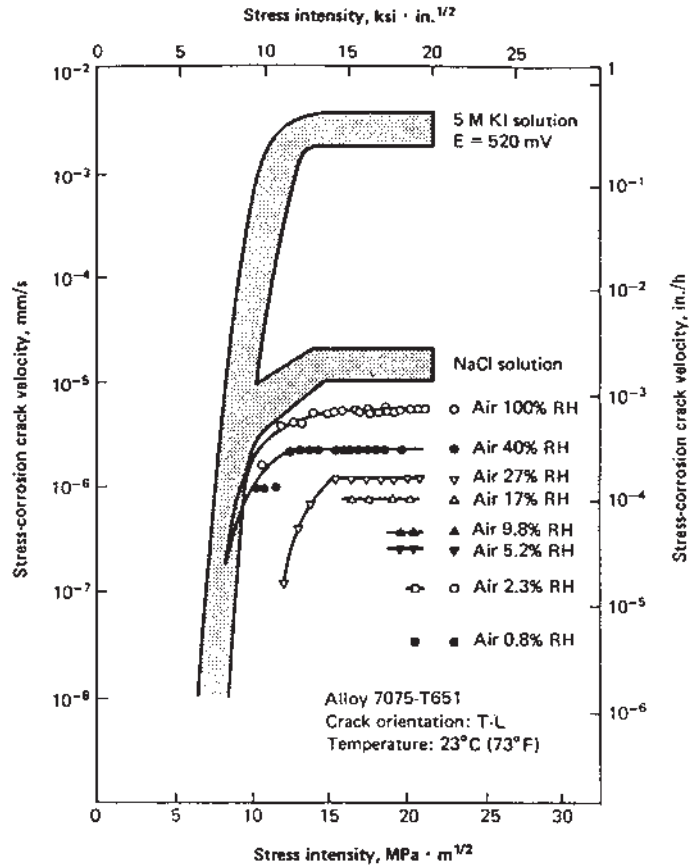
Figure 29 The crack growth velocity for 7075-T6 increases with increasing humidity. (From Ref. 20b, with permission from ASM International.)

cycles, high purity 7075-T6 alloy can withstand a stress of 270 MPa in a vacuum or in dry nitrogen, but can withstand only 180 MPa in humid air or humid nitrogen [27]. The presence of water vapor decreases the fatigue life.

Every aluminum alloy is affected by corrosion fatigue. Figure 35 shows the ratio of fatigue strength (after 10^7 cycles) in air versus in 3% NaCl solution for alloys 5087-H34, 5086-H36, 6061-T6, 7075-T73, and 2024-T3 [23]. Alloy 6061-T6 had only 55% of the fatigue strength as in air, while alloy 2024-T3 had only 25% of the fatigue strength.

3.10 Filiform Corrosion

Filiform corrosion (also known as wormtrack corrosion) is a cosmetic problem for painted aluminum [28]. Pinholes or defects in the paint from scratches or stone bruises can be the initiation site where corrosion begins with salt water pitting (Fig. 36). Filiform corrosion requires chlorides for initiation and both high humidity and chlorides for the propagation of the track. The propagation depends on where
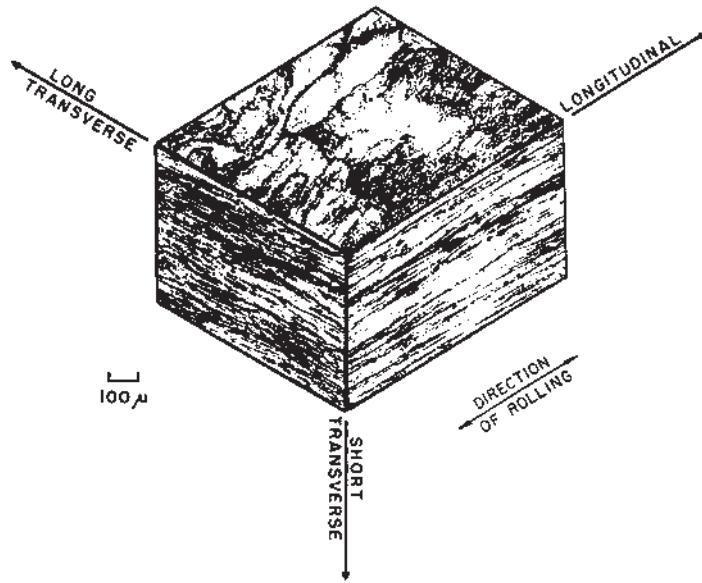
Figure 30   Rolling gives the unrecrystallized aluminum plate, different grain structures in the different directions. (From Ref. 23, with permission from ASM International.)
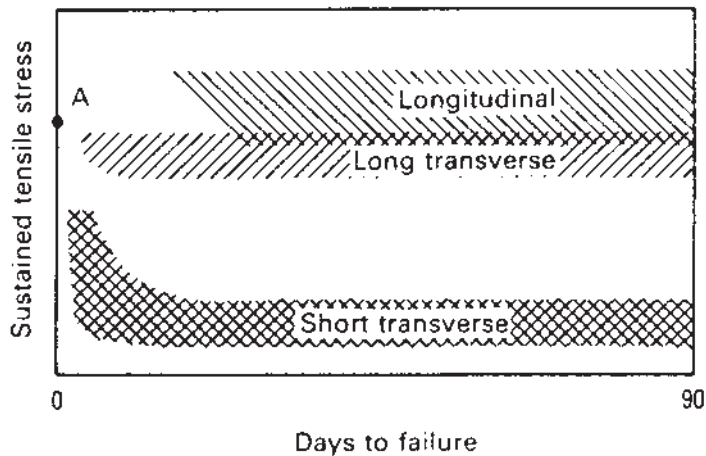
Figure 31   Stresses in the short transverse direction have the lowest threshold stress for SCC. (From Ref. 23, with permission from ASM International.)

and how the alloy is used. Maximum rate of propagation occurs at 40°C and 85% relative humidity [29]. Alloy 2036 was used for the engine hood of the Lincoln Towncar for over a decade without any incidence of filiform corrosion because of the hood's location. The heat from the engine keeps the relative humidity low, and the hood is not prone to scratches and chipping by stone bruising.
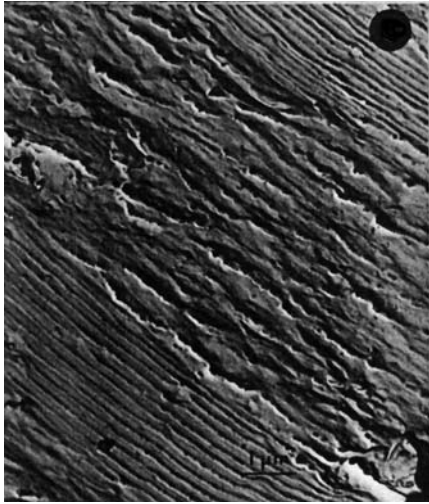
Figure 32 Characteristic fatigue striations found on the fracture face of 2024-T3 fatigued in dry air. (From Ref. 25, with permission from ASTM.)

Figure 37 is a schematic showing the mechanism of filiform corrosion. The filament must be initiated by chlorides, and then it proceeds by a mechanism similar to crevice corrosion. The head is acidic, high in chlorides, and deaerated and is the anodic site. Oxygen and water vapor diffuse through the filiform tail, and drive the cathodic reaction. Filiform corrosion can be prevented by sealing defects with paint or wax, and keeping the relative humidity low.

3.10 Microbiological Induced Corrosion

Microbiological Induced Corrosion (MIC) applies to a corrosive situation which is caused or aggravated by the biological organisms. A classic case of MIC is the growth of fungus at the water/fuel interface in aluminum aircraft fuel tanks [30]. The fungus consumes the high octane fuel, and excretes an acid which attacks and pits the aluminum fuel tank and causes leaking. The solution for this problem is to control the fuel quality and prevent water from entering or remaining in the fuel tanks. If fuel quality control is not feasible, then fungicides are sometimes added to the aircraft fuel.

4 CORROSION PREVENTION

In the previous section, the forms of aluminum corrosion and a few selected methods for corrosion prevention were discussed. This section describes the general methods for corrosion prevention of aluminum alloys.
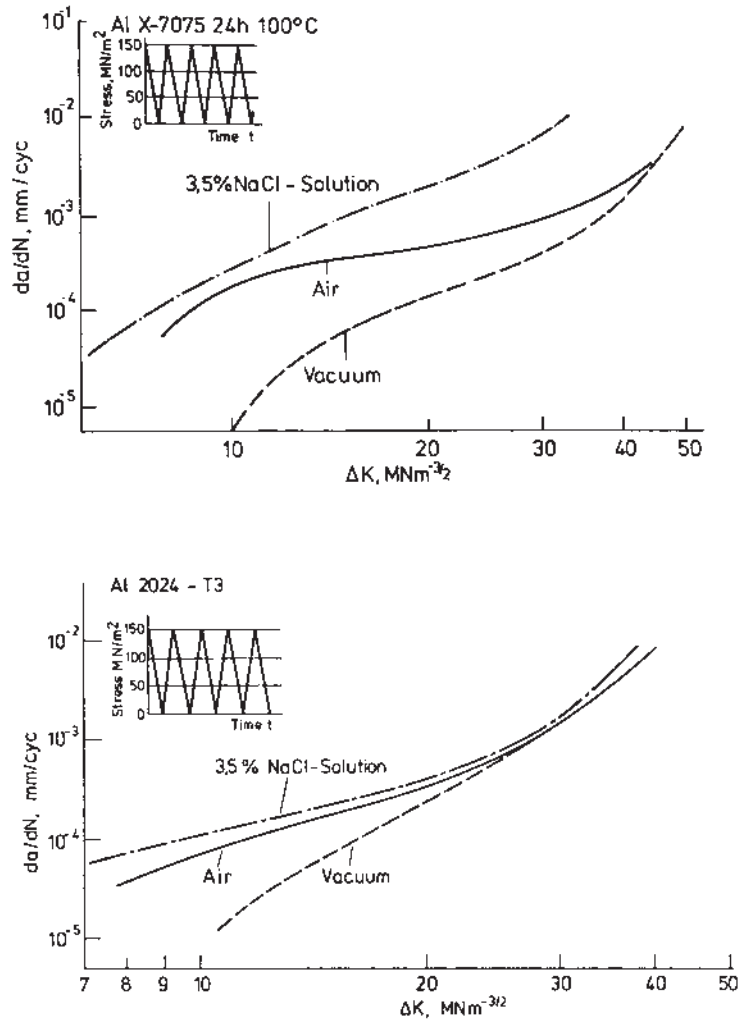
Figure 33 Fatigue cracks grow fastest in the NaCl solution, and slowest in a vacuum. The crack growth rate for in air is between the two extremes. (From Ref. 26, with permission from ASTM.)

4.1 Alloy Selection

The design stage is one of the most critical stages for corrosion prevention. A dominant factor is the specification of alloy and temper (heat treatment). Alloys with good corrosion resistance are the 1000 series, the 3000 series (in particular 3003 and 3004), the 5000 series with less than 3% Mg, and the 6000 series alloys, in particular 6061which contains Mn. Alloy 6063 (without Mn) shows poorer corrosion resistance so the Mn appears beneficial. Alloys which have poorer corrosion resistance (although higher strength) are the 2000 series, the 5000 series with $> 3\%$ Mg, and the 7000 series which contain Cu in addition to the Zn (e.g. 7075). A very
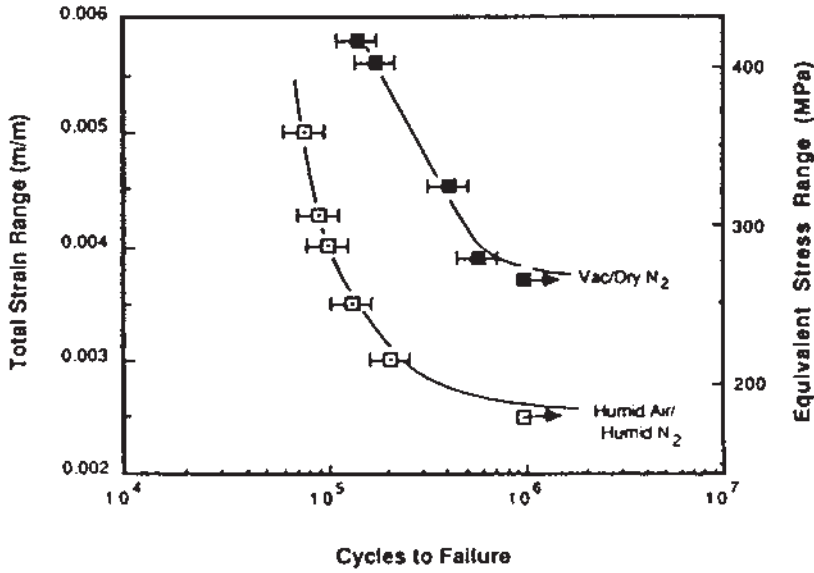
Figure 34 The high purity 7075-T6 alloy has a lower fatigue strength in humid air than in dry nitrogen, illustrating the impact of water in corrosion fatigue. (From Refs. 20b and 27, with permission from ASM International.)
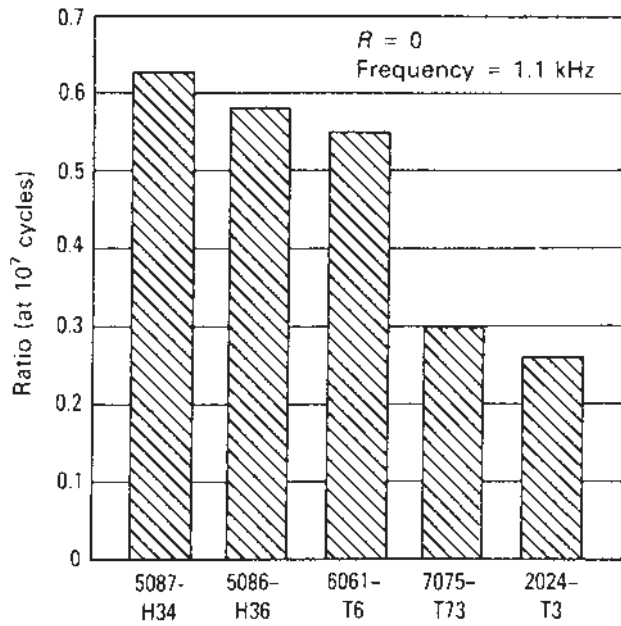
Figure 35 The ratio of fatigue strength in air versus the fatigue strength in 3% NaCl solution, after 10^7 cycles. (From Ref. 23, with permission from ASM International.)
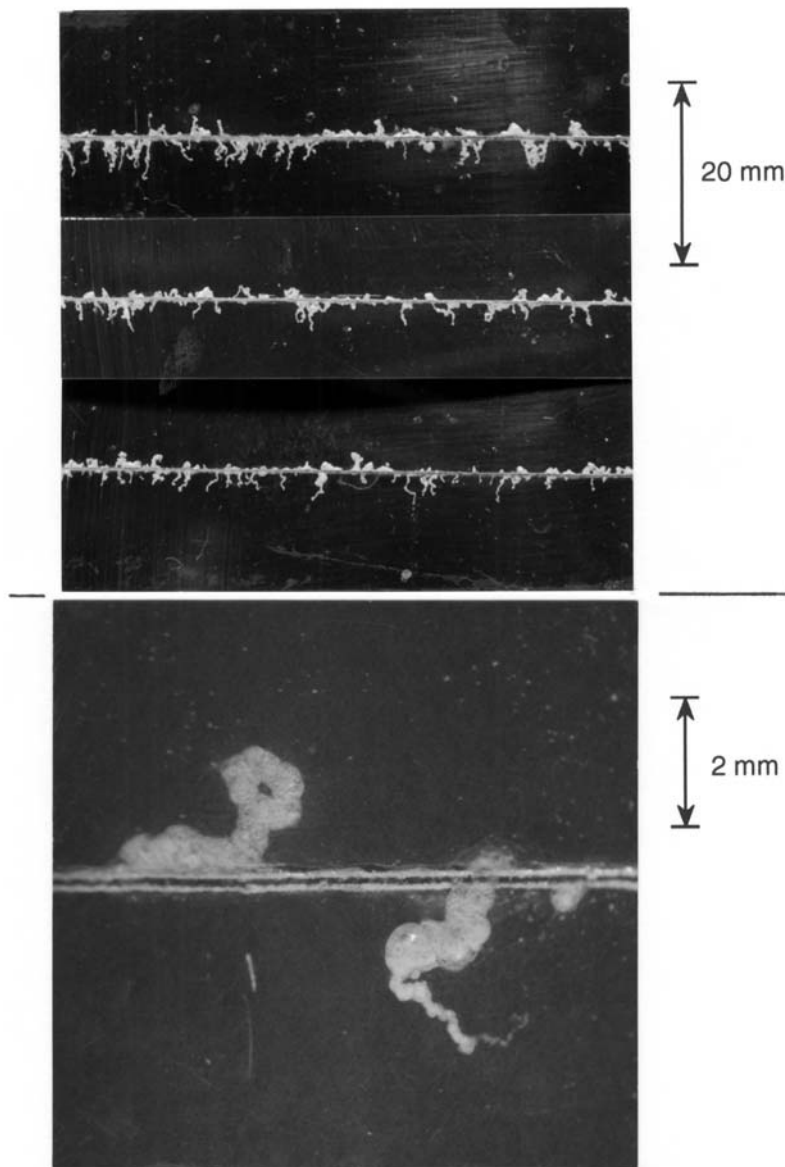
Figure 36   Filiform corrosion on clear coated aluminum. The filiform track initiates at a break in the coating (the scribe line) and wanders across the surface.

large factor in corrosion resistance is the effect of temper and presence of second phase particles. The Fe inclusions reduce the corrosion resistance by acting as pitting initiation sites. Copper precipitates in the 2000 alloys are also detrimental since these Cu containing particles also act as cathodic sites for oxygen reduction. For the 5000 series alloys with greater than 3% Mg, heat treatment can precipitate MgAl continuous precipitates along the grain boundaries. These continuous MgAl precipitates can dissolve preferentially, causing to IG attack or exfoliation.
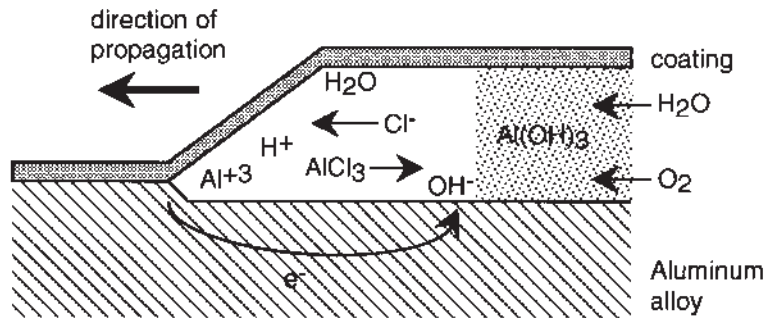
Figure 37 A cross-section of the filiform corrosion head illustrates that it is actually a type of crevice corrosion under a coating.

4.2 Product Design

The second factor to be considered during the design stage is the shape of the final product. The design should provide water runoff (no standing water) without crevices or sharp corners where the coating can be damaged. Figure 38 illustrates different designs to avoid liquid entrapment [2]. The designer should also avoid dissimilar metal contact and provide for the sealing of all fasteners, rivets and crevices.

4.3 Welds and Heat Affected Zones (HAZ)

Welded joints are a weak spot for many aluminum alloys. It is not possible to weld many aluminum alloys such as the 2024 and the 7075 due to phase segregation and cracking. For other alloys the heat affected zone (HAZ) is particularly problem-prone due to the precipitation of second phases. The HAZ is the zone between the weld metal and the unaffected base metal, where the metal has been heated, but not melted. This heating has lead to the precipitation of many second phases along the grain boundaries and the corresponding depleted zone adjacent to the grain boundaries. This HAZ is prone to intergranular corrosion.

To prevent preferential dissolution of the weld metal, the weld metal should be slightly more noble than the base alloy. Hatch [2] provides extensive tables for weld metal specifications.

4.4 Regular Maintenance

The aging aircraft program has demonstrated that a aircraft must be maintained adequately or the plane will be lost to corrosion. If the plane is carefully maintained, then its service life can be indefinite. Maintenance consists of inspecting for corrosion damage, and repairing the damage before the problem gets significant. The commercial airlines dismantle and inspect the aircraft during the ''D-Check'' [31]. Unpublished tests reported that a bare aluminum panels which were wiped with #00 steel wool and automotive wax on a regular basis maintained its shine for over ten years of semi-industrial outdoor exposure [15].
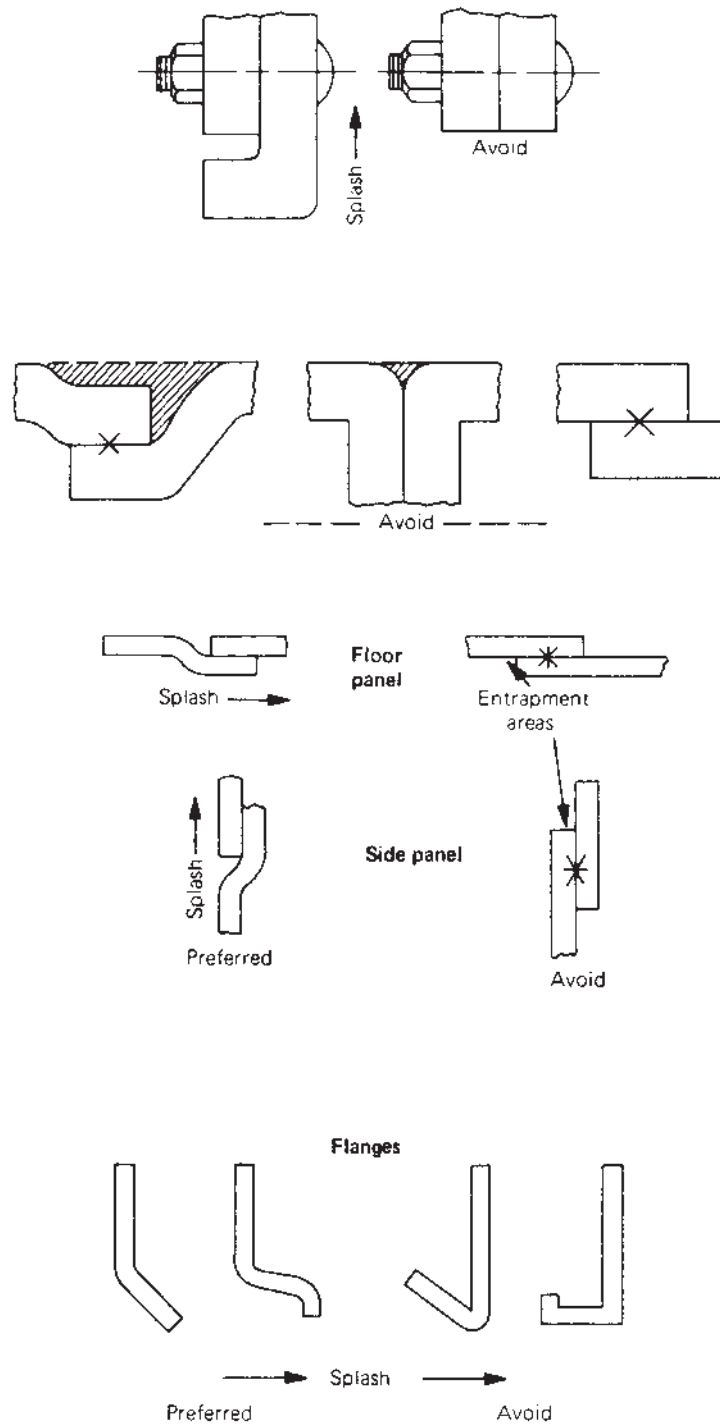
Figure 38 Designs and construction should be done to avoid liquid entrapment and corrosion. (From Ref. 2, with permission from ASM International.)

4.5   Coatings

Coatings are a standard method of protecting an aluminum structure against corrosion. (The old corrosion prevention adage, ''Paint it!'' is valid.) Polymers, (e.g. acrylic or polyurethane) are commonly used for the top coat. However the problem is keeping the polymer attached to the aluminum base metal. For this reason, a good coating usually consists of multiple layers, each with their own particular function. The initial coating on the bare aluminum is a primer or a conversion coating. The function of the primer is to chemically react with the base metal to form a rough surface layer with which the polymer coating can bond chemically and mechanically. Chromate conversion coats such as Alodine were used in the past, but these are being replaced by non-chromate conversion coats. The automotive industry uses zinc phosphate conversion coats because it works for both steel and aluminum body sheet. For use with aluminum, the zinc phosphate bath must contain fluorides in order to precipitate the dissolved aluminum which can poison the reaction.

4.6   Anodizing

Other types of conversion coating are anodizing and the electrocoat (E-coat). The science of anodizing has been covered by many authors (Wernick, [32]). In anodizing, the aluminum is immersed in an electrolyte (phosphoric acid, borate-boric buffer solution, sulfuric acid, oxalic acid, or chromic acid) and an anodic voltage is applied to the aluminum part. The aluminum reacts and grows a thick anodic oxide on the surface which consists of a barrier layer and a porous layer (Fig. 39). The porous layer can be colored with dyes and sealed by boiling in water to close the pores with aluminum hydroxide. The electrocoat is a rapid form of anodizing and polymer coating simultaneously. The E-coat is sprayed on to the positively charged aluminum, and the anodized layer is grown with the E-coat polymer.

Another form of anodizing has been reported by Kuznetsova et al. [33]. The group exposed pure aluminum in a vacuum to water vapor, and then to an electron beam at 100 V and the resulting oxide was thin but highly resistant to corrosion.

4.7   Alcladding

Alcladding is the term which first described the cladding of 2024-T3 aircraft fuselage skin with a commercial purity aluminum sheet. Alloy 2024-T3 contains 4% Cu, which age-hardens and forms Al_2Cu precipitates, which provide strength, but deteriorate the corrosion resistance. To prevent the high corrosion rate, 2024 is clad with pure aluminum, which is galvanically more anodic than the copper containing alloy (see Table 1), and so it sacrificially corrodes and protects the 2024 base alloy, Fig. 40 [34]. The alcladding is produced by hot rolling thin aluminum sheets on to the top and bottom of the 2024-T3 sheet. This aluminum sandwich is reduced in thickness, and the oxide film breaks, allowing nascent aluminum to meet and form a metallurgical bond. The pure aluminum outer layer has a more negative corrosion potential than the 2024-T3, so it acts as a sacrificial anode, preferentially corroding and protecting the 2024. This sacrificial protection is seen in Fig. 40 [34].
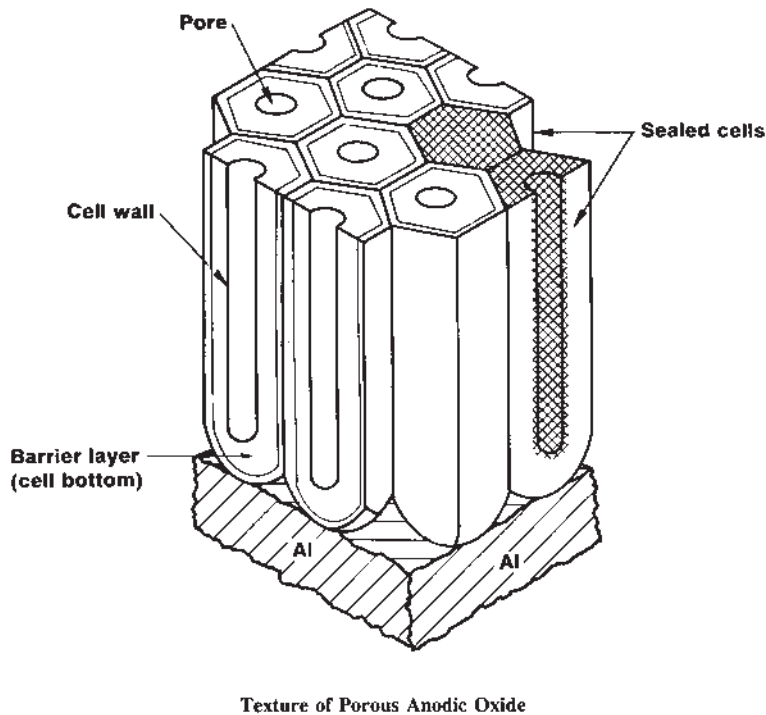
Texture of Porous Anodic Oxide

Figure 39   Cross-section of porous anodized aluminum. (From Ref. 3, with permission from Alcoa.)

The alcladding process is limited to sheet products, however a new electroplating process has been developed in which all shapes of aluminum products, forgings and machined screws, can be plated with pure aluminum, Fig. 41. This electroplating process uses an electrolyte of aluminum alkyls and metal fluorides dissolved in toluene at 90–100°C in an inert atmosphere [35]. As long as the base metal is more noble than the plating aluminum, then the plated pure aluminum will protect the base alloy from corrosion by acting as a sacrificial anode. The base metal can be an aluminum alloy, or any metal (steel or copper) which is more noble than the pure aluminum (see Fig. 13 and Table 1) In aerospace applications, plating the steel fasteners with pure aluminum can prevent galvanic corrosion of aluminum structural members.

4.8   Corrosion Inhibitors

Corrosion inhibitors are defined as any substance which when added in small concentrations to the corrosive electrolyte, lowers the corrosion rate of the metal. Chromates, silicates, polyphosphonates and soluble oils are common inhibitors for aluminum [36]. Chromates (containing hexavalent chromium) however are now known to be carcinogens and are being phased out of use. Inhibitors for mixed metal systems contain combinations of polyphosphonates, nitrites, nitrates, borates,
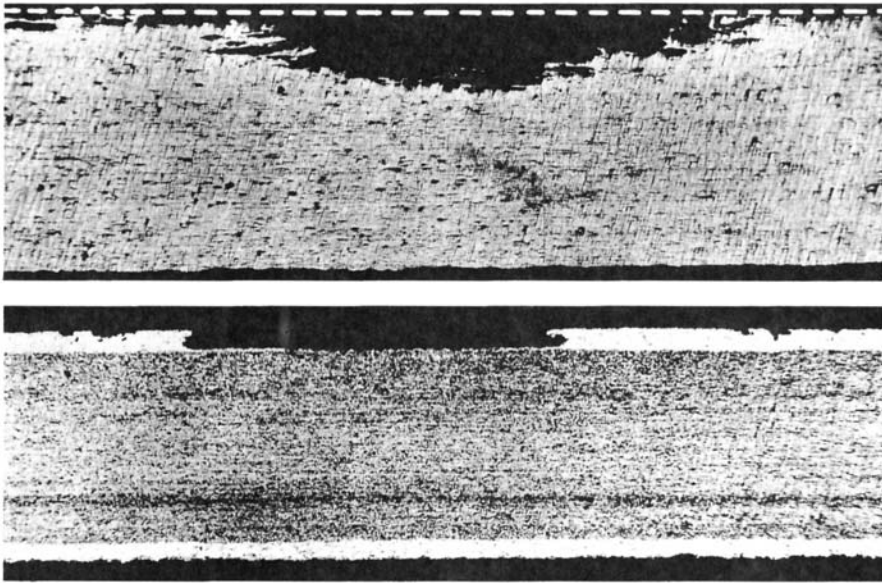
Figure 40 Alcladding sacrificially protects the underlying metal. (From Ref. 34, with permission from MIT Press.)
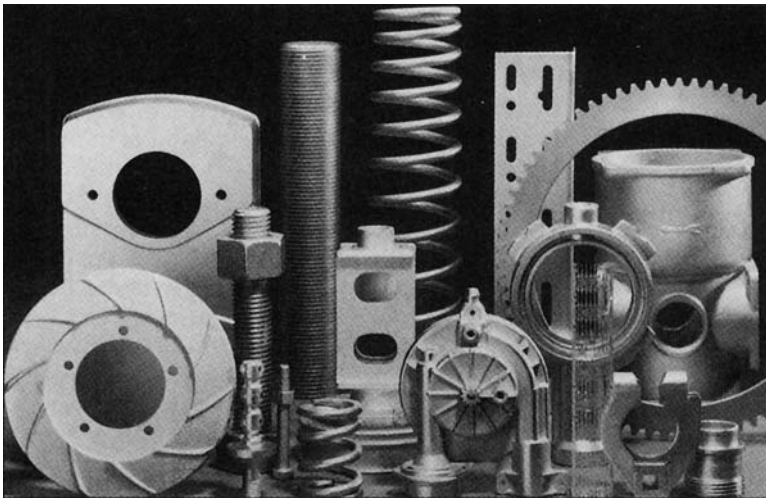
Figure 41 Electroplating aluminum from a toluene solution can coat all shapes of aluminum products, forgings and machined screws. (From Ref. 35, with permission from ASM International.)

silicates, and mercaptobenzothiazole [36]. One of the most complete references for corrosion inhibitors for aluminum is, "Korrosions-Inhibitoren für Aluminium, Blei, Kupfer, und Zink," by Witt [37].

4.9   Cathodic Protection

Aluminum may be cathodically protected by applying a negative voltage. The window of protection is very narrow; a slight voltage will protect, but a voltage more negative than -1.2 V versus the copper sulfate electrode will cause a large amount of hydrogen evolution which will form an alkaline atmosphere which dissolves aluminum. Cathodic polarization may be accomplished in soil by connecting a sacrificial zinc anode to the aluminum pipe or structure. The corrosion of the zinc imposes a slight negative potential on the aluminum and prevents it from corroding. Magnesium anodes are normally not recommended because they would apply a larger negative potential (-1.5 V) and cause caustic attack of the aluminum pipe. However magnesium anodes have been used in flowing irrigation systems and fish hatcheries since the alkaline atmosphere is continually swept away.

4.10   Sputter Deposited Aluminum

A recent development is the "stainless or sputter deposited Aluminum Alloys containing Mo, Ta, Cr, or W [38]." A potentiodynamic polarization is shown in Fig. 42. The potential required for pitting of these alloys is very high. These new alloys are not feasible via ingot metallurgy, but should have many applications in electronic circuits and thin films where corrosion resistance is required but bulk size is not.
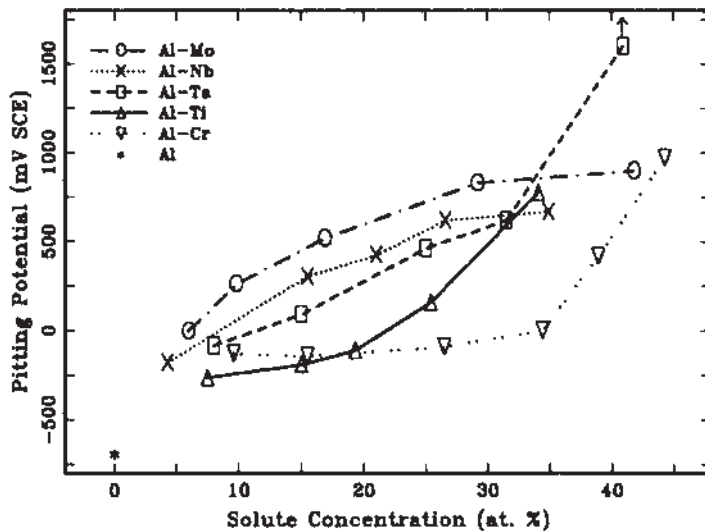
Figure 42   "Stainless aluminum" can be sputter deposited using aluminum and other elements. (From Ref. 38, with permission from The Electrochemical Society.)

4.11 The Ideal Corrosion Resistant Aluminum Alloy

The ideal corrosion resistant aluminum alloy would be single phase and homogenous. However, strengthening of aluminum requires second phases for grain boundary pinning, and precipitation hardening. Therefore the ideal aluminum alloy would contain inert, insulating, second phase particles (e.g. aluminum oxide) to avoid galvanic corrosion problems.

5 CORROSION TEST METHODS

5.1 Introduction to Test Methods

Corrosion tests are necessary in order to predict the service life of an aluminum alloy or product in a given environment. However, it is often difficult to design a correct corrosion test, and to correctly interpret the results. Every metal will corrode somewhere, especially when given sufficient time. The choice of the correct corrosion test depends on service environment where the aluminum product will operate. For example, an aluminum automobile wheel will be subjected to corrosion and cyclic fatigue in all four seasons, ranging from the humid and hot summer drive to the beach, to the salt saturated slushy snow spray on the city streets in the winter. How does one choose a corrosion test to evaluate the aluminum wheel in this service environment? Because of the complexity of service environments, numerous corrosion tests have been designed. Below is a brief review of the common tests for aluminum alloys and products. The author recommends using triplicate samples in triplicate tests before making any decision which carries a significant economic price tag. Making an important alloy decision based on the result of one or two samples tested with one method is inviting disaster.

5.2 Exposure Tests

The Alternate Immersion (ASTM G44) [17] is a cyclic test with the aluminum samples hourly immersed in 3.5% NaCl solution (10 mins) and then drying in air (50 mins). This cycle continues for 24 hr a day for 20–90 days. The American Society for the Testing of Materials (ASTM) has an extensive series of corrosion tests for aluminum alloys. Because aluminum is most susceptible to chloride pitting, the ASTM tests generally include some form of saltwater.

The classic corrosion test is the salt spray test in the saltwater fog cabinet, ASTM B117 [17]. The Saltwater Fog consists of leaving the sample continuously exposed to a mist or spray of saltwater for up to six weeks. This test imitates the conditions near a beach or in an automobile wheel well, and is used commonly for painted panels. This is an ideal test if the customer plans to use the aluminum product in a saltwater fog, but this does not necessarily correlate to any other environments. For example, filiform corrosion can initiate in the saltwater spray, but it does not propagate under the spray because of the 100% humidity.

Seacoast exposure is conducted by exposing the samples on open-air racks within sea-spray range of the ocean. The Aluminum Company of America (Alcoa) operates a seacoast exposure site at Point Judith, Rhode Island. The LaQue Center for Corrosion Technology has a commercial seacoast exposure site in North Carolina.

Colvin et al. [38b] reported that for accelerated corrosion tests of aluminum with automotive coatings, that a HCl dip test correlated well with seacoast exposure, while the Dry Bottom MASTMASSIS (ASTM G85 [17]) correlated well with highway underbody exposure. They reported that the saltspray ASTM B117 [17] did not correlate well to either seacoast nor highway exposure.

A humidity cabinet can be used to propagate filiform corrosion. Maximum growth occurs at a relative humidity of 80% and a temperature of 40°C, however the initiation must be initiated with chlorides, either with a saltwater spray, or by holding over HCl acid [29].

The intergranular corrosion test (ASTM G110) [17] uses the same NaCl-H_2O_2 solution as the ASTM G69, however after 2 hr of immersion, the aluminum sample is removed from the test, cleaned and cross-sectioned and polished to reveal the mode of corrosion attack, either pitting alone, or pitting and intergranular corrosion. This test is useful for revealing the presence of sensitized grain boundaries susceptible to intergranular corrosion.

The EXCO test (ASTM G34) [17] is an immersion test which accelerates the exfoliation of susceptible aluminum alloys. The ASTM G34 test consists of a 5–72 hr immersion in a solution containing sodium chloride, potassium nitrate, and nitric acid at room temperature. Typical EXCO results are shown in Fig. 43 [17].

5.3 Electrochemical Tests

The alloy composition or temper can be approximated by the aluminum alloy's corrosion potential measured by the ASTM G69 [17]. For this test, the aluminum sample is cleaned and then immersed in 1M NaCl, with 9 ml of 30% hydrogen peroxide (H_2O_2) per liter of solution for 1 hr. The corrosion potential is the average
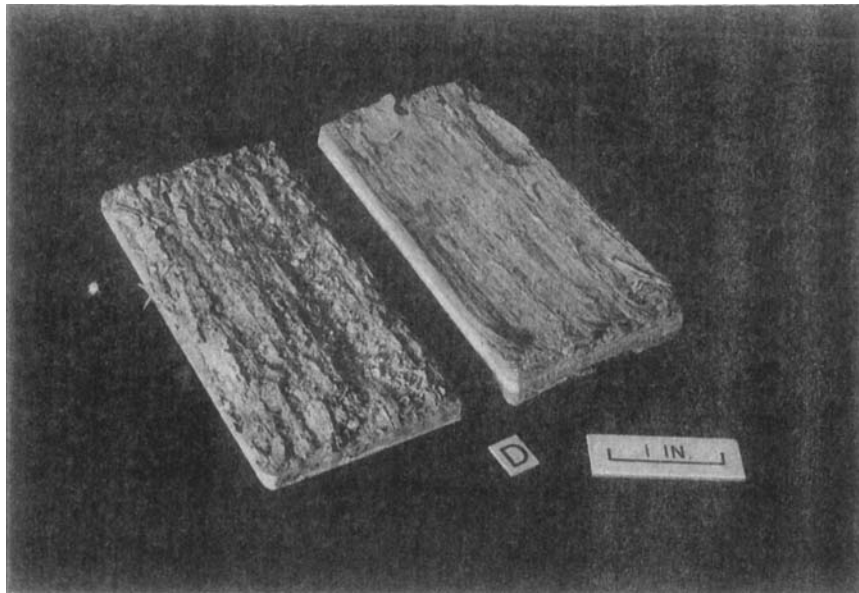
Figure 43 Typical EXCO results. (From Ref. 17, with permission from ASTM.)
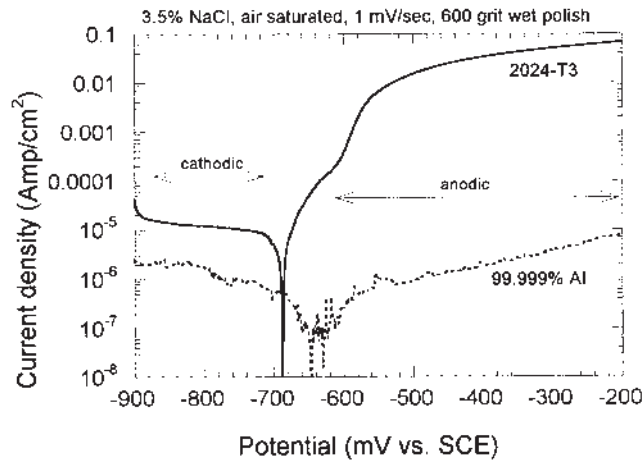
Figure 44 Potentiodynamic polarization of pure Al and alloy 2024-T3 in 3.5% NaCl solution.

open circuit potential recorded for the last half hour versus the saturated calomel electrode (SCE). The corrosion potential may be compared to the values shown in Table 1.

Potentiodynamic polarizations (ASTM G5) [17] may be used understand the corrosion mechanisms of an aluminum alloy. Figure 44 shows the potentiodynamic polarization of 99.999% pure aluminum (dotted line) versus alloy 2024-T3 (solid line) in a 3.5% NaCl solution saturated with air. The sample surfaces were wet polished with distilled water and 600 grit SiC paper prior to testing, which caused the natural formation of a oxide-hydroxide surface film. The scan rate was 1 mV/sec, starting at −900 mV versus SCE. The cathodic reaction, oxygen reduction (Eq. (3)) occurs to the left and the anodic reaction, aluminum oxidation, occurs to the right. The pure aluminum exhibits a cathodic current approximately ten times lower than the cathodic current on the 2024-T3 because there a very few second phases which break through the film and provide an electrical path for the electrons. The high currents for the 2024-T3 in the anodic region indicate a high rate of pitting corrosion. The pure aluminum did not show pitting but stayed passive because there were no breaks in the passive film which would allow the chloride ion ingress. By extrapolating the cathodic and anodic currents to their intersection, one may estimate the corrosion rate of each system. The 2024-T3 exhibits a corrosion rate of about 10 μA/cm^2, while the corrosion rate of the pure aluminum is closer to 0.1 μA/cm^2.

Electrochemical Impedance Spectroscopy (EIS) is a non-destructive method which can be used to look for pinholes or defects in coatings on aluminum [39] or measure the thickness of coatings or anodized layers [40]. An EIS test is performed by passing a small sine wave voltage (±5 mV) through an immersed sample and measuring the resulting current. The AC resistance of a system is called the impedance, /Z/. The lag between the maximum current and the maximum voltage is known as the phase shift, θ. These parameters, the phase shift, θ, and the impedance, /Z/, generated from EIS testing may be used to estimate the film quality
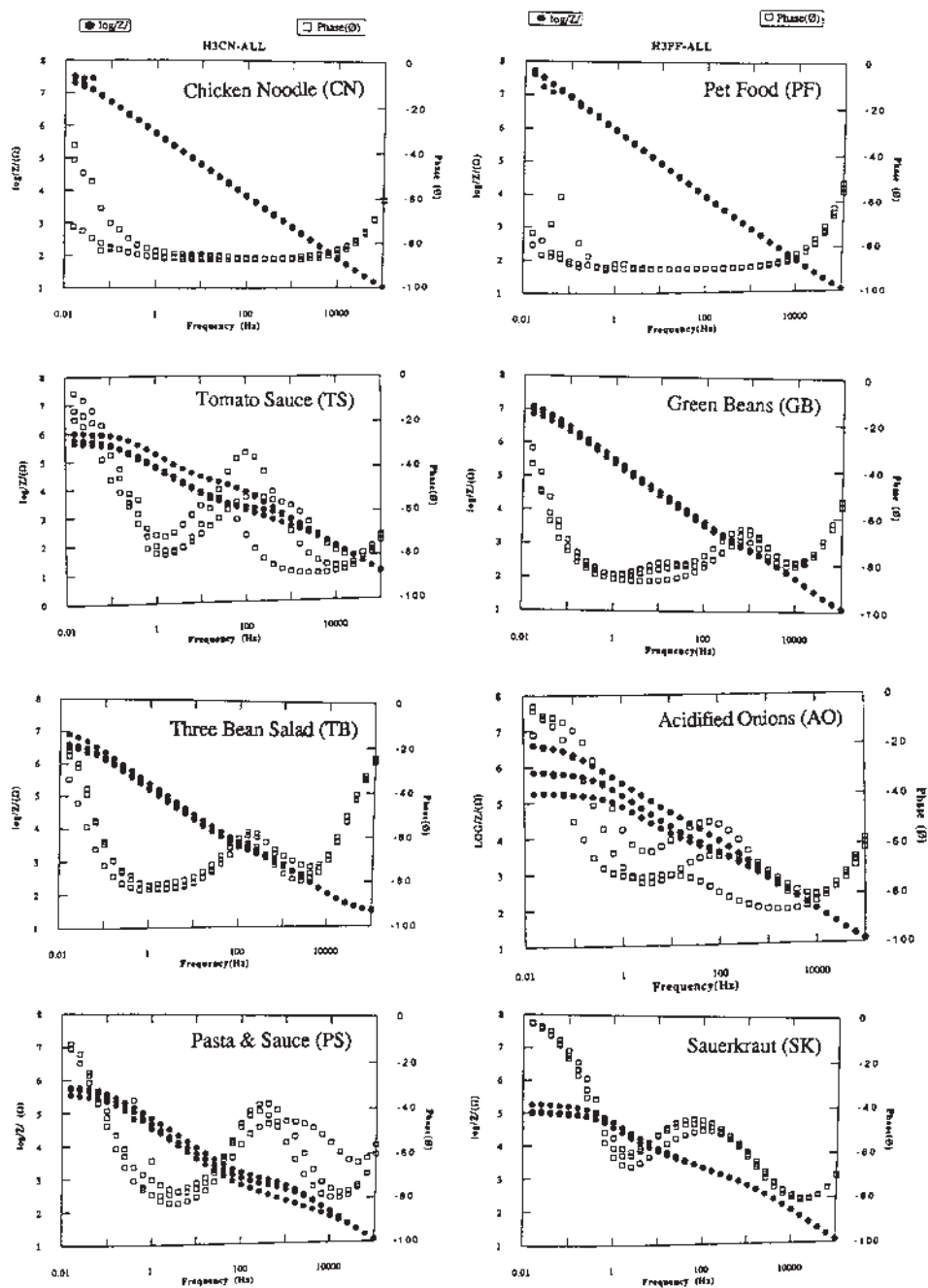
Figure 45 Electrochemical impedance spectroscopy can be used to detect pinholes in coatings. (From Ref. 39, with permission from Electrochimica Acta.)

and thickness. The part of the curve that gives the most information about the quality of the layer lies at a low frequency, 0.1 Hertz. The data at 0.1 Hz can show the presence of pinholes in the coating by a flattening of the impedance or a decreasing phase angle. The quality of the coating is maximized at the point where the impedance, /Z/, and the phase shift, θ, are at their maximum. A plot of log(f) versus log/Z/ can be interpreted to provide the coating thickness and the average corrosion rate under the coating. The thickness of the coating can also be estimated from the capacitance of the coating by approximating the layer as a parallel plate capacitor at frequencies where the phase angle approaches $-90°$. With this approximation the following equation can be utilized in order to determine the thickness of the oxide layer [40]:

$$d = \varepsilon_1 \varepsilon_0 A 2\pi f (Z - Z_0) \qquad \text{(parallel plate capacitor)} \qquad (10)$$

where d = thickness of the layer, ε_1 = dielectric constant of the layer, ε_o = 8.854×10^{-12} F/m (dielectric of free space), A = surface area (cm^2), f = frequency frequency (Hz), Z = impedance (Ω) at frequency f, and Z_0 = impedance at high frequency (a.k.a. solution resistance). Typical EIS results are shown in Fig. 45 for coated aluminum, in which the coatings are ranked from the best to the worst, from the top to the bottom. The impedance /Z/ for triplicate coated aluminum cans is shown as the solid circles. The phase angle is shown as the open squares. The impedance shows almost linear behavior after two years of storing chicken noodle soup and pet food, but a decrease in the impedance, and in the coating quality, after storing sauerkraut in similar coated cans for the same period. The phase angle drops from $-90°$ for the two top curves, to smaller values, also indicating a change from capacitive behavior to resistive behavior.

For additional information on corrosion tests for aluminum, a recent review article has been written by Lifka [41]. Additional reviews on the corrosion of aluminum alloys may be found in the *ASM Handbook*, Vol. 13 [23], in *Aluminum*, edited by J. E. Hatch [2], and in the *Corrosion Handbook* [42].

6   SUMMARY

Aluminum alloys will provide indefinite service life in certain environments, such as at room temperature in air or in a neutral aqueous solutions free from chlorides. In these benign environments, aluminum forms a protective oxide/hydroxide film on its surface which prevents further corrosion. However, in acidic, or caustic environments, aluminum can corrode. Additionally in neutral pH aqueous solutions which contain oxygen and chlorides, aluminum will corrode due to the formation of acidic occluded cells. However, with careful design, careful choice of alloys, coatings and maintenance, aluminum can provide a long and indefinite service life.

REFERENCES

1. M. Pourbaix, "Atlas of Electrochemical Equilibria in Aqueous Solutions," NACE and Cebelcor, Houston, 1974.
2. J. E. Hatch, (ed.), *Aluminum, Properties and Physical Metallurgy* 1984, ASM International, Materials Park, OH, 44073-0002, p. 295 (Fig. 19), p. 261 (Fig. 3), p. 303 (Fig. 27).

3.   K. Wefers and C. Misra, Oxides and Hydroxides of Aluminum—Technical Paper #19, 1987, Alcoa Laboratories, Alcoa Center, PA.

4.   H. L. Ewalds, F. C. van Doorn, and W. G. Sloof, *Corrosion Fatigue*, ASTM STP 801, (T. W. Crooker and B. N. Leis, eds.), 1983, ASTM, pp. 115–134.

5.   W. F. Linke, *Solubilities of Inorganic and Metal-Organic Compounds*, 1958, Van Nostrand Company, Inc., NY.

6.   T. P. Hoar, J. Electrochem. Society, 1970, 117(1), pp. 17C–22C.

7.   K. P. Wong and R. C. Alkire, J. Electrochem. Soc., 1990, 137(10), pp. 3010–3016.

8.   H. P. Hack, "Evaluation of Galvanic Corrosion," *ASM Handbook*, Vol. 13, Corrosion 1987, ASM International, Materials Park, OH, 44073-0002, p. 235 (Fig. 1).

9.   T. D. Burleigh, R. C. Rennick, and F. S. Bovard, "Corrosion Potentials for Aluminum Alloys measured by ASTM G69," Corrosion 1993, 49(8), pp. 683–685.

10. E. L. Colvin, private communication.

11. P. Schmutz and G. S. Frankel, "Corrosion Study of AA2024-T3 by Scanning Kelvin Probe Force Microscopy and In Situ AFM Scratching," J. Electrochem. Soc., 1998, 145, p. 2295.

12. T. D. Burleigh, E. Ludwiczak, and R. A. Petri, "Intergranular Corrosion of an Al-Mg-Si-Cu Alloy," Corrosion 1995, 51(1), pp. 50–55.

12b. T. D. Burleigh, E. Ludwiczak, and R. A. Petri, "ERRATA," Corrosion 1995, 51(3), p. 248.

13. M. C. Reboul, T. J. Warner, H. Maye, and B. Baroux, "A Ten-Step Mechanism for the Pitting Corrosion of Aluminum," *Materials Science Forum*, 1996, Vols. 217–222, Transtec. Publication, Switzerland, pp. 1553–1558.

14. W. H. Aylor, "Ten-year Seawater Test on Aluminum," *Corrosion in Natural Environments*, STP 558m, ASTM, 1974, p. 117.

15. A. J. Becker, private communication.

16. J. J. Liput, Jr. (ed.), *Alcoa Case Histories*, Alloy Technology Division, Alcoa Laboratories, Alcoa Center, PA (1991).

16b. D. A. Jones, "Effect of Water Chemistry on the Erosion-Corrosion of Aluminum in High Temperature High Velocity Water," Corrosion, October 1981, 37(10), p. 563–569.

16c. Y. Li, G. T. Burstein, and I. M. Hutchings, "The influence of corrosion on the erosion of aluminum by aqueous silica slurries," Wear, 1995, 186–187, pp. 515–522.

17. *Annual Book of ASTM Standards*, Vol. 3.02, Wear and Erosion: Metal Corrosion, 1998, ASTM Philadelphia, PA.

18. T. D. Burleigh, E. H. Gillespie, and S. C. Biondich, "Blowout of Aluminum Alloy 5182 Can Ends Caused by Transgranular Stress Corrosion Cracking," *Aluminum Alloys for Packaging* (J. G. Morris, H. D. Merchant, E. J. Westerman, and P. L. Morris, eds.), 1993, The Minerals, Metals, and Materials Society, pp. 323–332.

19. D. O. Sprowls, "High strength aluminum alloys with improved resistance to corrosion and stress-corrosion cracking," Aluminium (Duesseldorf), 1978, 54(3), pp. 214–217.

20. N. J. H. Holroyd, "Environment Induced Cracking of High Strength Aluminum," in *Environment Induced Cracking of Metals*, Proceedings, 1990, NACE International, pp. 311–345.

20b. A. J. McEvily, Jr., *Atlas of Stress-Corrosion and Corrosion Fatigue Curves*," 1990, ASM International, Materials Park, OH, 44073-0002, pp. 268, 288, 343.

21. T. D. Burleigh, "The Postulated Mechanisms for Stress Corrosion Cracking of Aluminum Alloys—A Review of the Literature 1980–1989," Corrosion 1991, 47(2), pp. 89-98.

22. D. A. Lukasak and R. M. Hart, "Aluminum Alloy Development Efforts for Copmpression Dominated Structure of Aircraft," Paper 1985, 50th Conf., Society of Allied Weight Engineers, Inc., (S.A.W.E.), May 1991, San Diego, CA.

23. E. H. Hollingsworth and H. Y. Hunsicker, "Corrosion of Aluminum and Aluminum Alloys," *ASM Handbook*, Vol. 13, Corrosion, 1987, ASM International, Materials Park, OH, 44073-0002, p. 592 (Fig. 11).

24. B. W. Lifka and D. Sprowls, Paper presented at the 7th Annual TMS-AIME Spring Meeting, Toronto, Ontario, May 19, 1975.

25. C. Q. Bowles and J. Schijive, "Experimental Observations of Environmental Contributions to Fatigue Crack Growth," *Corrosion Fatigue*, ASTM STP 801, (T. W. Crooker and B. N. Leis, eds)., 1983, ASTM, pp. 96–114.

26. K. Schulte, H. Nowack, and G Lütjering, *Environmentally Assisted Cracking*, ASTM STP 1049, (W. B. Lisagor, T. W. Crooker, and B. N. Leis, eds)., 1990, ASTM, pp. 347–373.

27. R. E. Ricker and D. J. Duquette, Met. Trans. A, 19A, July 1988, pp. 1775–1784.

28. H. Leth-Olsen, and K. Nisancioglu, "Filiform Corrosion Morphologies on Painted Aluminum 3105 Coil Material," Corrosion, 1997, 53(9), 705–717.

29. W. H. W. Slabaugh, S. E. Dejager, S. E. Hoover, and L. L. Hutchinson, "Filiform Corrosion of Aluminum," J. Paint Technol., 1972, 44, 566, pp. 76–83.

30. S. C. Dexter, "Localized Biological Corrosion," *ASM Handbook*, Vol. 13, Corrosion, (formerly 9th Edn, *Metals Handbook*), ASM International, Metals Park, OH, pp. 114–122.

31. E. E. Murphy, "Aging Aircraft: too old to fly?" IEEE Spectrum, June 1989, 26, pp. 28–31.

32. S. Wernick, R. Pinner, and P. G. Sheasby, "Surface Treatment and Finishing of Aluminum and its Alloys," 5th Edn, Vol. 1, 1987, ASM International, Metals Park, OH.

33. A. Kuznetsova, T. D. Burleigh, V. Zhukov, J. Blachere, and J. T. Yates Jr., "Electrochemical Evaluation of a New Type of Corrosion Passivation Layer: Artificially Produced Al_2O_3 Films on Al," Langmuir, 1998, 14(9), pp. 2502–2507.

34. R. H. Brown, "Aluminum Alloy Laminates," in *Composite Engineering Laminates*, (A. G. H. Dietz, ed.), 1969, MIT Press.

35. W. Frömberg and F. A. S. Donaldson, "Electroplating with Aluminum," Advanced Materials & Processes, 1996, 149(2), pp. 33–35.

36. A. H. Roebuck, "Inhibition of Aluminum," in *Corrosion Inhibitors* (C. C. Nathan, ed.), 1973, NACE, pp. 240–244.

37. C. A. Witt, "Korrosion-Inhibitoren für Aluminium, Blei, Kupfer und Zink," 1999, Aluminium-Verlag, Düsseldorf.

38. G. S. Frankel, M. A. Russak, C. V. Jahnes, M. Mirzamaani, and V. A. Brusic, "Pitting of Sputtered Aluminum Alloy Thin Films," J. Electrochem. Soc., 1989, 136(4), pp. 1243–1244.

38b. E. L. Colvin, L. F. Vega, M. W. Egbert, T. Nakayama, K. Ikeda, H. Shige, T. Ueda, and M. Inman, "Filiform Corrosion of Aluminum Auto Body Sheet in Accelerated and Outdoor Environments," Society of Automotive Engineers, 1997, Paper 970735.

39. F. S. Bovard, T. D. Burleigh, and A. T. Smith, "Electrochemical Impedance Spectroscopy of Electrocoated Aluminum Food Cans," Electrochimica Acta, 1995, 40(2), pp. 201–207.

40. T. D. Burleigh and A. T. Smith, "A Simple Impedance Spectra Method to Measure the Thickness of Nonporous Anodic Oxides on Aluminum," J. Electrochemical Society, 1991, 138(8), pp. L34–35.

41. B. Lifka, in *Corrosion Engineering Handbook* (P. A. Schweitzer, ed.), 1996, Marcel Dekker, NY, p. 129.

42. R. B. Mears, "Aluminum and Aluminum Alloys," in *Corrosion Handbook* (H. H. Uhlig ed.), 1948, John Wiley & Sons, Inc., NY.

12

Surface Chemistry of Adhesion to Aluminum

MARGARET M. HYLAND

University of Auckland, Auckland, New Zealand

1 INTRODUCTION

Understanding the microstructure and chemistry of the aluminum surface is the key to designing coating and structural bonding systems that endure. In important practical applications, bonding of structural components in aircraft, lacquer systems in packaging, paint, and decorative coatings for architectural, automotive, and general engineering components, the adhesion is generally of a polymer to the aluminum surface. The polymer layer may be a phenol or epoxy structural adhesive, a polyethylene, or other polyolefin coating in flexible packaging systems, or a paint or lacquer. Examples of systems involving bonding of a polymer layer to an aluminum substrate are shown schematically in Fig. 1. In each case, the polymer and the aluminum substrate are linked through the oxide layer on the surface of the metal. This layer, which may be as thin as a few nanometers, plays a crucial role in determining the strength and endurance of the bonding system, and much effort is spent on designing and maintaining a pretreatment schedule which ensures that it is engineered to have the optimum composition and microstructure. The 'optimum' composition and microstructure will vary from system to system and will depend on the nature of the polymer layer, the mechanical requirements of the bond, and the environment it will be exposed to in service.

Surface treatments usually involve a number of steps carried out under strictly controlled conditions and times. They can be roughly broken down into a cleaning step — to remove oils, greases, and dirt, and a step in which the existing oxide is removed; in the same or a following step, a new, tailored oxide or conversion coating
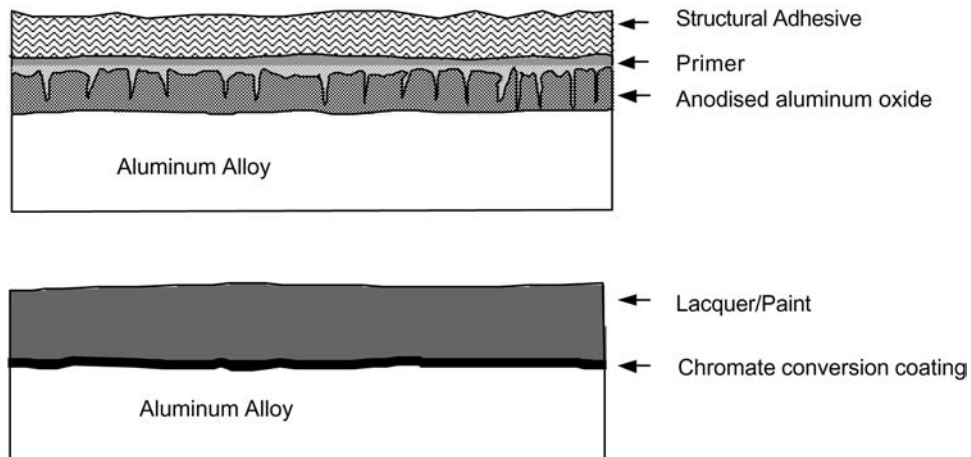
Figure 1 Schematic of the interface region for two polymer/aluminum systems.

is formed. This may be followed by sealing or priming of the surface. Examples of surface preparations [1–4] are given in Table 1. In general, the more demanding the environmental and/or mechanical performance required, the more extensive the pretreatment procedure. The requirements placed on a structural bond in an aircraft, for example, will be significantly greater those that for a packaging laminate, and the pretreatment process reflects this. The focus of this chapter is to examine key concepts common to all polymer/aluminum bonding applications and to examine how some common surface treatments alter the surface chemistry and microstructure and effect adhesion.

2   THE UNTREATED ALUMINUM SURFACE

The surfaces of aluminum alloys after forming are not conducive to adhesion. They will be covered with an aluminum oxide of varying thickness and composition

Table 1 Examples of Surface Treatments for Coating and Structural Applications

Food packaging laminates [1]	Anneal 280°C
Structural bonding in aircraft components [2]	Vapor degrease or alkaline clean; vacublast (dry alumina); vapor degrease or alkaline clean; chromic/sulfuric etch; chromic acid anodize; chromate primer; bond with epoxy-based adhesive
Screen cloth [3]	Alkaline clean; rinse; phosphate/chromate conversion coat; rinse; dry; varnish/paint
Landing gear [3,4]	Clean in non-etching cleaner (alkaline); rinse; deoxidize with chromic/sulfuric etch; rinse; chromate conversion coat; rinse; apply primer; polyurethane topcoat

depending on the preceding forming steps. Oxides formed at low temperature, below about 375°C, are composed of a thin, 1–2 nm amorphous γ–Al_2O_3 adjacent to the metal which is overlaid with progressively hydrated surface oxides and hydroxides [5]. The low-temperature hydrated overlayer is typically composed of hydrated gel-like pseudoboehmite (AlOOH), crystalline boehmite, and/or trihydroxides, $Al(OH)_3$, bayerite, or gibbsite depending on the relative humidity and exposure time [5–11], and the total oxide thickness may be between 2 and 60 nm [8,12–14]. The thicker oxides, in particular, lack the microstructure, porosity, mechanical strength, and hydration resistance to form durable bonds with a coating or adhesive. If higher temperature treatments such as annealing/degreasing have taken place, then the oxide will also contain oxides or hydroxides of the alloying elements. Magnesium, Si, Li, Na, Mn, and Cu preferentially segregate to the surface, having detrimental effects on the hydration/corrosion resistance of the interface [15–19]. Precipitates of intermetallics can also form, which, in the case of Fe intermetallics, reduce the interfacial hydration resistance or, in the case of Cu intermetallics, reduce the interfacial corrosion resistance [19,20].

In addition to the oxide/hydroxide layer, the surface will be contaminated with hydrocarbon and silicone processing aids, lubricants, rolling oils, etc., which hamper wetting of the surface by the adherent [21]. Below the oxide, a layer of deformed, micrograined metal will be present, especially in rolled, then heat-treated alloys. During rolling, surface oxides will become incorporated into this deformed substrate layer, which, researchers have shown, contributes to reduced corrosion resistance [22]. Thus, pretreatments are designed to remove the organic contaminants, remove the existing oxide layer, and develop a tailored surface layer with the required microstructure and chemistry to promote bonding and resist environmental attack. However, once the desired surface structure has been obtained, it will still be susceptible to attack by moisture and should be used promptly, or stored under appropriate conditions to prevent a friable, hydrated oxide from reforming.

3   ADHESION TO ALUMINUM SUBSTRATES: GENERAL PRINCIPLES

Strong bonds between an aluminum surface, or, more correctly, between the oxide layer on the aluminum surface and an organic coating or adhesive, are generally acknowledged to be the result of both mechanical interlocking and physico-chemical bonding [23,24]. Effective mechanical interlocking requires the surface oxide to be sufficiently rough to allow the coating/adhesive to key into the aluminum surface. The greater the interlocking, the more difficult it will be for a crack to propagate at the interface, and the greater the amount of plastic deformation the substrate and/or polymer must undergo for the bond to rupture, thereby increasing the energy needed to separate the interface. Surface roughness also increases the area of contact between the aluminum substrate and the polymer, maximizing the number of potential sites for physico-chemical bonding. Treatments such as grit blasting and acid etching will increase the macroroughness of the surface by introducing surface features such as scallops and whiskers, but this can be increased much more by generating 'microroughness,' in other words, fine porosity, in the oxide layer itself. Figure 2 is a schematic of the macroroughness and microroughness generated by chromic acid etching and anodization. The oxides produced by anodization, particularly in phosphoric acid, contain such porosity in the form of channels
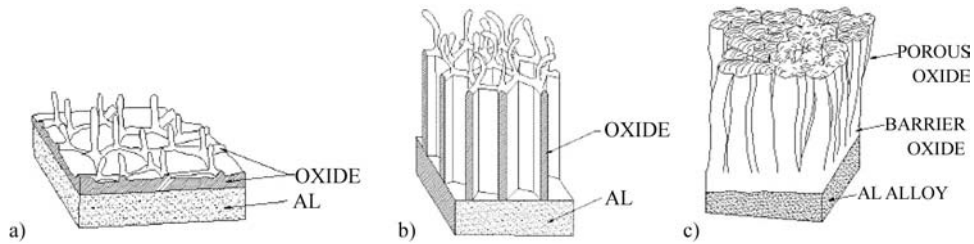
Figure 2 Surface morphologies of (a) FPL etched aluminum, (b) phosphoric acid anodized aluminum with FPL pretreatment, and (c) chromic acid anodized aluminum. Reprinted from Ref. [48] with permission from Elsevier Science.

extending from the outer surface almost to the surface of the metal. Anodization is the preferred surface treatment in the preparation of structurally bonded joints in aircraft because of the strength and durability of the bond formed. When the polymer can penetrate well into the surface oxide, physico-chemical bonding and mechanical interlocking are maximized, forming what is sometimes referred to as a *microcomposite*. Figure 3a shows an example of microcomposite formation in a structural bond [25]. The image is a cross-section of a phosphoric acid anodized 2014 A aluminum alloy surface to which a waterborne epoxy primer was applied. The oxide layer (green) has clearly been infiltrated by the primer (blue).

For mechanical interlocking to be effective, the polymer must be in intimate contact with the surface oxide. In other words, the polymer must be able to wet the surface oxide, penetrate the pores, and displace air from the interface. The extent of pore penetration in a fixed time interval will depend on the viscosity of the organic, the size and shape of the pore, and the polymer/substrate contact angle [23]. The contact angle is a measure of how well a liquid wets a solid or, for bonding to aluminum, how well a polymer coating or adhesive wets the surface oxide on an aluminum substrate. It is measured by placing a droplet on a substrate; as shown in Fig. 4, the smaller the angle, the greater the wetting. Aluminum oxide surfaces have high surface energies and are, generally speaking, more wettable than low surface
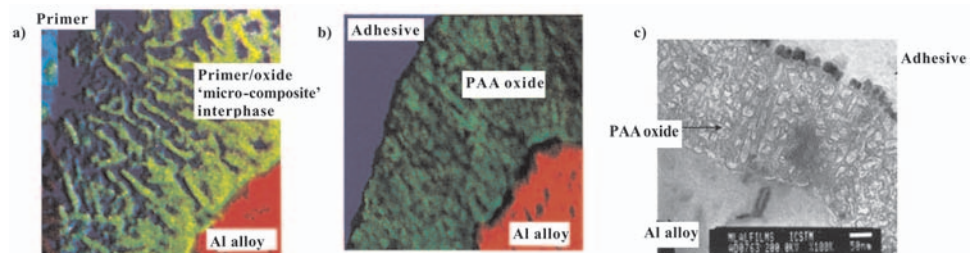
Figure 3 EFTEM and TEM micrographs of cross-sections of PAA–epoxy joints [25]. (a) EFTEM of a primed PAA joint showing infiltration of the oxide by the primer, forming a 'microcomposite' interphase above the Al alloy. (b) EFTEM of a non-primed PAA joint. The epoxy has not penetrated the oxide. (c) TEM of non-primed joint after wet cyclic fatigue testing. The PAA oxide has been hydrated, especially at the adhesive/epoxy interface. Reprinted from Ref. [25] with permission from Elsevier Science.
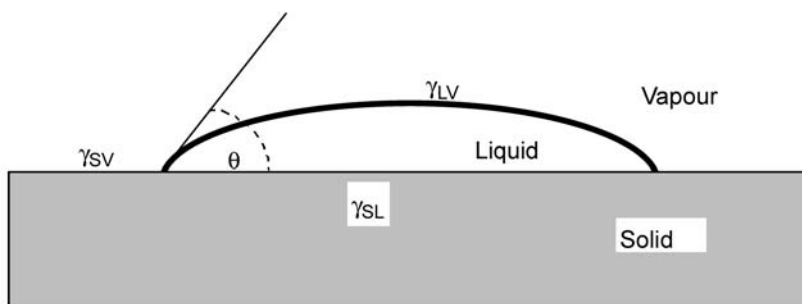
Figure 4 Schematic of droplet wetting a surface. The contact angle is θ, and γ_{LV}, γ_{SV}, and γ_{SL} are the interfacial tensions between the liquid–vapor, solid–vapor, and solid–liquid phases, respectively. When the contact angle is zero, the liquid spontaneously spreads at a rate determined by the liquid viscosity and surface roughness.

energy surfaces like teflon [23], but the wettability of aluminum oxide is reduced when the surface is covered by contaminant hydrocarbon. The common water break test gives a measure of this. Where the aluminum surface is clean, the contact angle is small, and water spreads across the surface freely. Where there is surface contamination (usually hydrocarbon), the contact angle is high, and the water droplets bead and do not spread. On a properly prepared aluminum surface, the polymer coating or adhesive should also spread across the surface, but it should be noted that a positive water bead test is not necessarily a good indicator of whether the polymer will also wet the surface. It signals the presence or absence of contaminant hydrocarbon, which will interfere with the wetting of the aluminum surface by water or the polymer coating or adhesive.

Kinloch et al. demonstrated the importance of wetting on bond strength and durability by comparing the fatigue behavior of adhesive joints prepared using a variety of metal pretreatments [25]. Aluminum surfaces that had been degreased only prior to bonding exhibited voids at the adhesive/oxide interface indicative of poor adhesive/oxide wetting (Fig. 3). These joints had lower fracture energies and strain release rates in both dry and humid cyclic fatigue testing compared to chromic acid etched and phosphoric acid anodized surfaces, where the adhesive wetted the oxide surfaces. Voids at the interface will act as stress concentrators, reducing the mechanical strength of the interface. In addition, such unfilled pores at the interface will allow moisture to more readily penetrate the interface – moisture at the interface is one of the major causes of bond failure, as discussed in the following section.

Physico-chemical bonds can range in strength from weak van der Waals interactions to acid–base interactions and covalent bond formation. The van der Waals forces are relatively weak and operate over extremely short distances – the strength of the interaction falling off as a function of distance to the sixth power [23]. They are due to the presence of dipoles – unequal electron sharing between an electronegative atom, such as fluorine or chlorine, and a more electropositive atom, such as carbon, causing one atom to be slightly negative, the other slightly positive. Positive dipoles in the organic will interact with negative dipoles in the oxide (or vice versa). The dipole may be permanent, because of the polar nature of the atoms in the bond, or may be induced in non-polar systems by interaction with a permanent

dipole or by interaction with another induced dipole. Thus, van der Waals interactions are usually subdivided into *dipole–dipole*, *dipole–induced dipole*, and *induced dipole–induced dipole* (also called *dispersion*) forces between atoms of the substrate and those of the polymer. Although this type of interaction is weak, dispersion forces, in particular, are ubiquitous, and will be significant when the substrate and polymer are in intimate contact and when the number of such bonds is large − i.e., when the contact area is high [23,24]. Hydrocarbon polymers such as polyethylene and polypropylene that lack polar groups adhere to aluminum foil primarily through dispersive forces.

Acid–base interactions can take place at interfaces and play a role in adhesion: the work of adhesion increases when an acidic substrate is coupled with increasingly basic liquids or vice versa [23]. A variety of spectroscopic techniques have shown that the organic compounds which contain nitrogen functionalities and which act as Lewis bases bond to the Lewis acid sites on the aluminum oxide surface [26,27]. Amine cured epoxies will also be able to bond through this mechanism.

One rule of thumb [23] for adhesive bonding is to choose an adhesive with the opposite acid/base character to that of the substrate. Unfortunately, determining the acid/base character of the oxide layer on an aluminum substrate is not straight-forward, since the oxide is amphoteric. The surfaces of aluminum oxides contain a variety of oxide, hydroxyl, and metal cation sites with acid or basic character depending on the bonding arrangement [5,28]. The number and type of these groups, and therefore the net acid or basic character of the surface, will vary with the specific aluminum oxide or hydroxide phase present at the surface, degree of hydration of that phase [5,28], and the presence of oxides of alloy elements or impurities, such as magnesium or silicon [29,30]. These, in turn, will be strongly dependent on the surface treatment regime. Thus, the precise role of acid–base interactions in adhesion to aluminum is not well understood, but the development of surface analytical techniques to measure the acido-basicity of aluminum surfaces offers promise [29–32]. Oxygen binding energies from x-ray photoelectron spectroscopy (XPS) correlated well with other spectroscopic and wet analytical techniques to give a measure of the acid/base nature of the surfaces of aluminum alloys treated using common pretreatment methods. The acid/base nature is apparently quite dependent on the specific treatment and alloy − there are no obvious trends as a function of treatment type (acid etch, alkaline clean, anneal, corona discharge) or alloy [29–33]. Nevertheless, changes in the surface acidity have been correlated with changes in bond strengths [32,33]. The adhesive strength of amine cured epoxy and polyurethane primers, for example, has been shown to vary with the acid/base character of the oxide layer [32], moving from adhesive to cohesive failure as the surface became more basic. However, it would be unwise to generalize these results, since a number of contributing factors, such as surface morphology and chemical reactions in the polymer, also contribute to the measured strengths.

All other factors being equal, covalent bonding at the interface has the potential to generate the strongest interfacial bonds [23,24]. One of the primary targets in the development of adhesive systems is to engineer the surface chemistry of the components to allow covalent bonds to form across the interface [23,34]. The surface aluminum oxide has the potential to form covalent bonds with some organics, but direct evidence of such bond formation is not readily found in the literature with the exception of a few important systems, carboxylates and organosilanes. One of the

few clear examples of direct chemical bond formation is the interaction of carboxylic acid groups with surface hydroxyls on aluminum oxide layers. The COO− functional group forms a monodentate bond with basic surface groups forming Al–O–C bonds, which can be evidenced directly by infrared spectroscopy [34–36]. The incorporation of the carboxylate functional group through copolymerization [36], grafting [34,37], or plasma treatment [38] has been used to enhance the adhesion of polyolefins to foil surfaces for flexible packaging applications.

The other well-studied example of covalent bonding at the aluminum interface is the silane coupling agent. Organosilanes hydrolyze in water, providing an OH group at one end of the molecule which will react with the oxide surface, forming strong Si–O–Al bonds [39,40]. The organic end of the molecule is compatible with the paint or primer, thus forming a link across the interface, increasing the adhesion strength in paint and structural bonding systems [41].

4 DURABILITY OF THE BOND − PREVENTING HYDRATION OF THE INTERFACE

These multilayered aluminum/aluminum oxide (or conversion coating)/polymer structures can fail within the polymer layer, within the oxide layer, or at the metal/oxide or oxide/polymer interfaces. A variety of test methods and geometries have been developed to measure the strength of the polymer/aluminum bond and its durability in humid or corrosive environments (see, for example, Ref. [42]). Detailed discussion of the methods is outside the scope of this review, but in addition to the numerical treatment of the data obtained from the tests, the fractured specimens are analyzed to determine the locus of failure. Failure within the polymer is termed *cohesive*, and any failure at or below the polymer/oxide interface is often referred to as *interfacial* or *adhesive*. Occasionally, a distinction is made between failures at the polymer/oxide interface (*adhesive failure*) and those at the oxide/metal interface (*interfacial*) [41]. In this chapter, interfacial will refer to any failure below the polymer/oxide interface.

Polymer/aluminum bonds are notoriously susceptible to water − ingress of moisture hydrates the interface, initiating interfacial failure. This is the major cause of bond failure in such systems. Analyses of the two sides of the failed bond show that the failure is often initially cohesive − that is, it takes place within the adhesive, but with exposure to humid atmospheres, the locus of failure moves to the coating−oxide interface (see, for example, Refs. [7,24,25,41]).

Generally speaking, water damages the interface in two ways: it disrupts the oxide/polymer chemical interactions and it changes the microstructure and chemistry of the oxide layer itself, thereby disrupting mechanical bonding.

In many cases, the oxide/polymer bond is thermodynamically unstable in the presence of moisture − particularly if the bond is based on weak secondary, van der Waals interactions. For example, the work of adhesion for aluminum oxide/epoxy interface has been calculated in the absence and presence of moisture. The bond is stable (work of adhesion is $178 \, mJ/m^2$) unless water is introduced, in which case the work of adhesion becomes negative ($-137 \, mJ/m^2$) [23]. Both sides of the interface will react with water. X-ray photoelectron spectroscopy shows that with water exposure, carboxylate groups form in the epoxy and the aluminum oxide becomes increasingly hydroxylated [43].

The changes that take place within the oxide layer itself are equally destructive. These changes proceed in a manner similar to the reaction of aluminum surfaces immersed in water [8]. As described earlier, moisture at the interface initially converts the aluminum oxide to a gel-like pseudoboehmite, which progressively extends into the oxide layer. The outer oxide surface eventually converts to boehmite, crystalline boehmite, and/or one of the trihydroxides, gibbsite or bayerite. The hydrated oxide is reported to be mechanically weaker than the γ-Al_2O_3 formed from the surface treatments [21]. The volume expansion that accompanies the phase transformations described may also contribute to bond failure by increasing internal stresses [7].

Water can diffuse into the interface through the polymer layer and via voids at the interface or in the oxide layer. Eliminating voids by wetting of the oxide by the polymer and intimate contact of the two can help to slow the rate of moisture ingress and improve bond durability [25,44,45]. This is illustrated in Fig. 3 [25], which shows cross-sections of aluminum/epoxy structural joints prepared by phosphoric acid anodization (PAA) and fatigue tested under dry and wet conditions. Where a water-based primer was used prior to application of the epoxy adhesive, energy filtered transmission electron microscopy (EFTEM) micrographs of cross-sections of the joint show that a microcomposite interphase, described earlier, is formed (Fig. 3a). With no primer, the adhesive fails to penetrate the porous oxide, as shown in the EFTEM micrograph (Fig. 3b). Although the two had similar initial fracture energies and similar threshold strain–energy release rates (G_{th}) when tested under dry conditions, the unprimed joint has a much lower G_{th} under wet fatigue testing. Transmission electron microscope images of the wet cyclic-fatigue-tested specimen (Fig. 3c) show that the oxide layer has undergone hydration. A dense oxide is visible at the interface between the PAA oxide and the adhesive, which would have disrupted oxide/polymer adhesion. In contrast, the primed oxide, lacking open porosity, showed no evidence of a dense hydrated oxide at the interface.

Impurities at the interface can also reduce bond strengths. Magnesium and lithium oxides or hydroxides at the surface have been shown in some systems to be detrimental to bond durability [12,15,16,19]. These elements are concentrated at the surface, sometimes a thousand-fold over their bulk concentration, during annealing. Their hydroxides are soluble and will increase the local pH when water is introduced to the interface. This increase in pH enhances the solubility of aluminum oxide, disrupting the oxide layer [15]. However, other studies have found that the presence of Mg oxides and hydroxides enhances bond stability [46,47]. Other factors such as curing chemistry of the adhesive may play a role; amount and form of the magnesium may also play a role [47].

Fluoride ions at the surface, from insufficient rinsing or rinsing using fluoride containing tap water, can form soluble aluminum species and disrupt the oxide layer, which can be sufficiently extensive to dissolve the surface layer, altering surface topography [48].

A range of strategies are used to improve the water resistance of the interface: ensuring there is minimal porosity at the interface by maximizing oxide/polymer wetting; reducing the susceptibility of the polymer/oxide bond by designing for covalent bond formation between them (silane coupling agents); employing targeted cleaning steps to selectively remove detrimental impurities; reducing the rate of oxide hydration by using hydrophobic sealants or passive, insoluble chromate or phosphate conversion coatings.

5 COMMON PRETREATMENTS FOR ALUMINUM SUBSTRATES

A multitude of prebonding surface treatments procedures are available for aluminum. Critchlow and Brewis identify 41 individual pretreatments [49]. The combination of pretreatment steps used for any particular application will depend on the mechanical requirement of the bond in service, the severity of the service environment, the alloy in question, and the polymer coating/adhesive. Details of pretreatment procedures for various applications can be found in Refs. [1,4,41, 50,51]. The pretreatments may be divided into *mechanical*, *chemical*, and *electro-chemical* [49]. This section describes some of the common aluminum pretreatments from each of these categories and shows how these treatments alter the microstructural or chemical composition of the surface to improve the strength and environmental resistance of the aluminum/polymer interface. As a general rule, the order of bond strength and durability is electrochemical > chemical > mechanical. There will always be exceptions, and the quality of the preparation and the compatibility of the polymer with the surface oxide can change these relative orders.

5.1 Mechanical Pretreatments

Mechanical pretreatments are usually the first and occasionally the only stage in a pretreatment schedule. This category covers a broad range of processes, not all of which are strictly 'mechanical,' but have the common feature of generally being less sophisticated than the chemical and electrochemical treatments described later, both in the equipment required and in the changes they induce in the aluminum surface. This stage can fulfill multiple functions: removal of gross oxide layers, surface roughening, and degreasing; these may be carried out together or as separate steps. Common mechanical pretreatments include grit blasting, mechanical abrasion, solvent degreasing, and corona discharge/plasma.

Grit blasting, generally using silica or alumina particles, and *mechanical abrasion* physically remove gross surface contamination and some of the existing oxide and roughen the substrate, enabling mechanical keying. They are often followed by a degreasing step. Larger grit particles produce greater surface roughness, and the resultant bonds are also apparently less durable [46,52] than those on smoother surfaces. At the same time, grit blasting also changes the wettability of the surface — rougher surfaces (irrespective of grit type) have lower surface energies than smoother surfaces [52]. This effect is due, in part, to the surface features such as asperities, which reduce wettability by preventing liquid flow and droplet spreading, but may also be related to Na and Mg impurities from residual grit. Grit blasting with finer grit leaves higher levels of these surface contaminants, which leaves the surface with a higher surface energy [52].

Solvent or vapor *degreasing* with solvents such as trichloroethylene, methylethyl ketone, or methylene dichloride is a cleaning step to remove hydrocarbons. This step does not affect the surface morphology (unless it is coupled with mechanical abrasion or ultrasonics) or surface chemistry of the oxide layer, nor does it remove adherent inorganic contaminants such as metal oxides and hydroxides of alloy elements. It is an important step because removal of mill oils, processing aids, etc., increases the wettability of the oxide, allowing the adhesive/coating to come into intimate contact with the oxide. Because the chemistry

and morphology of the oxide are unchanged, this may not be sufficient to ensure strong bonding if this is the only treatment applied. More often, degreasing takes place prior to other processing steps such as alkaline clean or acid etch, and the improved surface wettability means that these solutions can evenly wet and uniformly act on the alloy. Acid and alkaline clean and etches will also remove surface hydrocarbons, but they are more effective if the surface is solvent degreased first.

Foils used for laminates in food packaging are degreased in the annealing step. At annealing temperatures, the hydrocarbons are oxidized, and their removal lowers the contact angle of the surface with water [53]. In addition to the removal of hydrocarbon residues, the oxide layer thickens, and there is a significant change in the oxide composition, with the migration of alkali and alkaline earth elements such as Na and Mg to the surface [12,15–17]. The oxide itself may be converted to boehmite [11].

Aluminum may also be pretreated by exposure to activated gases in corona discharge or plasma treatments [33,38,53]. The reactive atmosphere, typically oxygen or nitrogen, oxidizes hydrocarbon residues, removing the surface contamination, increasing the number of surface acid and base groups, and reducing the water contact angle [33,53]. The oxide layer thickens – the final depth depending on treatment time and energy.

5.2   Chemical Pretreatments

5.2.1   Acid/Alkaline Etches and Cleaning

Acid and alkaline cleaners and etches are frequently used to prepare the aluminum surface. They may be used as the final treatment (for example, chromic acid etches), or used as part of a multistage process prior to conversion coating or anodizing. A degreasing step is generally carried out before acid/alkaline treatment; though these treatments can remove hydrocarbon contaminants, they are more effective if the surface has been degreased first. Acid and alkaline solutions act by chemical dissolution of the surface oxide and, to a greater or lesser extent, the metal as well. The distinction between 'cleaning' and 'etching' is therefore a matter of degree of aggressiveness and selectivity of attack. While both acid and alkaline treatments 'clean' the surface (aluminum oxide is soluble in either highly acidic or highly basic solutions), the resulting surface chemistries and textures differ, and therefore acid and alkaline treatments have different functions in a carefully tailored treatment process. For example, many pretreatment procedures employ alkaline solutions to remove the air-formed aluminum oxide, followed by an acid etch, its function being to remove oxides of alloying elements not dissolved in the alkaline treatment and/or develop a surface microstructure to enhance mechanical interlocking.

5.2.2   Alkaline Clean/Etch

Alkaline cleaners based on sodium carbonates of phosphates are less aggressive than sodium hydroxide, which tends to etch the metal significantly. Typical formulations of etching and non-etching alkaline treatments are given in Ref. [1]. Alkaline cleaners can be further inhibited by the addition of silicates, which form a protective film over the surface. Rinsing to remove the film and residual basic solution is important.

After alkaline cleaning, the surface oxide is considerably thinner than the native oxide − less than 10 nm [47]. Alkaline cleaning and etching can leave the surface enriched in alloying elements, Si, Fe, Cu, Mg, Mn, Mg [1,18,19,54], and in capacitor foils, Pb [55]. These may be as a uniform oxide or hydroxide layer, or as localized stains or 'smut.' An acid 'de-smut' treatment dissolves the residue, which, as discussed earlier, could otherwise have detrimental effects on the bond strength and corrosion resistance. Aside from this, little other information about the surface chemistry of the alkaline treated surface is available. Alkaline etched surfaces may develop a 'scalloped' structure [47,55,56]. This treatment can be used to remove the micrograined structure that develops in cold rolled alloys and which has been implicated in their poor corrosion resistance after heat treatment [19].

5.2.3 Acid Etches

Acid etches are used for a variety of purposes: to remove the native aluminum oxide; to de-smut an alkaline cleaned surface; to develop a specific surface microstructure for bonding; and as a pretreatment for anodizing because the nature of the acid etch will affect the morphology of the anodized film. The most common acid treatments for bonding applications are the chromic acid etches (CAE) − mixtures of chromic and sulfuric acids, although HF/HNO_3, HF/H_2SO_4, and chromic/phosphoric combinations are also employed. The drive to reduce the use of toxic, hexavalent chromium has led to the development of other etching systems − most notably the 'P2' etch and $FeSO_4/H_2SO_4$. Details of acid etch solutions and procedures can be found in Refs. [1,4].

Acid etching results in a thin, 5−40 nm thick oxide film [40,44,48,57]. The microstructure of such oxides formed by P2 and chromic acid etching, in particular the Forest Product Laboratories (FPL) and modified FPL chromic acid etches, have been well characterized [25,40,44,48]. A schematic of the morphology of the FPL etch is shown in Fig. 2a. The surface is covered with shallow scalloping, approximately 40 nm wide. Fine 'whiskers' protrude from the surface to a height of about 40 nm which are believed to be important in mechanical interlocking of coating/adhesives to the aluminum surface. The oxide is an amorphous Al_2O_3, and small quantities of ions of the treatment solutions are adsorbed onto the surface: Fe and sulfate in the case of the P2 etch, and Cr and F in the case of the FPL etch [44]. The surfaces of FPL etched substrates are susceptible to damage by hydration and presence of impurities. Fluoride contamination, for example, causes the whiskers to be dissolved [48].

Even when acid etches are used as an intermediate step in the treatment process, they may still influence the chemistry and structure of the final bonding surface. Depending on whether an FPL or P2 etch is carried out prior to anodization, more 'open' or 'closed' pore morphologies, respectively, are observed [44]. Acid de-smutting changes the surface chemistry by removing Mg, Si, Cu, Mn, and Fe oxides and hydroxides, which would otherwise remain at the bonding interface and degrade bond strength and durability.

5.2.4 Conversion Coatings

Conversion coatings are adherent layers of insoluble chromium oxides/hydroxides or Al, Cr, Zn, and Mn phosphates, which largely replace the aluminum oxide. They are typically used for surfaces that are to be lacquered or painted and enhance bond

durability of the polymer coating to aluminum alloys by providing corrosion protection to the metal through formation of a passive layer and presentation of a surface that enhances polymer adhesion.

Chromate conversion coatings are amorphous passive layers of two common types: chromium phosphates or chromium oxide/hydroxides. These systems provide excellent corrosion resistance and are less expensive than anodizing, which is also used as a paint or lacquer base. However, hexavalent chromium is carcinogenic, and there are significant pressures to reduce or eliminate its use. Chromium-free alternatives are being actively sought. Twite and Bierwagen [58] and Metroke et al. [59] review some of the alternative systems, which include cerium hydroxides (CEROX), silica, titania–silica, zirconia sol–gels, direct electrodeposition of coatings (e-coat), plasma polymerized coatings, and conducting polymer coatings.

Chromium oxide/hydroxide conversion solutions typically contain soluble Cr(VI), added as either CrO_3 or Na_2CrO_7, fluoride ion, usually as NaF, and sometimes ferricyanide, $Fe(CN)_6^{3-}$ accelerants, at pH 1–2. The fluoride activates the surface by dissolving the surface aluminum oxide. Once it is sufficiently thinned, reduction of Cr(VI) to Cr(III) can take place, and chromium oxide/hydroxide is deposited. The exact form of the oxyhydroxide is unclear, but it is variously represented as CrOOH [60] or hydrated chromium oxide, $Cr_2O_3 \cdot nH_2O$ [61]. Varying amounts of aluminum hydroxide will also be precipitated and incorporated into the growing chromate film [60]. A thin aluminum oxide layer remains immediately adjacent to the aluminum surface overlaid with a thicker more porous layer, giving a total thickness typically between 60 and 200 nm [60,62]. Fluoride [63], ferricyanide [64], and unreduced Cr(VI) [65,66] are also found in the coating. Chromium (IV) is important for the 'self-healing' ability of these coatings. If the conversion coating is disrupted and the metal exposed, Cr(IV) can be reduced to Cr(III) oxide/hydroxide, reforming the passive layer. The pH of the chromating solution plays an important role in the quality of the coating. As the pH is reduced, thicker coatings can form [60,61], and more Cr(VI) is incorporated into the layer [65]. At the same time, the coating becomes more porous and more permeable to ion transport [61]. Chromating can also be done using alkaline solutions, hydroxyl serving the same function as fluoride, that is, to activate the surface by dissolving the residual oxide.

The uniformity of the coating is dependent on surface treatments prior to chromating and on the microstructure, in particular the presence on intermetallic precipitates [63,67–69]. The chromium oxide/hydroxide forms as isolated deposits, which grow in size and number with the treatment time [62]. Degreasing and uniform removal of the surface oxide (deoxidizing) by alkaline or acid etching are essential to ensure uniform wetting by the chromating solution and a uniform rate of deposition across the surface.

Chromium phosphate conversion coatings are formed in acidic solutions containing CrO_3, F^-, and PO_4^{3-}. Like chromium oxide/hydroxide coatings, the existing surface oxide is thinned by fluoride, allowing electron tunneling and reduction of Cr(VI) to Cr(III), but in these solutions, $CrPO_4$ is deposited. The coating is duplex, having a more porous, amorphous overlayer with a dense layer at the metal interface [70]. The dense layer contains less phosphate and relatively more Cr, O, Al, and especially F, which may be mixtures of Cr_2O_3 and AlF_3 [71]. The overlayer is predominantly chromium phosphate, but also contains some Al and F [71]. In the early stage of the deposition process, small nodules are formed, which agglomerate,

forming filaments — the final texture being dependent on substrate chemistry and topography [69].

Zinc phosphate is used as a conversion coating prior to painting for parts where aluminum is combined with steel or galvanized steel [72]. These treatments differ in several fundamental ways from the chromium coatings — they are crystalline and offer less corrosion resistance than the chromium-based conversion coatings. The crystalline structure provides a barrier layer to moisture and assists the mechanical keying of the polymer coating.

It is not clear how the chromate and phosphate coatings enhance adhesion. One suggestion is that the porosity of the layer allows penetration of the polymer coating/adhesive system [69], enhancing mechanical and van der Waals interactions. Stronger bonding interactions could also take place through the surface hydroxyl groups, which will be present on the oxide/hydroxide and phosphate surfaces. Hydroxyl groups, which may be acidic or basic in nature, may participate in hydrogen bonding or acid/base interactions with the polymer, just as they have been implicated in the adsorption of Cr(VI).

5.3 Electrochemical Pretreatments: Anodizing

Anodizing refers to the electrochemical formation of an oxide layer when aluminum is anodically polarized in (usually) aqueous solution. Anodizing of aluminum alloys is widely used to form a corrosion and abrasion resistant coating in its own right, as a base for paints, organic finishes, and adhesive bonding, and in countless specialist applications. The diversity of applications is a reflection of the range of oxide microstructures, thicknesses, and compositions that can be generated. These depend on the specific anodizing procedure — voltage, amperage, solution composition, time, alloy pretreatment — giving rise to a bewildering array of anodizing systems. Reviews of the structure and formation mechanism of anodized aluminum films can be found in Ref. [73] and descriptions of specific procedures and technologies in Refs. [1,72,74]. Fortunately, the resulting oxides can be broadly divided into the non-porous, barrier oxides and the porous films. It is the latter that are used in adhesive/coating applications; the two principal procedures for producing porous oxide films are chromic acid (CAA) and phosphoric acid anodization. Anodization is used extensively in the aerospace industry to prepare aluminum for structural bonding and painting.

5.3.1 Phosphoric Acid Anodization

Phosphoric acid anodization was developed to create surfaces amenable to structural bonding in aerospace applications. The surfaces are degreased, chromic/sulfuric etched, and then anodized in 10 wt% H_2SO_4 at 10 V. Under these conditions, PAA films are typically 0.5–1 μm thick, with a thin (< 100 nm) barrier oxide adjacent to the metal underlying a distinctive pore structure [1,25,48]. This is shown schematically in Fig. 2b. Film growth takes place at the metal surface as aluminum is converted to oxide and the metal consumed. The vertical pores are hexagonal in cross-section and roughly 40 nm in width, with pore walls of 10 nm [48]. The thickness of the pore wall increases with applied voltage, thus reducing the amount of open porosity [1]. This open structure offers potential for significant mechanical interlocking and microcomposite formation, provided the adhesive/primer/coating

wets the surface and penetrates the pores. The surface morphology depends on the pretreatment used. When the surfaces are FPL pretreated, the outer surface morphology is similar to the whisker-like FPL morphology, providing additional interlocking capacity [44,48]. P2 etched specimens have been observed to have a denser, less open surface structure [44]. The oxide is amorphous Al_2O_3[25], with phosphate incorporated in the outer sections [57,75]. Davis has shown that the phosphate incorporated at the surface is in the form of a low solubility aluminum phosphate, which helps to seal the surface [76]. Continued exposure to moisture will eventually dissolve the phosphate, and bond failure is correlated with the loss of phosphate at the interface allowing moisture to hydrate the aluminum oxide to boehmite and eventually bayerite [76].

5.3.2 Chromic Acid Anodization

Chromic acid anodized oxides are formed in electrolytes containing 2.5–10 wt% CrO_3 [72]. Like PAA oxides, they are amorphous Al_2O_3, with a columnar pore structure that grows from the aluminum surface into the metal. However, they have a thicker barrier layer, are thicker overall (1.5–7 μm) [1,4,72], and denser than PAA oxide films, and few, if any, chromium ions are incorporated in the oxide [1,64]. The pore structure is less regular than for PAA oxides and the cell walls thicker [48] (Fig. 2c). The density contributes to the mechanical strength of these oxides and goes some way to slowing moisture ingress in the absence of a passive phosphate. It is expected that the less open structure hinders the interpenetration of adhesive/polymer, but this depends on the organic used — examples of microcomposite formation have been found [45]. As with PAA oxides, the surface morphology can be altered by pre- and postanodization procedures (for example, Ref. [45]).

REFERENCES

1. S. Wernick, R. Pinner, and P. Sheasby, *The Surface Treatment and Finishing of Aluminium and Its Alloys*, 5th Edn., Vol. 1, ASM International, Materials Park, OH, 1987, p. 661.
2. G. Schmiedel and G. Schubert, "Aluminium Foil Laminates for Flexible Packaging," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.
3. A. Higgins, "Adhesive Bonding of Aircraft Structures," International Journal of Adhesion and Adhesives 2000, 20, pp. 367–376.
4. J. R. Davis, ed., *Aluminum and Aluminum Alloys*, ASM Specialty Handbook, ASM International, Materials Park, OH, 1993, pp. 438–482.
5. K. Wefers and C. Misra, *Oxides and Hydroxides of Aluminum*, Aluminum Company of America, 1987, p. 92.
6. N. A. Thorne, et al., "Hydration of Oxide Films on Aluminium and Its Relation to Polymer Adhesion," Surface and Interface Analysis 1990, 16, pp. 236–240.
7. J. D. Venables, "Adhesion and Durability of Metal–Polymer Bonds — Review," Journal of Materials Science 1984, 19, pp. 2431–2453.
8. M. R. Alexander, G. E. Thompson, and G. Beamson, "Characterisation of the Oxide/Oxyhydroxide Surface of Aluminium by X-Ray Photoelectron Spectroscopy," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.
9. S. Thomas and P. M. A. Sherwood, "Valence Band Spectra of Aluminum Oxides, Hydroxides and Oxyhydroxides Interpreted by $X\alpha$ Calculations," Analytical Chemistry 1992, 64, pp. 2488–2495.

10. J. B. Metson and H. M. Dunlop, "New Insights into the Surface Chemistry of Aluminium and Its Oxides with Static TOF–SIMS and Scanning Probe Microscopies," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.

11. H. M. Dunlop, et al., "Static TOF–SIMS Study of Cluster Ion Emission from Aluminium Oxides and Hydroxides," in: *SIMS X*, J. Wiley and Sons, Muenster, 1995.

12. M. Amstutz and M. Textor, "Applications of Surface-Analytical Techniques to Aluminium Surfaces in Commercial Semifabricated and Finished Products," Surface and Interface Analysis 1992, 19, pp. 595–600.

13. D. F. Allgeyer and E. H. Pratz, "XPS Analysis of Thin Oxide Films on Chemically Treated Aluminium Alloy Surfaces Using a High-Energy Mg/Zr Anode," Surface and Interface Analysis 1992, 18, pp. 465–474.

14. B. R. Strohmeier, "An ESCA Method for Determining the Oxide Thickness on Aluminum Alloys," Surface and Interface Analysis 1990, 15, pp. 51–56.

15. M. Textor and R. Grauer, "A Photoelectron and Secondary Ion Mass Spectrometric Study of the Chemical Composition of Thermal Oxide Layers on Technically Pure Aluminium," Corrosion Science 1983, 23(1), pp. 41–53.

16. M. Textor and M. Amsutz, "Surface Analysis of Thin Films and Interfaces in Commercial Aluminium Products," Analytica Chimica Acta 1994, 297, pp. 15–26.

17. I. Olefjord and A. Karlsson, "Surface Analysis of Aluminium Foil," in: *Aluminium Technology '86*, London, 1986.

18. C. E. Moffitt, D. M. Wieliczka, and H. K. Yasuda, "An XPS Study of the Elemental Enrichment on Aluminum Alloy Surfaces from Chemical Cleaning," Surface and Coatings Technology 2001, 137(2–3), pp. 188–196.

19. A. Afseth, et al., "Influence of Heat Treatment and Surface Conditioning on Filiform Corrosion of Aluminium Alloys AA3005 and AA5754," Corrosion Science 2001, 43, pp. 2359–2377.

20. A. Afseth, et al., "Filiform Corrosion Studies of AA3005 Analogue Model Alloys," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.

21. G. D. Davis, "Surface Treatments of Aluminum and Titanium: From Basic Research to Production Failure Analysis," Surface and Interface Analysis 1991, 17, pp. 439–447.

22. G. Plassart and M. Aucouturier, "Effect of Cold Rolling on the Surface Structure and Chemistry of an Al–4.7 Wt% Mg Alloy," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.

23. A. Pocius, *Adhesion and Adhesives Technology*, Hanser-Gardner, New York, 1997, p. 279.

24. A. J. Kinloch, "The Science of Adhesion," Journal of Materials Science 1980, 15, pp. 2141–2166.

25. A. J. Kinloch, M. S. G. Little, and J. F. Watts, "The Role of the Interphase in the Environmental Failure of Adhesive Joints," Acta Materialia 2000, 48(18–19), pp. 4543–4553.

26. W. Possart and S. Dieckhoff, "Adhesion Mechanisms in a Cyanurate Prepolymer on Silicon and on Aluminium," International Journal of Adhesion and Adhesives 1999, 19(6), pp. 425–434.

27. J. Marsh, et al., "Interaction of Epoxy Model Molecules with Aluminium, Anodised Titanium and Copper Surfaces: An XPS Study," Applied Surface Science 1998, 133(4), pp. 270–286.

28. J. A. Davis and J. Hem, "The Surface Chemistry of Aluminum Oxides and Hydroxides," in: *The Environmental Chemistry of Aluminum* (G. Spositi, ed.), CRC Press, Boca Raton, 1989.

29. S. Lopez, et al., "Acid–Base Properties of Passive Films on Aluminum. I. A Photoelectrochemical Study," Journal of the Electrochemical Society 1998, 145(3), pp. 823–829.

30. S. Lopez, et al., "Acid–Base Properties of Passive Films on Aluminum. II. An X-Ray Photoelectron and X-Ray Absorption Near Edge Structure Study," Journal of the Electrochemical Society 1998, 145(3), pp. 829–834.

31. W. M. Mullins and B. L. Averbach, "The Electronic Structure of Anodised and Etched Aluminium Alloy Surfaces," Surface Science 1988, 206, pp. 52–60.

32. P. Michelin and E. Ollivier, "Paint Adhesion on Clad 2024 Aluminium Alloy," in: Aluminium Surface Science and Technology, Manchester, U.K., 2000.

33. M. H. Hansen, et al., "The Role of Corona Discharge Treatment in Improving Polyethylene–Aluminum Adhesion: An Acid–Base Perspective," TAPPI Journal 1993, 76(2), pp. 171–177.

34. L. Ulren, T. Hjertberg, and H. Ishida, "An FT–IR Study on Interfacial Interactions in Ethylene Copolymers/Aluminium Laminates in Relation to Adhesion Properties," Journal of Adhesion 1990, 31, pp. 117–136.

35. M. R. Alexander, et al., "Interaction of Carboxylic Acids with the Oxyhydroxide Surface of Aluminium: Poly(Acrylic Acid), Acetic Acid and Propionic Acid on Pseudoboehmite," Journal of Electron Spectroscopy and Related Phenomena 2001, 121(1–3), pp. 19–32.

36. D. H. Kim and H. J. Won, "Studies on Polymer–Metal Interfaces. Part 1. Comparison of Adsorption Behaviour Between Oxygen and Nitrogen Functionality in Model Copolymers onto Metal Surfaces," Polymer 1999, 40, pp. 3989–3994.

37. T. Funasaka, et al., "Adhesive Ability and Solvent Solubility of Propylene–Butene Copolymers Modified with Maleic Anhydride," International Journal of Adhesion and Adhesives 1999, 19, pp. 367–371.

38. I. Di Iorio, et al., "Plasma Treatments of Polymeric Materials and Al Alloy for Adhesive Bonding," Journal of Materials Processing Technology 1997, 68, pp. 179–183.

39. P. R. Underhill, G. Goring, and D. L. DuQuesnay, "A Study of the Deposition of 3-Glycidoxypropyltrimethoxysilane on Aluminum," International Journal of Adhesion and Adhesives 1998, 18, pp. 307–311.

40. M.-L. Abel, et al., "Evidence of Specific Interaction Between γ-Glycidoxypropyltrimethoxysilane and Oxidized Aluminium Using High-Mass Resolution ToF–SIMS," Surface and Interface Analysis 2000, 29, pp. 115–125.

41. G. D. Davis, "Surface Considerations," in: Adhesives and Sealants (H. Brinson, ed.), ASM International, 1990.

42. D. A. Dillard and D. Lefebvre, "Testing and Analysis," in: Adhesives and Sealants (H. Brinson, ed.), ASM International, 1990.

43. A. Raveh, et al., "Aging Effects of Epoxy Resin on Joining of Aluminum Plates," Journal of Adhesion 1991, 36, pp. 109–124.

44. R. P. Digby and D. E. Packham, "Pretreatment of Aluminium: Topography, Surface Chemistry and Adhesive Bond Durability," International Journal of Adhesion and Adhesives 1995, 15(2), pp. 61–71.

45. X. Zhou, G. E. Thompson, and G. Potts, "Effects of Surface Pretreatment on Adhesion and Durability of Polymer Aluminum Bonds," in: Aluminium Surface Science and Technology, Manchester, U.K., 2000.

46. G. W. Critchlow and D. M. Brewis, "Influence of Surface Macroroughness on the Durability of Epoxide–Aluminium Joints," International Journal of Adhesion and Adhesives 1995, 15, pp. 173–176.

47. O. Lunder, B. Olsen, and K. Nisancioglu, "Pre-Treatment of AA6060 Aluminium Alloy for Adhesive Bonding," International Journal of Adhesion and Adhesives 2002, 22, pp. 143–150.

48. J. D. Venables, et al., "Oxide Morphologies on Aluminum Prepared for Adhesive Bonding," Applications of Surface Science 1979, 3, pp. 88–98.

49. G. W. Critchlow and D. M. Brewis, "Review of Surface Pretreatments for Aluminium Alloys," International Journal of Adhesion and Adhesives 1996, 16(4), pp. 255–275.

50. M. Davis and D. Bond, "Principles and Practices of Adhesive Bonded Structural Joints and Repairs," International Journal of Adhesion and Adhesives 1999, 19, pp. 91–105.

51. E. W. Thrall and R. Shannon, eds., *Adhesive Bonding of Aluminum Alloys*, Marcel Dekker, New York, 1985.

52. A. F. Harris and A. Beevers, "The Effects of Grit-Blasting on Surface Properties for Adhesion," International Journal of Adhesion and Adhesives 1999, 19, pp. 445–452.

53. B. R. Strohmeier, "Improving the Wettability of Aluminum Foil with Oxygen Plasma Treatment," in: *Contact Angle, Wettability and Adhesion*, VSP, Utrecht, Netherlands, 1993.

54. A. J. Davenport, et al., "Surface Finishing of Aluminum Aerospace Alloys," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.

55. S. Ashitaka, et al., "Effects of Lead and Copper During Processing Treatments of Aluminium Capacitance Foils," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.

56. E. V. Koroleva, et al., "The Effect of Microstructure on the Surface Morphology of Dilute Binary Al–Si and Al–Mn Alloys During Alkaline Etching," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.

57. J. S. Soloman and D. E. Hanlin, "AES and SEM Characterisation of Anodized Aluminium Alloy Adherents for Adhesive Bonding Application," Applications of Surface Science 1980, 4, pp. 307–323.

58. R. L. Twite and G. P. Bierwagen, "Review of Alternatives to Chromate for Corrosion Protection of Aluminium Aerospace Alloys," Progress in Organic Coatings 1998, 33, pp. 91–100.

59. T. L. Metroke, R. L. Parkhill, and E. T. Knobbe, "Passivation of Metal Alloys using Sol–Gel–Derived Materials − A Review," Progress in Organic Coatings 2001, 41, pp. 233–238.

60. H. A. Katzman, et al., "Corrosion-Protective Coatings on Aluminum," Applications of Surface Science 1979, 2, pp. 416–432.

61. P. Campestrini, et al., "Investigation of the Chromate Conversion Coating on Al Clad 2024 Aluminium Alloy: Effect of the pH of the Chromate Bath," Electrochimica Acta 2002, 47, pp. 1097–1113.

62. G. M. Brown, et al., "The Morphology, Structure and Mechanism of Growth of Chemical Conversion Coatings on Aluminium," Corrosion Science 1992, 33(9), pp. 1371–1385.

63. G. M. Brown and K. Kobayashi, "AFM Study of the Conversion Coating Growth on AA2024-T3 Alloy," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.

64. A. E. Yaniv, et al., "Oxide Treatments of Al 1100 for Adhesive Bonding − Surface Characterization," Applications of Surface Science 1985, 20, pp. 538–548.

65. J. Wan, et al., "Chromium Valence State in Chromate Conversion Coatings on Aluminium," Physica B 1995, 208 and 209, pp. 511–512.

66. A. Hovestad, et al., "The Use of Electrochemical Measurement Techniques for the Monitoring of Conversion Processes for Aluminium," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.

67. P. Campestrini, E. P. M. van Westing, and J. H. W. de Wit, "Characterisation of Chromate Conversion Coatings on Aluminium Alloy 2024 Using EIS," in: *Aluminium Surface Science and Technology*, Manchester, U.K., 2000.

68. R. C. Furneaux, G. E. Thompson, and G. C. Wood, "An Electron Optical Study of the Conversion Coating Formed on Aluminium in a Chromate/Fluoride Solution," Corrosion Science 1979, 19, pp. 63–71.

69. J. A. Treverton, M. P. Amor, and A. Bosland, "Topographical and Surface Chemical Studies of Chromate–Phosphate Pretreatment Films on Aluminium Surfaces," Corrosion Science 1992, 33, pp. 1411–1426.
70. G. Geominne, H. Terryn, and J. Vereecken, "Characterisation of Conversion Layers on Aluminium by Means of Electrochemical Impedance Spectroscopy," Electrochimica Acta 1995, 40(4), pp. 479–486.
71. W. Vanhoolst, et al., "Study of Conversion Treatments on Aluminium by Means of Complementary Surface Analytical Techniques," in: *SIMS X*, J. Wiley and Sons, Muenster, 1995.
72. *Surface Engineering*, 10th Edn., ASM Handbook (A.I.H. Committee, ed.), Vol. 5, ASM International, Materials Park, OH, 1994.
73. G. C. Wood, "Porous Anodic Films on Aluminum," in: *Oxides and Oxide Films* (J. W. Diggle and A. K. Vijh, eds.), Marcel Dekker, New York, 1976.
74. S. Wernick, R. Pinner, and P. Sheasby, *The Surface Treatment and Finishing of Aluminium and Its Alloys*, 5th Edn., Vol. 2, ASM International, Materials Park, OH, 1987, pp. 682–728.
75. G. E. Thompson, R. C. Furneaux, and G. C. Wood, "Electron Microscopy of Ion Beam Thinned Porous Anodic Films Formed on Aluminium," Corrosion Science 1978, 18, pp. 481–498.
76. G. D. Davis, et al., "Application of Surface Behaviour Diagrams to the Study of Phosphoric Acid Anodised Aluminium," Journal of Materials Science 1982, 17, pp. 1807–1818.

13

Surface Modification

KIYOSHI FUNATANI

IMST Institute, Nagoya, Aichi, Japan

MASAYUKI YOSHIDA

Nihon Parkerizing Company Ltd., Hiratsuka, Kanagawa, Japan

YOSHIKI TSUNEKAWA

Toyota Technological Institute, Nagoya, Aichi, Japan

1   INTRODUCTION

The aluminum and alloys can offer a wide variety of mechanical, chemical, electrical properties and decorative finishes by surface modifications.

In comparison to many other metals which rely on various applied coatings for corrosion protection, many aluminum components are used without applied coating. However, there are common mechanical processes such as sand or grit blasting, scratch brushing and buffing, and following chemical finishes as various types of dip or spray coatings or treatments have been used widely.

The recent technical needs to reduce wear and friction problems use of organic or solid lubricant film coatings is becoming popular.

Anodic coatings and various type of additional finishing are used commonly to provide good protection against corrosion and are also good base for subsequent coloring by dying and paint coatings.

Electrolytic or electroless plating gained commercial use for components which need wear, seizure and corrosion resistance. While the environmental restrictions on the use of cadmium, lead and other toxic materials and hazardous waste originated from chromium plating is becoming a serious problem and enhanced the research and development of various nickel based alternative coating technologies.

Especially, groups of plating methods dispersing particles as ceramics or PTFE etc., and harder matrix by addition of phosphorus or some other alloying elements are becoming popular in various wear parts.

Thermal or explosive spray coating of finely powdered metals or ceramics can improve mechanical, chemical and physical properties to be used for elevated temperature applications.

Remelting of the surface layer of aluminum and its alloys is used to refine microstructure or to form durable alloyed layer to improve mechanical strength and wear life.

1.2   Types and Processes of Surface Modification Methods

The aluminum and alloys can offer a wide variety of mechanical, chemical, electrical properties and decorative finishes by surface modifications. In comparison to many other metals which rely on various applied coatings for corrosion protection, many aluminum components are used without applied coating. However, there are common mechanical processes such as sand or grit blasting, scratch brushing and buffing, and following chemical finishes as various types of dip or spray coatings or treatments have been used widely.

The recent technical needs to reduce wear and friction problems have been satisfied by the use of organic or solid lubricant film coatings.

Anodic coatings and various type of additional finishing are used commonly to provide good protection against corrosion and are also good base for subsequent coloring by dying and paint coatings.

Electrolytic or electroless plating gained commercial use for components which need wear, seizure and corrosion resistance. While the environmental restrictions on the use of cadmium, lead and other toxic materials and hazardous waste originated from chromium plating is becoming a serious problem. This has and enhanced the research and development of various nickel based alternative coating technologies.

Especially, groups of plating methods dispersing particles as ceramics or PTFE etc., and harder matrix by addition of phosphorus or some other alloying elements are becoming popular in various wear parts.

Thermal or explosive spray coating of finely powdered metals or ceramics can improve mechanical, chemical and physical properties to be used for elevated temperature applications.

Remelting of the surface layer of aluminum and its alloys is used to refine micro-structure or to form durable alloyed layer to improve mechanical strength and wear life.

There are various processing methods and technologies to improve the surface properties of aluminum and its alloys as Fig. 1.

1.3   Mechanical Surface Treatment and Finishes

Aluminum and aluminum alloys are cleaned and finished by various methods to enhance their appearance, functional properties or of both.
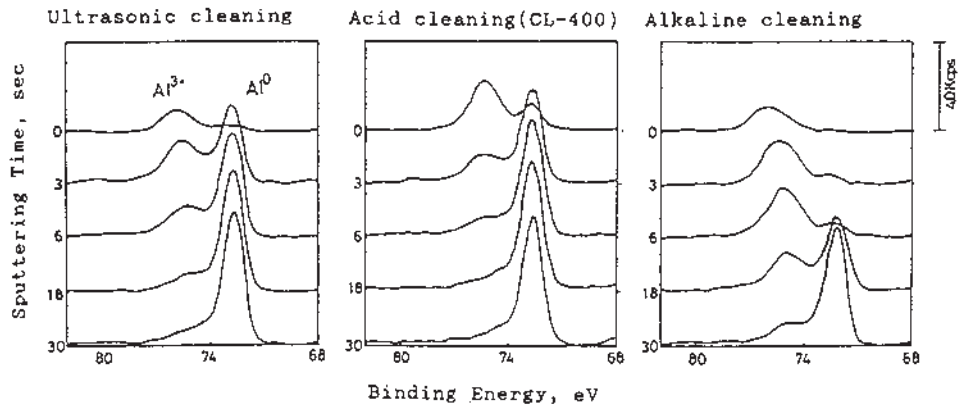
Figure 1 The influence of cleaning on surface state (Al)[2].

1.3.1 Abrasive Blast Cleaning [1]

Abrasive blasting is one of the efficient method to remove scale, sand and molding residues from cast products. It is readily adaptable to cleaning of castings, because they are usually thick enough so that no distortion results from blasting.

Blast cleaning is not recommended to finish thin sectioned parts, because such parts may warp by compressive stresses resultant from blasting. Blasting of thin sections with coarse abrasive is not recommended because the coarse abrasive can wear through the aluminum. Typical dry blasting applies pressurized air nozzle blast methods. Their blasting condition with several grit size of silica abrasives (20–over 200 grit), nozzle size (10–13 mm diameter) and pressure level (205–620 kPa) and are selected depend on application need.

Table 1 Types and Processes of Surface Modification Methods

SURFACE MODIFICATION	
METHODS	PROCESSES
Chemical conversion coatings	Alkaline chromate, Cr, Zt, Ti-chromate, Phosphating, Boehmite, Electro-phoretic chromate, etc.
Anodizing, Hard Anodizing, Coloring	Acidic (Sufuric, Chromic, Oxalic, Boric), Alkaline, Hard anodizing, Self-color, Dye, and base for coatings, etc.
Conversion plating	Zincating, Stanate process
Electroless Plating	Ni-P-SiC, Ni-P-B-PTFE, etc
Electrolytic Plating	Cu, Ni, Cr, Zn, Sn, Fe, Fe-P, or Ni-P-SiC or hBN and etc, Ni-B, Ni-W-B, etc.
Enameling and Lining	Vitreous Enameling, Glass Lining etc
Vapor deposition	PVD, CVD,
Plasma spray coatings	Fe,
Painting and Coatings	Painting, Organic spray, Solid Lubricant Coatings
Melting, Alloying and or Diffusion	Laser, Electron Beam, TIG, MIG, alloying and refining, Nitriding, Electro-Discharge alloying

In wet blasting, a slurry containing fine abrasives such as Alumina, Silica flour, Glass or Diatomite which mixed with water is blasted through nozzles and directed at the part. Abrasive grits from 220 to 5000 mesh may be used. Wet blasting is generally used for decorative applications where a fine-grain matte finish is desired. Wet blasting is used also for preparing surfaces for organic or electroplated coatings.

1.3.2   Barrel Burnishing [1]

Barrel burnishing can produce a smooth mirror-like surface on aluminum parts. Bright dipping just before burnishing and other preliminary treatments also are helpful in specific instances. One of these pretreatments to produce smooth surface on aluminum castings use etching in alkaline solution and acid cleaning to remove cast skin of the castings.

The main mechanism of barrel burnishing is to make the metal to flow, rather than to remove metal from surface by using lubricating compounds and specially designed soaps. Many of such solution have a pH of about 8, although more acidic materials can be used.

1.3.3   Polishing and Buffing [1]

In many cases aluminum parts requires polishing before final buffing because aluminum is soft and easily marred. This is also done to remove burrs, flashes or surface imperfections. Usually, buffing with a sisal wheel is commonly used before final buffing.

Polishing

Generally, polishing operations can be performed using either belts or setup wheels. For rough polishing work setup wheels are superior to belts when canvas wheels in relatively crude setup can be used. For fine polishing work, a specially contoured wheel may be more recommendable than a belt.

The typical wheel polishing by felt with Tallow grease is used to finish die cast aluminum products such as soleplates of steam irons. The medium-hard felt polishing wheel double coated with alumina abrasive bonded with hide glue is used for polishing. Also many belt polishing methods are used for bright finishing of aluminum die cast steam iron sole plates. The polishing condition such as type and hardness of wheel, and abrasive type and its mesh size of belt size are selected depending on the area and steps of polishing processes.

Buffing

Selection of buffing methods depends mainly on cost, because it is usually possible to obtain the desired results by any one of several, different methods. For hand buffing, appropriate equipment should be selected depending on types and condition of products, ranging from simple, light-duty machines to heavy-duty, variable-speed, double-control units.

Automatic buffing machines with many heads are used to bright finish die cast aluminum soleplates for steam iron. A liquid buffing compound was fed by guns to each wheels.

2 CHEMICAL PROCESSING

2.1 Pre-Treatment

The chemical conversion coating and dry-in-place coating for aluminum and its alloy are explained below. In practice, metal surfaces are treated with a cleaner or surface conditioner for the purpose of either degreasing or improving the surface condition prior to the conversion coating. The surface of most metal materials widely used in the industry is not composed of a simple single and uniform layer but is a mixture of such substances as oils, coolants, lubricants, fines, smuts, inter-metallic compounds, segregated alloy components, corrosion products and an aluminum oxide layer. The formation of conversion coating and its performance is susceptible to these substances. First, the pretreatment prior to conversion coating or dry-in-place coating is discussed.

2.1.1 Solvent Degreasing

Solvent degreasing is the most commonly used method to remove soil from the surface. Solvent degreasing with chlorinated hydrocarbons such as trichloroethylene and 1–1–1 trichlorethane is widely used in many industrial fields. The process and equipment for solvent degreasing are simple because it requires no rinsing and drying. Solvent degreasing is quite useful since the material to be degreased is not subject to corrosion by solvents. However many industrial solvents are toxic or narcotic and highly inflammable. Environmental regulations forced to keep these solvents out in use.

2.1.2 Alkaline Cleaning

Alkaline cleaners, alkaline aqueous solutions, are used for aluminum material worldwide. There are many kinds of cleaners. The merits of alkaline cleaners are low cost, easy control of solution, high degreasing ability and no corrosion of equipment. However, a subsequent water rinsing has to be employed after alkaline cleaning stage, thus requiring additional waste water treatment.

Usually, alkaline cleaners consist of an alkaline builder, chelating agent and surfactant. The performance of aluminum etching and degreasing largely depends on the alkaline builder, a major ingredient, consisting of sodium hydroxide, potassium hydroxide, potassium phosphate, potassium silicate, potassium carbonate and so on. The pH value of the alkaline cleaner solution plays an important role for aluminum etching.

The type of cleaner in use affects the surface conditioning. Figures 1 and 2 show the surface conditions of aluminum magnesium alloy analyzed by XP [2] after cleaning. The surface conditions are different by cleaner type. Aluminum hydroxide was formed and alloy components were segregated when it was cleaned with a sodium hydroxide-based cleaner. On the other hand, chelating agent such as hexagluconate prevents the formation of aluminum hydroxide and some chelating agents also remove alloy components. The surfactants also play an important role for degreasing. The determining factor in the surfactant selection is the degree of degreasing ability and the load for waste water treatment. The biodegradable property of the surfactant, the biological decomposition property, will become more important in the future.
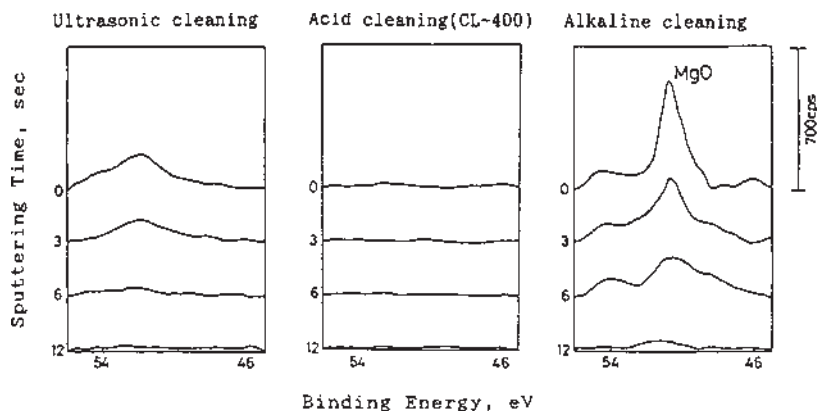
Figure 2   The influence of cleaning on surface state (Mg)[2].

2.1.3   Acid Cleaning

Acid cleaners such as sulfuric acid-based cleaners are used in the field of aluminum beverage can manufacturing. Fluoride cleaner is a typical acid cleaner and it is used worldwide. This cleaner consists of sulfuric acid, fluoride and surfactant. In contrast to alkaline cleaners, acid cleaners have some drawbacks. These include the cost, the difficulties of bath control due to the shorter life of the electrode, equipment corrosion, harder waste water treatment and severe working atmosphere due to acid mist. But, the acid cleaner provides a good clean surface, forming thinner hydroxide layer and no segregated alloy components as shown in Figs. 1 and 2. These clean surfaces provide suitable condition for chemical conversion coating and its performance.

2.1.4   Surface Conditioning

Surface conditioner is sometimes applied after alkaline cleaning in order to remove the aluminum hydroxide layer, segregated alloy components and smut. The surface conditioners are acid based solutions such as sulfuric acid or nitric acid. Figure 3 shows the effect of surface conditioning on the paint adhesion property of chromium phosphate conversion coating [3]. A peeling test was used to evaluate the paint adhesion. The paint adhesion property was improved by surface conditioning.

2.2   Chemical Conversion Coating

2.2.1   Alkaline Chromate Conversion Coating

Background

The first chemical conversion coating for aluminum was developed by Bauer and Vogel in 1915 [4]. This coating method is known as BV process. In 1930, Dr. Gustav Eckert produced the MBV (Modified Bauer Vogel). Subsequently the EW process, Pylumin process and Alork process were developed. All of them are called alkaline chromate conversion coating. However, none of these methods were used worldwide since the alkaline chromate conversion coating required relatively higher temperature treatment and longer treatment time.
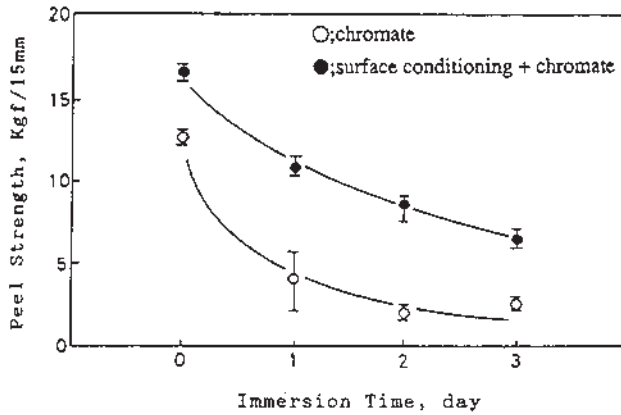
Figure 3 The effect of surface conditioning on paint adhesion[3].

Solution

The alkaline chromate conversion coating solution is made of sodium carbonate and dichromate. The treatment temperature ranges from 90°C to 100°C and treatment time is in the range of 3–20 min. The application method is only by dipping.

Coating Film

The coating film is mainly aluminum oxide with a mixture of about 30% of chromium oxide. Coating weight is in the range of 1–5 g/m^2 and the color is iridescent olive green or brownish green depending on the coating weight. The coating film provides corrosion resistance.

Reactions

The reactions are shown below. Aluminum is oxidized by this process.

$$2Al + NaCO_3 + 3H_2O \quad 2NaAlO_2 + 6H^+ 6e \tag{1}$$

$$2NaAlO_2 + 6H^+ + 6e \quad Al_2O_3 + 2NaOH \tag{2}$$

$$2Na_2CrO_4 + 6H^+ + 6e \quad Cr_2O_3 + 2Na_2O + 3H_2O \tag{3}$$
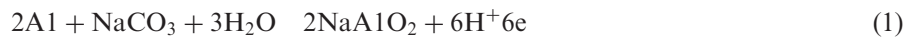
2.2.2 Chromium Chromate Conversion Coating

Background

Chromium chromate conversion coating was first introduced in the 1950s. Because this coating film provides high corrosion resistance, it is still used for construction materials, aircraft and automotive aluminum wheels. The Cr coating weight ranges normally from 100 to 300 mg/m^2. The surface color changes from a light yellow to golden yellow as the coating weight increases. Therefore, this coating is sometimes called yellow chromate. The chromium coating weight is usually measured by XRF (X-ray Fluorescence Spectroscopy) and chromium weight represents coating weight.

Solution

The chromium chromate conversion coating solution consists of chromic acid and hydrofluoric acid, usually with the addition of accelerators. The original accelerator used was potassium ferricyanide. However, due to environmental requirements, non-ferricyane chemicals were put in use and chemicals with tungsten or molybdenum accelerator have been developed.

The treatment temperature is normally from 35°C to 55°C and treatment time from 5 to 120 sec. Spray or dip methods are both possible.

Coating Film

The model of coating film which was made with ferricyanide was reported by Treverton and Davies. They analyzed coating film by XPS [5,6]. After that, some analytical works were carried out [7,8] and the depth profile by XPS is shown in Fig. 4 [9]. The model structure of current coating film, ferricyan free, is shown in Fig. 5. The compositions of coating film are mainly hydrated amorphous chromium oxide and a small amount of chromium fluoride. It is considered that about 10% of chromium is hexavalent and this hexavalent chromium provides high corrosion resistance. When tungstate is used as an accelerator, hexavalent tungstate becomes a part of the film, thus resulting in a more uniform film than a film which was made without tungstate.

Reactions

Basically, it is an oxidation-reduction reaction [Eq. (4) and Eq. (5)]. The chemical reactions are shown below. The first etching reaction is the rate-determining step of coating film formation. Chromium is reduced and deposited on the surface. The reaction ratio of CR/Al, calculated by aluminum etching weight and chromium depositing weight is about 1.1. It indicates that hexavalent chromium became a part of the film [10].
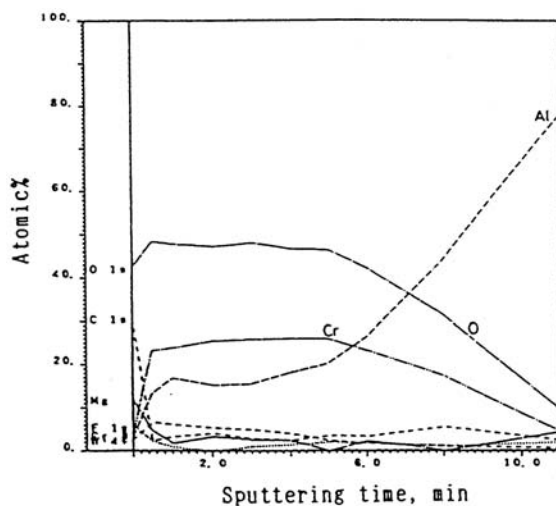
Figure 4 The depth profile of chromium chromate coating film by XPS[9].
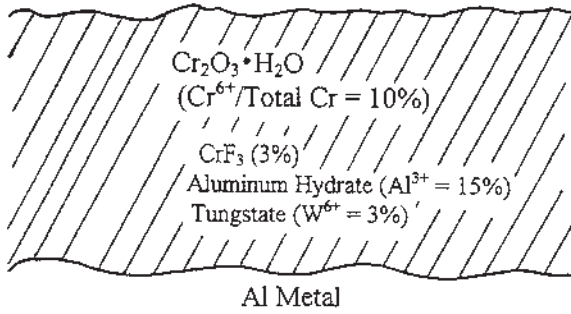
Figure 5 The structure model of chromium chromate conversion coating film on aluminum.

The tungstate accelerates film formation about 1.5 times faster, as shown in Fig. 6 and the corrosion resistance is greatly improved by the application of tungstate.

The reaction ratio without tungstate is 0.75 and it with tungstate is 1.1. This indicates that the tungstate helps the depositing on the surface. [9].

$$Al \quad Al^{3+} + 3e \tag{4}$$

$$CR^{6+} + 3e \quad Cr^{3+} \tag{5}$$

$$2Al + 6HF \quad 2AlF_3 + 6H^+ + 6e \tag{6}$$

$$H_2Cr_2O_7 + 6H^+ + 6e \quad 2CrOOH \quad + 3H_2O \tag{7}$$

$$2H^+ + 2e \quad H_2 \tag{8}$$

$$CrOOH + 3HF \quad CrF_3 + 2H_2O \tag{9}$$
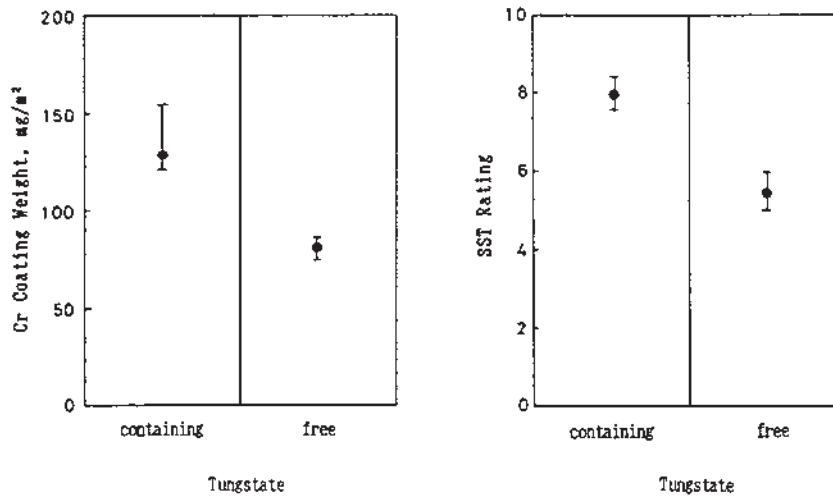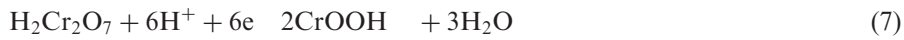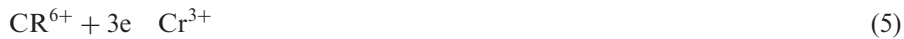
Figure 6 The effect of tungstate accelerator on chromium chromate coating weight and corrosion resistance[9]. Coating weight was controlled by treatment time. $Cr = 80$ mg/m^2.

2.2.3   Chromium Phosphate Conversion Coating

Background

The chromium phosphate conversion coating was developed in 1945 [11]. This coating film shows green color when the Cr coating weight is more than 100 mg/m^2, and is sometimes called green chromate. The coating film is mainly hydrated amorphous chromium phosphate without hexavalent chromium, so, the film is considered safe [12,13]. Therefore, this coating has been widely applied to the aluminum beverage can body and the aluminum lid of beverage can, which is generally called the aluminum end.

Solution

The chromium phosphate conversion coating solution consists of chromic acid, phosphoric acid and hydrofluoric acid. The treatment temperature is normally from 35°C to 55°C and treatment time is from 2 to 30 sec. Spray or dip methods are both possible. The aluminum coil (i.e. end stock) is only treated by spray method.

Coating Film

The compositions of coating film are mainly hydrated amorphous chromium phosphate and chromium hydroxide and a small amount of aluminum fluoride which is concentrated between coating film and aluminum substrate [6,7,14]. Model structure of coating film is shown in Fig. 7 [2]. Cr coating weight can be controlled from 10 to 50 mg/m^2. The film thickness is about 80 nano meters [15] when the Cr coating weight is 40 mg/m^2.

Reactions

The chemical reactions are shown below. The reaction is based on an oxidation-reduction reaction which is the same as chromium chromate conversion coating. The first etching reaction is the rate-determining step of coating film formation. Therefore, the coating weight depends on hydrofluoric acid concentration. Chromium is reduced and deposited on the surface as hydrated chromium phosphate.
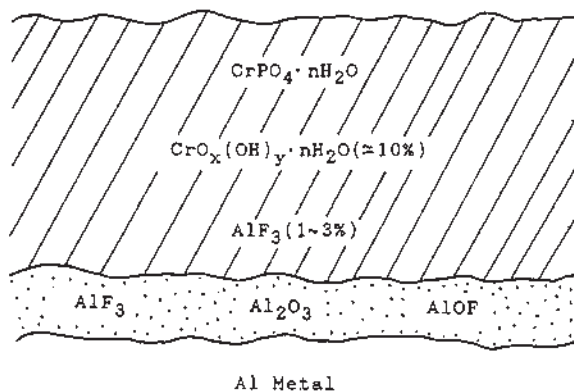
Figure 7   The structure model of chromium phosphate conversion coating film on aluminum[2].

The reaction ratio of Cr/Al, calculated by aluminum etching weight and chromium depositing weight is about 0.4. It indicates that the generation of hydrogen gas is much more than in chromium chromate conversion coating reactions.

$$2Al + 6HF \quad 2AlF_3 + 6H^+ + 6e \tag{10}$$

$$H_2Cr_2O_7 + 6H^+ + 6e \quad 2Cr(OH)3 + H_2O \tag{11}$$

$$2Cr(OH)_3 + 2H_3PO_4 \quad 2CrPO_4 + 6H_2O \tag{12}$$

$$2H^+ + 2e \quad H_2 \tag{13}$$
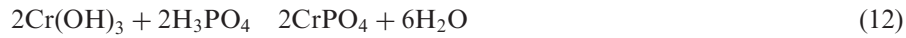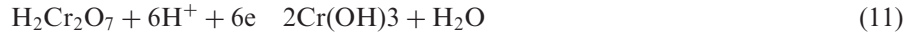
2.2.4 Zirconium Phosphate Conversion Coating

Background

The zirconium phosphate conversion coating was developed in 1976 [16]. The coating film is mainly hydrated amorphous zirconium phosphate and zirconium oxide. This coating has been applied to the aluminum beverage can body for 20 years around the world.

Solution

The zirconium phosphate conversion coating solution consists of fluoro-zirconium acid, phosphoric acid, nitric acid and hydrofluoric acid. The pH of treatment solution ranges from 2.6 to 3.3. The treatment temperature is normally from 35°C to 45°C and time is from 10 to 30 sec. The method of application for aluminum can bodies is the spray method.

Coating Film

Based on XPS analysis the coating film model is shown in Fig. 8. The composition of coating film is mainly hydrated zirconium phosphate, which is concentrated on the towered surface, and hydrated zirconium oxide and a small amount of aluminum fluoride [2].
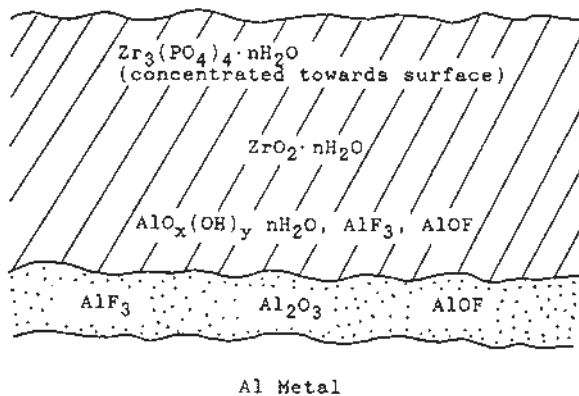
Figure 8 The structure model of zirconium phosphate conversion coating film on aluminum[2].

The coating weight can be controlled from 8 to 16 mg/m^2 as Zr content. The purpose of conversion coating for aluminum beverage cans is to provide bottom staining resistance. Good corrosion resistance in the course of sterilization process of cans contributes to achieve good bottom staining resistance. Generally, high coating weight provides better corrosion resistance. But, in the case of paint adhesion, lower coating weight shows better result. Therefore, finding the best coating weight is important issue. Actually, the Zr coating weight is limited and the control is very severe.

Reactions

The chemical reactions are shown below. Hydrolysis of zirconium acid plays an important role. The depositing reaction [Eq. (16)] caused by hydrolysis is the rate-determining step of coating film formation.

$$2Al + 6HF \quad 2AlF_3 + 3H_2 \tag{14}$$

$$2Al(OH)_3 + 6HF \quad 2AlF_3 + 6H_2O \tag{15}$$

$$H_2ZrF_6 + 2H_2O \quad ZrO_2 \quad + 6HF \tag{16}$$

$$3ZrO_2 + 4H_3PO_4 \quad Zr_3(PO_4)_4 \quad + 6H_2O \tag{17}$$
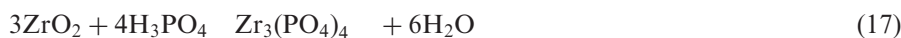
2.2.5   Titanum Phosphate Conversion Coating

Background

The titanium phosphate conversion coating was widely applied in the 1980s. The coating film is mainly hydrated amorphous titanium phosphate and titanium oxide. This coating has been applied to aluminum beverage can end stocks in the US.

Solution

The titanium phosphate conversion coating solution consists of fluoro-titanium acid, phosphoric acid and hydrofluoric acid.

Coating Film

The composition of coating films are mainly hydrated titanium phosphate and titanium oxide.

Reactions

The chemical reactions are almost the same as zirconium phosphate conversion coating.

$$2Al + 6HF \quad 2AlF_3 + 3H_2 \tag{18}$$

$$2Al(OH)_3 + 6HF \quad 2AlF_3 + 6H_2O \tag{19}$$

$$H_2TiF_6 + 2H_2O \quad TiO_2 \quad + 6HF \tag{20}$$

$$3TiO_2 + 4H_3PO_4 \quad Ti_3(PO_4)_4 \quad + 6H_2O \tag{21}$$

2.2.6 Phosphating

Background

Original phosphating was developed in the 1920s [17]. The phosphating process has been improved and it has been applied to steel substrate such as in the automotive. It is still very useful for steel. The phosphating could have been applied to aluminum, however, it was not actually applied in the industrial use because dissolved aluminum ions prevent coating film formation.

Since the 1980s, aluminum has been used for automotive parts in order to reduce the weight to improves the fuel consumption. This is particularly important because global warming poses a very serious problem for all of us. Assembled automotive bodies had to be treated by phosphating and the modified phosphating has improved for aluminum [18].

Solution

The phosphating solution for aluminum consists of phosphoric acid, metal ion, such as zinc, nickel and manganese ion, nitric acid as oxidant and hydrofluoric acid. It is almost the same for the phosphating solution used for steel materials.

In fact, the aluminum ion, which dissolve from the substrate adversely affects the phosphating bath. Therefore, sufficient fluoride must be added to keep the aluminum out of the bath solution as complex substance. If the free fluoride concentration is lower than the optimum value, the proper phosphate coating weight cannot be obtained and corrosion resistance is reduced accordingly. On the other hand, when the concentration of free fluoride is higher, the performance, especially secondary adhesion, deteriorates. This is caused by deposition of soluble salts containing Na_3AlF_6 (cryolite) together with the phosphate film.

Coating Film

The composition of coating films is mainly crystalline phosphate such as zinc phosphate. In the case of automotive, the main component is $Zn_3(PO_4)_2$ $4H_2O$ (hopeite) with about 2 g/m^2 of coating weight.

Reactions

The chemical reactions of phosphate are based on precipitation reaction. It depends on the following basic equilibrium.

$$\text{Primary Phosphate (soluble)} \quad \text{Tertiary Phosphate (insoluble)} \tag{22}$$

In the case of zinc phosphate, solution of zinc primary phosphate dissociates particularly under the influence of rise in pH to form the tertiary salts together with free phosphoric acid.

$$Zn(H_2PO_4)_2 \quad Zn_3(PO_4)_2 + 4H_3PO_4 \tag{23}$$

2.2.7 Boehmite Treatment

Background

The boehmite treatment is very simple. Aluminum materials are treated with boiling water or steam. Aluminum is converted to aluminum oxide by oxidation. It is an environmental friendly treatment as there are no specific chemicals required such

as chromium based chemicals and no need for waste water treatment. This process is applied to aluminum tableware, aluminum heat exchangers and electrolytic capacitors.

Solution

In the industry, deionized water with triethanolamine and/or aqueous ammonia are used. The treatment temperature is normally from 90°C to 100°C and treatment time is from 10 to 30 min.

Coating Film

The coating film is mainly boehmite (γ-AlOOH) and bayerite ($Al(OH)_3$). The thickness of film is from 0.2 to 2.0 μm. When the treatment temperature is below 80°C the film is mainly bayerite and its corrosion resistance is not satisfactory.

Reactions

The reactions are shown below.

$$6Al \quad 2Al^{3+} + 6e \tag{24}$$

$$6H_2O + 6e \quad 6OH^- + 3H_2 \tag{27}$$

$$2Al^{3+} + 6OH^- \quad 2AlOOH + 2H_2O \tag{26}$$

2.2.8 Electrolytic Chromate Coating

Background

One of the applications of this coating is for aluminum wheels, requiring high corrosion resistance and colorless surface. Usually, chromium chromate with lower coating weight is used in order to achieve colorless surface.

The electrolytic chromate coating was developed for aluminum wheels in 1992 [19]. This method provides good corrosion resistance and paint adhesion as a pretreatment for painting. Even though the Cr coating weight is 150 mg/m^2, the film is colorless. This is particularly suitable for clear painting.

Solution

The solution is chromium acid and sulfuric acid based. The electrolytic polarization with aluminum materials as cathode is carried out. The electric current density is from 0.5 to 15 A/dm^2 and treatment time is more than 30 sec.

Coating Film

The coating film consists of granular metallic chromium and a very thin chromate film which covers the whole surface [20].

Reactions

The chemical reactions are shown below. Chromium is reduced and deposited on the surface.

$$H_2Cr_2O_7 + 6H^+ + 6e \quad 2Cr(OH)_3 \quad H_2O \tag{27}$$

$$Cr(OH)_3 + 3H^+ + 3e \quad Cr \quad + 3H_2O \tag{28}$$
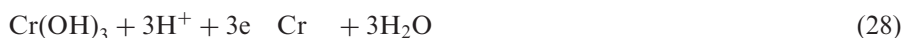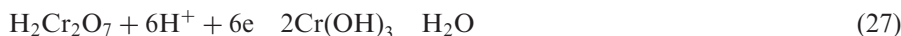
2.3 Dry-in-Place Coating

2.3.1 Chromate Coating

Background

Chromate coatings are applied to aluminum sheets or coils, which are used for aluminum fins of heat exchangers and construction materials in order to improve corrosion resistance and paint adhesion.

The application of dry-in-place process has been increasing, because it needs no water rinsing and waste water treatment.

Solution

There are a few kinds of chromate coatings. They are water based and consist of chromate and/or resins.

Process

The typical process is shown below. Usually, chromate coating is applied by roll coater. Wet coating weight of chromate coating solution is about 5 mL/m^2 and Cr coating weight is about 20 mg/m^2. After the coating, the treated materials are baked. The baking temperatures are from 150°C to 200°C and the time is from 5 to 10 sec.

Cleaning   Water Rinse   Chromate Coating   Baking

Scheme 1#Typical Process of Chromate Coating

2.3.2 Hydrophilic Coating

Background

Aluminum is used for the heat exchangers of air conditioners. Recently, reduction of evaporator size has been an important issue, especially, for automotive uses. So, reducing the size of aluminum fin pitch is important to reduce the evaporator size. Because hydrophilic coating improves the heat exchanging rate, the application of this coating has been gradually increasing.

The hydrophilic coatings are roughly divided into two types. One of them is water glass based and the other is resin based [21].

There are two processes for hydrophilic coating. They are pre-coat process and post-coat process.

Pre-Coat Process

In this process, hydrophilic coating is applied to aluminum coils or sheets. The typical process is shown below. The heat exchanger for room air conditioners is made by this process.

Post-Coat Process

In this process, hydrophilic coating is applied to heat exchanger cores. The typical process is shown below. The heat exchanger for automotive air conditioners is made by this process.

Aluminum Coil

Coil Cleaning Water Rinse Pre Treatment (i.e. chromate coating) Water Rinse Drying Hydrophilic Coating Baking *Treated Coil*

Treated Aluminum Coil

Treated Coil Applying Oil Press Forming Degreasing Welding *Heat Exchanger Core*

Scheme 2#Pre-Coat Process

Aluminum Heat Exchanger Core

Core Cleaning Water Rinse Conversion Coating Water Rinse Hydrophilic Coating Baking *Treated Core*

Scheme 3#Post-Coat Process

3 ANODIZING

The fundamental reaction in all anodizing processes is thickening of surface oxide film by converting aluminum to aluminum oxide while the part is anode in an electrolytic cell.

There are various anodizing processes and patent [22–24], using chromic acid [25,31], sulfuric acid [25–28], oxalic acid [29,30], phosphorus acid, boric acid [32] and some of those acid mixes. Among those electrolytes three typical processes are commonly used in industry. They are (1) chromic acid; (2) sulfuric acid; (3) oxalic acid; and (4) boric acid. Hard anodizing solution uses sulfuric acid alone or with additives.

Except for those produced by hard anodizing processes, most anodic coatings range in thickness from 5 to 18 μm (0.2–0.7 mm). Especially, thin film thickness anodized film are favored for pretreatment of painting and organic coatings. The particular nature of anodized film having pores in the oxidized anodic film have profound influence on processing methods, post treatment and hybrid after finishing of anodized components.

A new boric and sulfuric mixed acid process enables the reduction of toxic effluents during the anodizing process with better fatigue strength is used in aircraft industry.

3.1 Anodizing Processes

The general anodizing operations typically employed is illustrated in Fig. 9.

3.1.1 Surface Preparation

The basic requirement for successful anodizing process is to prepare chemically clean surface free of oxide. The cleaning method is selected on the basis of the type of soils or contaminants that must be removed. Usually, the cleaning step consists of removing the major organic contaminants by vapor degreasing or solvent cleaning, then making the surface chemically clean so that acid pickling or etching solutions can react uniformly over the entire surface. However, the recent environmental necessity change degreasing solutions from solvent to other alkaline degreasing methods to meet pollution regulations.

Racking → Emulsion → Rinse → Inhibited → Wash
　　　　　or Solvent　　　　　Alkaline　　↓
　　　　　Clean　　　　　　　Clean　　　Rinse

Chemical → Rinse → Desmut → Wash ← Desmut ← Chemical ← Nitric
Etch　　　(double)　　　　　　↓　　　　　　　　Electrolytic　Acid
　　　　　　　　　　　　　　Rinse　　　　　　　Brighten　　Dip

Anodizing → Wash → Rinse → Seal → Rinse → Unrack
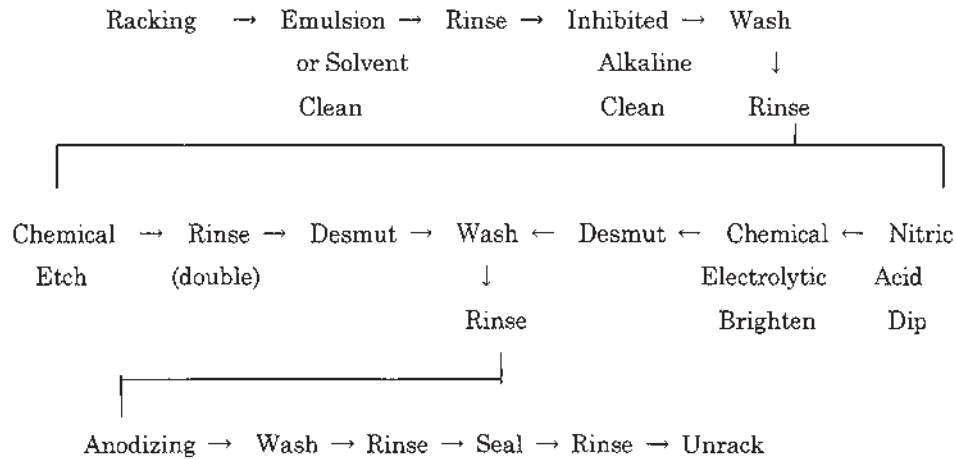
Figure 9 Anodizing operations [22].

3.1.2 Types of Anodizing Processes

There are various types of anodizing processes, the selection of anodizing process depend on types of aluminum materials or alloys, application such as decorative, protective or base for organic or coating, purpose of anodizing and components or part condition as wire, sheets or finished components.

Depend on materials and surface finish condition the anodized film defers in colors and properties caused from the condition of reaction of materials surface with treating solution, and electrical conditions. Especially, the surface reactivity with treating solution and electrolytic conditions varies largely depend on types of alloying element and their content and necessitate through studies to optimize treatment condition. Also the purpose of applying anodizing film is important to select processing methods whether it is for decorative, protective or barrier coatings.

Among many types of electrolytic anodizing methods, three typical processes are commonly used in industry. They are chromic acid, sulfuric acid and hard anodizing process.

Chromic Acid Process

The operational steps used in the chromic acid process depends on the type of part, the alloy to be anodized, and the main purpose to form anodized surface. This process is applied for part it is difficult to remove all of the electrolyte. It will color the finish from yellow to dark olive depending on film thickness. Table 2 gives typical sequence of operations that meets the requirements of military specification MIL A-8625.

Chromic acid anodizing solutions (Table 3) contain from 3 to 10 wt% CrO_3, which is made up in tank about half full of water to dissolve acid, and then adding water to adjust concentration to the specified operational condition. The Chloride contamination in the solution as NaCl should be less than 0.20 g/l and sulfate as Na_2SO_4 should not exceed 0.50/l and free chromic acid should not lower than 30 g/l. Current density should be kept under 0.1 A/dm by voltage control from

Table 2 Types of Anodizing Processes

Type	Solution	Method	Composition (%)	Current density (A/dm²)	Volts (V)	Temp. (°C)	treat. time (min.)	Color	Thickness (μm)	Application
Acidic bath [23]	Sulfuric acid	Alumilite	H_2SO_4: 10–20	DC, 1–2	10–20	20–30	10–30	transparent	5–30	Hard, protective. Seal for corrosion resist.
		Hard Anodizing	H_2SO_4: 7	DC, 2–5	23–120	-5–+5	10–60	transparent colorless	3–35	wear resistant, sealed for max. corr. resist. good base for dyes
	Oxalic acid	Almite, JAPAN	$(COOH)_2$ $2H_2O$: 2–4	AC, 1–2 DC, 0.5–1	80–120 25–30	20–29	20–60	yellow brown, half transparent	> 3	corrosion & wear resistance, decorative
		UK, U.S.A. [28]	$(COOH)_2$: 5–10	DC, 1–1.5	50–65	30	10–30	half transparent	15	corrosion resistance. decorative
		Eloxal GXh [29]	$(COOH)_2$:3–5	DC, 1–2	40–60	18–20	40–60	yellowish	10–62	
		Eloxal GXL [29] Eloxal WX [29] Eloxal WGX [29]	$(COOH)_2$:3–5 $(COOH)_2$:3–5 $(COOH)_2$:3–5	DC, 1–2 AC, 2–3 AC, 1–2 DC, 2–3	30–35 40–60 30–60 40–60	35 25–35 20–30	20–30 40–60 15–30	colorless yellowish yellowish		
	Chromic acid	Bengough-Stuart [31]	CrO_3: 2.5–3.0	DC, 0.1–0.5	0–>40 40 40–>50 50	40	10 20 5–40 5	opaque gray	2.5–15	protective decorative not fit for heavy metal alloy
		accelerated chromate	CrO_3: 5–10	DC, 0.15–0.3	40	35	30	opaque gray	2–3	not recommended for sealing
	Boric acid	condenser process [29,31]	H_3BO_3: 9–15 Borax: 0–0.25	DC DC	50–500 230–250	90–95			2.5–7.5	electrolytic condenser dielectric thin film
	Phosphoric acid	Boeing process	H_3PO_4: 10	DC,	10–12	23–25	20–30	colorless	1–2	thin film for adhesive bonding purpose
Alkaline bath [24]	Ammonia-fluoride peroxide		NH_4OH: 15	DC, 2 NH_4F :1	—	220	—	—	—	
	Sodium phosphate		NaOH, H_2O_2 Na_3PO_4 : 25	DC, 2 DC, 3	— —	20 20	30 30	transparent transparent, milky yellowish	6 7	for assembled parts good for dyes
Miscellaneous [24]	boric acid, formamide		H_3BO_3 25	DC, 1–4	—	17–15	—	dark brown	—	
	molten salt		$(NH_4)HSO_4$ $NaHSO_4$: 1:1	DC, 1	80–180	melt	—	colorless	—	

Table 3 Sequence and Solutions in Chromic Acid Anodizing

Operation	Solution	Temperature (°C)	Treatment Time (min)
Vapor degrease	Suitable solvent	R.T.	–
Alkaline clean	Alkaline cleaner	As specification	As specification
Rinse (double)	Water (running or spray)	Ambient	1
Desmut	HNO_3 (10–25 vol %)	Ambient	As required
Rinse (double)	Water (running or spray)	Ambient	1
Anodize	CrO_3 46 g/L, pH 0.5–1.0	32–35	about 30
Rinse (double)	Water (running or spray)	Ambient	1
Seal (optional)	Water (with pH control 4–6)	90–100	10–15
Drying	Air	(as specification, 150 max)	As required

Source: Ref. 22.

0 to 40 V within 5–8 min and gradually reduce to zero after the specified treating time is over. The part is removed from the bath within 15 sec, rinsed and sealed should be 200 mg/m^2 minimum [22].

Sulfuric Acid Process

The operational steps used in the sulfuric acid process are the same as for the chromic acid processes. Care should be taken when assembled or welded parts contains joints or recesses are processed so as the parts are not entrap anodizing solution.

Table 4 gives typical sequence of operations that meets the requirements of MIL-A-8625.

Depending on the type of alloy, shape and specification voltage should be accurately adjusted to produce a optimum current density over the anodizing plane.

Oxalic Acid

Oxalic acid process can produce relatively hard coatings of attractive gold or light bronze color on some types of alloys, and have advantages for anodizing wire similar as Eloxal WX process listed in Table 1. Although, chemical and power consumption of the oxalic processes are more expensive than the sulfuric acid processes, thicker

Table 4 Sulfuric Anodizing Solutions for Architectural Parts

Process Step No.	Process	Composition of Solution	Temperature (°C)	Treatment time. (min)
1	Alkaline cleaning	Alkali, inhibited	60–71	2–4
2	Alkaline etching	NaOH, 5 wt %	50–71	2–20
3	Desmutting	HNO_3, 25–35 vol. %	room	2
4	Anodizing	H_2SO_4, 15 wt %	21–25	5–60
5	Sealing	Water(pH) 5.5–6.5	100	5–20

Source: Ref. 22.

coatings is obtainable without the use of special techniques and has been used to some extent for hard anodizing. A common post treatment following anodizing are steam sealing and light buffing to finish light gold aluminum products.

Hard Anodizing

The hard anodizing produces a heavier or the more dens coating than general sulfuric acid or chromic anodizing processes in a comparable length of time by lower operating temperature (0–10°C) to prevent the formation of soft and porous outer layers, and higher current density 5(2–3.6 A/dm^2). Higher operating temperatures may cause dissolution of coating and apt to burn and damage the work. The most popular bath used for hard anodizing is sulfuric acid (5–15 vol%) at current density of about 3 A/dm^2 and lower than 5 with mild agitation [1]. Typical hard anodizing processes are as listed in Table 5.

Many other hard anodizing bath are investigated and partly used in industrial market, such as modified sulfuric acid, several mixed acids, oxalic or glyoxylic acids and Tartalic acid base electrolyte.

Hard anodizing film are used for components which require a wear or seizure resistant surface such as engine pistons, cylinders and gears of hydraulic systems.

3.2 Applications of Anodizing

The application of anodizing was started from the dielectric films produced by the boric acid process used for electrolytic capacitors, and expanded to various fields such as protective (protection against corrosion, and abrasion) use, decorative (to protect polish and other mechanical finish surfaces or to provide color), and base for organic finishes and electrolytic plating finishes. Also the natural character and properties of anodic oxide film are used for special application such as electrical and thermal insulation,and hard, seizure and wear resistant coatings. Beside the natural property of anodic oxide coatings, the morphology of oxide film with pores formed during film growth have profound effects for their use for secondary coating after anodizing. Anodic oxide film with fine pores can be a excellent base for paints, lacquers, resins, solid lubricant coating and plating.

The general classification of anodizing today are expanded to various field as shown in Table 6.

3.3 Coloring and Sealing

There are various types of coloring methods for anodized film mainly classified to natural (as anodized), electrolytic (secondary coloring), dyeing (organic dye and inorganic dye) methods.

3.3.1 Color of Anodized Films

Compared with other coloring methods the dyeing methods can produce varieties of color, but their sunshine durability need careful test before final selection

Electrolytic or as anodized methods products having better sunshine durability are more recommendable for the application of outer architectural use.

Color of anodized films treated by sulfuric and oxalic acid methods are shown in Table 7.

Table 5 Typical Hard Anodizing Processes

Processes	Composition (wt%)	Current density (A/dm^2)	Voltage (V)	Temp. (°C)	Time (min)	Color	Thickness (μm)	Application
Sulfuric Acid process	H$_2$SO: 7–20	DC, 2–4.5	200–450	23–120	> 60	gray	30′: 15 60′: 34 100′:150	wear parts, Hv450–550
Sanford Process	H$_2$SO$_4$: 7, PE:3, MA:7, etc	DC, 1–2	15–60	–10 (0±2)	–60	gray brown	60	wear parts,
Sulfuric Dicarboxylic	H$_2$SO$_4$:10–15 Dicarboxylic acid	DC, > 4	–	> 10	60	gray brown	60	usable for alloys
Oxalic Process	Oxalic acid: 3–5	AC:1–20 BC:1–20	80–200 40–60	3–5	> 60	yellow brown	≧20	high-hardness
M.H.C. Process	H$_2$SO$_4$: 15	DC, 2–2.5	25–30	0	60	gray	∞60	thick wear film
Hard Alumilite	H$_2$SO$_4$: 12 H$_2$CrO$_4$:	DC, 3.6	–	9–11	60	gray	∞60	

Source: Ref. 24.

Table 6   Applications of Anodized Components

Application	Types of anodizing methods
Rust Proof	Sulfuric, Oxalic, Chromic and Organic acids.
Wear resistance	Hard anodizing by chromic(Oxalic and Sulfuric) acids)
Ornaments	(Elect. Chemical Polish) + sulfuric acid and coloring
Optical use	(Elect. Chemical Polish) + sulfuric acid and colouring
Pre Paints	Sulfuric and Chromic acid processes
Pre Plating	Phosphoric acid process
Architecture (color)	Natural or Electrolytic coloring

Source: Ref. 24.

Table 7   Color of Anodized Film

Alloy	Type	Sulfuric Acid	Oxalic Acid
Al	1050, 1200	Silver white	Gold, yellow brown
Al-Cu	2017, 2024	Gray white	Light brown, light rose
Al-Mn	3003, 3004	Silver white, light yellow	Yellow brown
Al-Si	4043	Gray to dark gray	Gray yellow to dark G.Y.
Al-Mg	5005, 5052	Silver white, light yellow	Gold
Al-Mg-Si	6061, 6063	Silver white, light yellow	Gold
Al-Zn	772, 7075	Silver white	Gold

Source: Ref. 24.

3.3.2   Coloring

Various methods are used to color finish aluminum products used for architecture applications as shown in Table 8. Integral color anodizing are performed by using special electrolytes containing organic acids, in which the coating becomes colored by effects of some secondary anode process involving decomposition of one component of the electrolyte.

Table 8   Coloring Method and Properties of Colored Anodized Films

Type of Method	Properties						
	Color	Fading life	Corrosion	Wear resist.	Color setting	Alloy select.	Process Cost
Organic dye	◎	△	◯	◯	◎	◎	◎
In-organic dye	△	◯	◯	◯	△	◎	◯
Electrolyte with Metallic salts	△	◎	◯	◯	◯	◯	◯
Alloy (Si, Cr, Mn)	△	◎	◯	◯	◯	×	△
Organic Electrolyte	△	◎	◎	◎	◯	△	△

Source: Ref. 24.

Several types of dyeing methods are used to color finish anodized products by sulfuric acids, such as coloring by organic dyestuff, inorganic pigments, in situ forming of inorganic pigment by double decomposition and electrolyte containing metallic salts to form colored deposits. The use of special composition alloy for sulfuric acid anodizing process enables colored finishes.

3.3.3 Sealing Process

Lose of absorptive power of anodic oxide film by contact with steam was recorded in 1929 by Setoh, S. and Miyata, A [33], that the anhydrous oxide pore is hydrate and filled out by steam treatment.

The hydration and sealing increase elasticity and decrease in hardness and wear resistance, and effectively seal the surface giving more resistance to leaching or fading and the corrosion resistance enormously. Sealing is possible by many other methods as illustrated in Table 9, and many other solutions containing Ni and many other alkali metal salts and other solution such as Ammonium acetate, Lead acetate, Triethanolamine, sodium molybdate, Barium nitrate and Electrophoretic method solutions [26].

3.3.4 Properties of Anodized Film

The main composition of anodized oxide are Al_2O_3 around 60–80% and aluminum compound formed with electrolyte, Al_2O_3 H_2O and H_2O, and their properties are influenced by various factors as alloy types anodizing solution and processing variables.

Generally, the film formed by anodic oxidation are almost insoluble in the electrolyte, strongly adherent and practically non-conduction when dry. The maximum thickness depend on types of electrolyte and its ability to dissolve the oxide, operating condition such as current density, temperature and concentration. The physical properties of anodized film are as shown in Table 10, and the main effects of those factors are as shown in Table 11.

3.3.5 Specification

To ensure high quality products of anodized aluminum automotive components ASTM offers specification and SAE Information report (SAE J399) and SAE Recommended practice (SAE J1974) is published in SAE handbook as listed below [34].

Anodized Aluminum Automotive Parts- SAE J399 FEB85.
ASTM B110, Dielectric Strength of Anodically Coated Aluminum.
ASTM B457, Measuring Impedance of Anodic Coatings of Aluminum.
ASTM B244, Measuring Thickness of Anodic Coatings on Aluminum with Eddy Current Instruments.
ASTM B136, Resistance of Anodically Coated Aluminum to Staining by Dyes.
ASTM B368, Copper-Accelerated Acetic Acid Salt Spray (Fog) Testing (CASS Test).
ASTM B538, Fact (Ford Anodized Aluminum Corrosion Test) Testing.
ASTM B580, Guide to the Specification of Anodic Oxide Coating on Aluminum.
ASTM B429, Measurement and Calculation of Reflecting Characteristics of Metallic Surfaces Using Integrating Sphere Instruments.

Table 9 Sealing Method for Anodized Film

Operating Condition	Sealing methods				
	Steam	Vapor	Ni-salt	Dichromate	Na-silicate
Sealing media Composition(g/l)	Pressurized steam	Pure water	Ni-acetate:5–5.8 Co-acetate: 1, H_3BO_3: 8–84	$H_2Cr_2O_7$: 15, Na_2CO_3: 4.	Na_2SiO_3:
Ph	–	6–9	5–6	6.5–7.5	–
Temperature (°C)	(0.2–0.5 MPa)	99–100	70–90	90–95	90–100
Time (min)	15–30	16–30	15–20	2–10	20–30
Character	Best corrosion resistance	Fit for large products	Good for organic dyeing process	Good for 2000 alloy, (yellow)	Good for Alkaline attack

Source: Ref. 24.

Table 10 Physical Properties of Anodized Film

Properties	Condition	Value
Specific gravity	20 °C	2.5–3 g/cm^3
Hardness Ms	–	7–8
Hardness Hv	–	300–600
Melting point °C	–	2100°C
Specific heat	20–100°C	0.200
Thermal expansion rate	20–100°C	5×10^{-4}
Thermal conductivity	–	0.05–0.015 (C.G.S)
Electric resistance	20°C (50μm)	4×10^{17}
($\Omega \cdot$cm^2/m)	100°C (50μm)	8×10^{18}
Voltage limit (V)	5–30 μm	μm $\times$ 20–40

Source: Ref. 24.

ASTM E430, Measurement of Gloss of High Gloss Metal Surfaces Using Abridged Goniophotometer or Goniophotometer.

Decorative Anodizing Specification for Automotive Applications SAE J1974 Jun93.

ASTM publications relate with anodizing processes are:

ASTM B 137, Measurement of weight of Coating on Anodically Coated Aluminum.

ASTM B 368, Copper-Accelerated Acetic Acid-Salt Spray (Fog) Testing (CASS Test)

ASTM B 487, Measurement of Metal and Oxide Thickness by Microscopical Examination of a Cross Section.

ASTM B 680, Seal quality of Anodic Coatings on Aluminum by Acid Dissolution.

ASTM B 681, Measurement of Thickness of Anodic Coatings on Aluminum and of Other Transparent Coatings on Opaque Surfaces, Using the Light-Section Microscope.

ASTM D 2197, Standard Test Method for Adhesion of by Scrape Adhesion.

ASTM E 430, Measurement of Gloss of High Gloss Surfaces by Goniophotometry.

ASTM G 23, Operating Light-Exposure Apparatus (Carbon-Arc Type) With and Without Water for Exposure of Nonmetallic Materials.

4 PLATING ON ALUMINUM

4.1 Introduction

The application of electrolytic or electroless plating on the surface of aluminum and its alloys enable the improvement of surface properties and enhance the use of a very attractive combination of valuable properties of the light metal with highly desirable surface characteristics. A great deal of efforts were devoted in the development of efficient processes of plating aluminum. Chromium plating of aluminum has been used widely for its attractive properties as hardness, ease of fabrication, lightness, attractive finish and relative economy but the recent health and environmental issues resulting in the replacement of the process by the more environmentally friendly methods.

The low specific gravity of aluminum, approximately one third weight of metals, is an important design factor which makes it well worth while investigating the possibility of replacing zinc base diecastings and steel components with

Table 11 Factors Influence the Properties of Anodized Film

			Factors		
Property	Materials	Solution	Electrolytic Condition	Coating Thickness	Post Treatment
Corrosion	Good with Pure Al, Al-Mg, Al-Mg-Si	Oxalic, organic acids are good. Chromic is good For Alkaline corrosion	No large effect.	better with thick film.	Affected By Sealing Condition.
Wear	Good with Pure Al, Al-Mg, Al-Mg-Si,	Oxalic, organic acids are good, (low temp. elect. Sulfuric)	lower Temp. is better for hardness.	better with thick film	No large Effect.
Color	No large Effect	Sulfuric acid is good.	Higher temp. and AC are better.	better with thick film.	–
Brightness	Hi-purity Al-5N01 Are good.	Sulfuric acid is good	No large effects	better with thin films	–

Source: Ref. 24.

aluminum. However, some aluminum alloys are not easily plated, wear and seizure properties are poor, and a careful study needs to be made to substitute the base metal in a manufacturing processes.

The market of plated aluminum are expanding recently. The aluminum products are plated both by electrolytic or electroless methods. Especially, in the disc market and in fast moving machinery fields to meet the demand of weight reduction and more fuel economy and better performance of aircraft and automobile components are favoring the high strength to weight ratio of aluminum alloys.

4.1.1 Plating of Metals on Aluminum

The most commonly deposited metals on aluminum are nickel and chromium, however for some special applications it is more desirable to apply other metal deposits, because aluminum is anodic to most of the metal commonly deposited on it. Therefore, it is important to deposited pore free metal film to achieve adequate protection. For decorative use, nickel–chromium plating have been applied to various types of aluminum and its alloy products for use as in aircraft, automotive, general machinery components, and a wide variety of other industrial products.

Hard chromium have been applied to various forged, cast and wrought aluminum components, such as dies and molds, printing rollers, pistons, cylinder liners, and aircraft landing gear parts, to give increased wear resistance.

Electroless nickel (chemical nickel plating) is similarly applied to a wide variety of engineering components to improve wear and seizure properties.

Varieties of metals as tin, copper, zinc, brass, silver and nickel can be applied on aluminum to provide a solderable surface, to enable the bonding of rubber to aluminum surfaces. Tin, copper and sometimes silver deposits are used on aluminum bus bar ends to improve the electrical conductivity at the joints, while precious metals such as gold and rhodium are used on aluminum components in electrical components and lead-tin alloy is plated on bearing surface.

Cadmium plating had been widely applied in aircraft components such as switch boxes, instrument panels where electrical earthing is needed, however, the pollution and hazardous problems caused the elimination of this plating.

Various plating methods are developed such as, to copper coat aluminum for high voltage electrical installations in place of solid copper wire, and an aluminum wire with a nickel and a few other metallic coatings are developed and used for aluminum conductors.

4.1.2 Difficulties of Plating on Aluminum

The surface oxide film of aluminum and the extreme reactivity of bare aluminum which results in rapid re-formation of the film on exposure to air or to aqueous solutions, which cause the difficulty of plating on aluminum. Additionally, the following factors should be considered:

1. *Effects of aluminum oxide film*. Aluminum oxide film have amphoteric nature and cause complicates the surface reaction in either during or subsequent to preparation processes for plating.
2. *Influence of electrochemical potential*. The electrochemical position of aluminum in series of metals and alloys, and aluminum matrix and secondary phases have a profound influence on deposition reactions and

Table 12   Electrode Potential of Metals, Second Phase and Alloys

Metals	Potential (v)	Second Phase	Potential (v)	Cast and wrought Alloys	Potential (v)
Ni	−0.07	Si	−0.26	2024, Al-Cu-Mg,T4	−0.69
Ag	−0.08	NiAl$_3$	−0.52	355, LM16, Al-Si-Cu-Mg, T4	−0.78
Cu	−0.20	FeAl$_3$	−0.56	355, LM16, Al-Si-Cu-Mg, T6	−0.79
Cu-39 Zn	−0.28	CuAl$_2$	−0.73	2024, Al-Cu-Mg, T6	−0.81
Cr(plate)	−0.37	–	–	356, LM25, Al-Si-Mg	−0.82
Sn	−0.49	–	–	443, LM18, permanent mold	−0.82
Mild steel	−0.58	–	–	443, LM18, sand mold	−0.83
Cd	−0.82	Al:99.5%	−0.83	3103, Al-Mn, N3	−0.83
–	–	Al:99.95%	−0.85	7075, Al-Zn-Mg-Cu,T6	−0.83
–	–	MnAl$_6$	−0.85	3031, Al-Mn,N3	−0.83
Zn	−1.10	CuMgAl$_2$	−1.00	6063, Al-Mg-Si,H9	−0.83
–	–	MgZn$_2$	−1.05	520, LM10, Al-Mg	−0.85
Mg	−1.73	Mg$_2$Al$_3$	−1.24	5251, N4, Al-Mg	−0.85

Source: Ref. 35.

adhesion, and makes the plating of some metals or alloys quite difficult [35]. In Table 12 the electrochemical potential of metals, secondary phases and alloys are listed.

Measured electrode potentials by wrt 0.1 N KCl Calomel Electrode, in an aqueous solution of 53 g/l NaCl + 3 g/l H$_2$O$_2$ at 25.

3. *Effects of coefficient of expansion on bond strength*. The coefficient of expansion of aluminum and its alloys are quite different from that of the metals commonly deposited on it as listed in Table 13. In applications where considerable temperature changes occur, the differential expansion of aluminum and the deposited metal film may cause sufficient strain to rupture the bond between the deposit and the substrate metal [36].

4. *Effects of atomic diameter and lattice structure*. The difference in atomic diameter and crystal lattice structure between aluminum substrate and the deposited metal film. The properties of metals and alloys commonly by plated on aluminum and closeness of the fit of atoms are summarized [37].

The effect of alloy composition is mostly due to variations in the potential differences between the alloys and zinc in the alkaline solution, e.g. aluminum and the aluminum-magnesium alloys have a higher potential difference with respect to zinc than the heavy metal containing alloys, resulting in a thicker zinc deposit for the same conditions.

Table 13   Mean Coefficient of Linear Expansion (COLE) of the Common Metals per °C.

Metal	Cr	Steel	Ni	Au	Brass	Cu	Ag	Al	Sn	Zn	Cd
COLE(10^{-6})	7	11	13	14	18	18	19	24	27	27	31

Source: Ref. 36.

Table 14 Metals Able to Directly Deposit on Aluminum.

Element	Crystal Structure	Lattice Diameter	Atomic Diameter	Valence	Solubility in Al%	Plating Process	Adhesion to Aluminum	References
Al	fcc	4.0413	2.8755	3	–	–	–	–
Cu	fcc	3.607	2.551	1, 2	0, 1	I*	poor	8, 10
Ag	fcc	4.077	2.583	1	1.0	–	–	–
Au	fcc	5.070	2.878	1	nil	–	–	–
Zn	cph	6.660	2.660	2	2.0	I&E	good	11–14
Cd	cph	7.973	3.973	2	nil	I	poor	10, 15
Sn	diamond	6.476	2.80	2, 4	nil	I	fair	10, 19
Sn	tetragonal	5.830				–		
Pb	fcc	4.940	3.493	2	nil	–	–	–
Cr	bcc	2.879	2.493	3, 6	0.07	E	good	16, 17
Fe	bcc	2.861	2.477	2, 3	nil	I	poor	18
Ni	fcc	3.517	2.487	2	0.005	I	poor	17, 18
Alloys								
Cu-Zn 7/3	fcc	3.693			0.2	I	good	19
Zn-Ni 94/6	cph					I	good	20
Cu-Sn 7/3	fcc/.ognl					I	good	21
Cu-Pb 8/2	fcc					E	–	22

fcc: face centered cubic, cph: close packed hexagonal, ognl: orthogonal.
I = Immersion, E = Elector-plating.
Source: Ref. 37.

4.1.3   Types of Plating Methods for Aluminum

Various methods have been developed for deposits metals on aluminum which are roughly classified as follows:

1. Direct plating methods.
2. Processes based on zincate immersion method.
3. Processes based on stannate immersion method.
4. Processes based on mechanical or chemical preparation.
5. Processes based on preparation by anodic or chemical oxidation.

The zincate and stannate processes are more common than the other methods for plating aluminum. Especially the zinc immersion method is the most widely used process and the stannate process is recommended for specific application.

4.2   Direct Plating Methods

Many direct plating methods are used for plating on aluminum which are classified as follows:

1. Direct copper plating.
2. Direct nickel plating.
3. Direct chromium plating.
4. Direct tin plating.
5. Direct zinc plating.

4.2.1   Direct Copper Plating [39]

Direct copper plating on aluminum is possible by following methods.

1. Ammonium pyrophosphate solution [38].
   $NH_4P_2O_7$: 100 g/l, $(COOH)_2$ $2H_2O$: 60 g/l, copper oxide: 20 g/l, TEA*: 50 g/l, pH: 5.6–5.8.
   *TEA: Triethanolamine; $(HOCH_2CH_2)_3N$.
2. Copper pyrophosphate solution [39].
   CuP_2O_7: 40 g/l, 2–5 A/dm^2 (1–2 min. after immersion) at $60°C$ with agitation, after degreasing and cleaning in a solution of 3% tri-sodium phosphate and 3% soda ash at 65, followed by electrolytic plating.
3. Cyanide copper bath [40] (copper strike plating).
   CuCN: 40 g/l, NaCN: 48 g/l, Na_2CO_3: 30 g/l, Rochelle Salt $(Na_2C_4H_4O_6)$: 650 g/l temp. 40–60°C, pH: 10–10.5., live entry, 2.4 A/dm^2 (1–2 min) and 1.2 A/dm^2 (3–4 min)
4. Copper sulfate solution [41].
   $CuSO_4$: 200–220 g/l, H_2SO_4: 50–55 g/l, $CdSO_4$ $8/3H_2O$: 0.3–0.4 g/l, $SnCl_2$: 0.3–0.4 g/l, temp: 20–22°C, 1–5 A/dm^2.
5. Alkaline bath of Cu(tetraethylenepentamin)$^{2+}$ solution for plating on Al wire [42]. Cu(tetraethylenepentamin)$^{2+}$: 70–90%, $Cu(NH_3)_4)^{2+}$: 10–30%, 0.5–50 A/dm^2.

4.2.2   Direct Nickel Plating [39]

There are several direct nickel plating methods as follows.

1. Low chloride nickel plating bath with fluoride addition [43].
   $NiSO_4$ $6H_2O$: 150–250 g/l, H_3BO_3: 23–30 g/l, NaF: 2–3 g/l, NaCl: 1–2 g/l, $K_2S_2O_8$: 1–2 g/l. pH: 4–5, temp.: 50–60°C, C.D.: 1–2 A/dm^2.
2. Fluoborate bath [44,45], (for pure Al, 5052, 6061, 7075 and 336, 339, 413 etc)
   $NiSO_4$ $6H_2O$: 200 g/l, $NiBF_4$: 66 g/l, $NiCl_2$: 1 g/l, H_3BO_3: 20 g/l, pH: 1.0, temp.: 60°C, C.D. 3–4 A/dm^2.

The adhesion strength of 27–35 MN/m^2 can be doubled by heating up to 500°C and quench result from the formation of Ni_2Al_3 diffusion layer [39].

4.2.3 Direct Chromium Plating

Chromium is directly deposited from following bath (American Duramir [46,47], containing, H_2CrO_4: 240 g/l, H_2SO_4: 1.65 g/l, with special additive: 1.75 g/l, C.D.: 0.1–0.15 A/dm^2, temp.: 60°C.

This deposit film have knoop hardness 332–389 with good corrosion resistance and used under polished condition or applied for combination with bright nickel-chromium plating [39].

4.2.4 Direct Tin Plating

Direct tin plating is possible both by immersion process and electrolytic methods, however the chemical conversion coating method have considerable cost advantages. Tin coatings is used for many applications such as deposit for soldering, electrical conductivity, metal plating, corrosion resistance and to give a low friction surface to aluminum pistons for internal combustion engines.

Immersion Tin Plating [49].

STANNATE SOLUTION.   Tin film deposition by dipping method is used for aluminum alloy pistons and cylinders containing silicon. Compared with general tin plating by dipping method, the dipping into sodium stannate solution can deposit rather thick film (about 5 μm) with good adhesion and contribute to reduce friction loss and to prevent seizure problem during the running in period.

The fundamental tin plating on pistons are processed in following sequence [48,49]. This process is popularly used in machining shop to tin plate on as machined piston in machining shop floors.

There are a few other stannate solutions used for tin plating other than above mentioned solution as listed in Table 15, but are mostly used for tin platings on steels.

STANNOUS CHLORIDE BATH..   Stannous chloride solution are partly used for strike plating before Ni or other metal plating, and pretreatment of metal plating on aluminum and or post treatment. Careful selection of solution and the optimization of pretreatment condition are important to get best results [49].

STANNOUS SULFATE-FLUORIDE SOLUTION [54].   A typical promising tin plating solution contains $SnSO_4$: 161.1 g/l (stannous sulfate), HF: 48% (Hydro-fluoric acid), hydrolyzed glue, goulac (A sulfite paper pulp waste containing lignin sulfonates) and purified residue acid operated within pH 0.2–1.0 and its improvement of adhesion strength is investigated by adding various additive agents [49,54].

Table 15 The Typical Tin Plating Sequence

Processes	Solution [49]	Typical SDTO [48]
1 Degreasing	○	○
2 Caustic etching at R.T.	NaOH: 50 g/l 5 min.	(○)
3 Rinse	○	○
4 Acidic phosphate bath treatment	H_3PO_4: 60-HNO_3: 40 60–65°C × 5 min.	HNO_3 (25 vol%)
5 Rinse	○	○
6 Immersion Tin plating	(Na or K)$_2$SnO$_3$ 50–70°C	$Na_2Sn(OH)_6$ (45–70 g/l) 55°C × 1 min.
7 Washing and rinsing	○	Cold water wash and rinse
8 Hot water dip and dry	○	○

Electrolytic Plating Stannate Bath

Several types of stannate base bath (potassium or sodium stannate) are used for tin plating.

Thickness range of 0.75–7.5 μm is used for general purposes and for corrosion resistant application much thick film of 25–75 μm tin film is deposited by number 1 sodium solution listed in Table 18.

ELECTROLYTIC STANNATE BATHES.

ACIDIC TIN PLATING BATH. Thin tin film (0.25–1.5 μm) are deposited by bath as listed in Table 19. Thicker film is also possible to deposit by longer treating time. Those plating bathes are used for plating on pistons, cylinders and many other mechanical parts, wires, base for soldering, refrigerator parts.

4.2.5 Direct Zinc Plating

Electrolytic Direct Zinc Plating

One of the direct zinc plating used commercially called Vogt process is operated after pretreatment (Cathodic clean, rinse, Acid etch, Rinse, Cathodic clean, rinse) in sol-

Table 16 Stannate Solution for Immersion Tin Plating

Composition (g/l)	1 [50]	2 [50]	3 [51]
K_2SnO_3 (potassium stannate)	100	200	–
Na_2SnO_3 (sodium stannate)	–	–	20–80
$K_4P_2O_7·3H_2O$(potassium pyrophosphate)	–	–	30–120
$Zn(C_2H_3O_2)_2·2H_2O$(zinc acetate)	2	–	–
*m-cresol-sulfonic acid	33	33	–
*$K_2H_2PO_4$ (potassium di-hydrogen phosphate)	–	100	–
$(NH_4)_2C_2O_4·H_2O$(ammonium oxalate)	–	–	10–20
NaOH(sodium hydroxide)	–	–	1.5–7
Temperature (°C)	55	60	–
Time (min)	2	5–10	–

Table 17   Stannous Chloride Solution for Immersion Tin Coating

Composition (g/l)	1 [52]	2–1 [53]	2–2 [53]	3 [49]
NaOH (sodium hydroxide)	25–37.5	85.0	100	(25)
CuCl (copper chloride)	37.5	–	–	–
$SnCl_2 \cdot 2HO$ (stannous chloride)	37.5	–	28.0	100
Na_2CO_3 (sodium carbonate)	–	7.0	7.0	–
Na_2SnO_3 (sodium stannate)	–	28.0	28.0	–
$NiCl_2$ (nickel chloride)	–	10.0	7.0	–
$NiCO_3$ (nickel carbonate)	–	–	7.0	–
Rochelle salt	–	–	–	200
hydrolysed glue	–	–	–	5
$Na_2CrO_4 \cdot 10H_2O$ (sodium chromate)	**	–	–	–
Temperature (°C)	95–96	65–82	65–82	70
Treatment time (min)	–	0.5–1	0.5–1	4

ution described bellow deposit zinc film necessary for electrolytic plating of brass and nickel [40,56].

Electrodeposition of zinc : $ZnCl_2$: 0.5 g/1, NaCN : 0.5 g/1, NaOH : 10 g/1

20 sec., 0.5 A/dm^2, R.T.

This direct zinc deposit method have two demerits such as need heat treatment for diffusion of plated metals with aluminum for adhesion, and it is difficult to get enough adhesion on higher magnesium alloys. However, the adhesion problem is solved by the use of zincate processes.

4.3   Zincate Processes

Most of the electroplating on aluminum uses chemical zinc immersion process modified zincate solutions relatively simple, cheap and reliable technique results from

Table 18   Alkaline Stannate Baths for Electrolytic Tin Plating

Plating solution (g/l)	Potassium bath low concentration	Potassium bath high concentration	Sodium bath
$K \cdot Na_2SnO_3$ (potassium or sodium stannate)	100	200	100
$K \cdot NaOH$ (sodium or potassium hydroxide)	15	22	10
$CH_3COOK \cdot Na \cdot 3H_2O$ (K or Na acetate)	0–15	0–15	0–15
metal Sn	40	80	45
H_2O_2 (hydrogen peroxide)	(0.5)	(0.5)	(0.5)
Temperature (°C)	66–85	75–90	60–80
Current density (A/dm^2)	1.5–4	1.5–5	0.5–3
Voltage (V)	4–6	4–6	4–6

Source: Ref. 55.

Table 19 Acidic Tin Plating Bathes

Constituents and condition	Fluoboride		Sulfonic acid	
	Bright (g/l)	Dull (g/l)	Bright (g/l)	Dull (g/l)
$Sn(BF_4)_2$ (stannic fluoborate)	50	200	–	–
free HBF_4 (hydrogen fluoborate)	100	100	–	–
free H_3BO_3 (boric acid)	25	25	–	–
free (Sn^{2+})	20	81	–	–
$SnSO_4$ (in sulfate)	–	–	40	40
H_2SO_4 (sulfate)	–	–	100	60
cresol-sulfonate	–	–	30	40
β-naphthol	–	1	–	1
gelatin	–	6	–	2
formalin (37%, ml/l)	10	–	5	–
brightener (ml/l)	20	–	10	–
disperser	10	–	20	–
Temperature (°C)	17	20	17	20
Current density (A/dm^2)	2	3	1.5	1.5
Tin anode purity (%)	99.9	99.9	99.99	99.99

Source: Ref. 55.

various improvement work. The process is less sensitive to alloy composition than processes based on anodizing pretreatments and is cheaper than the stannate immersion process [56].

The reactions in the alkaline zincate solution are electrochemical dissolution of aluminum in alkaline solution and the deposition of zinc.

The basic reaction of aluminum and zinc in the zincate solution may be as:

$$\text{Anodic}: \quad Al + 3OH- \rightarrow Al(OH)_3 + 3e-$$
$$Al(OH)_3 \rightarrow AlO_2 - +H_2O + H^+$$

$$\text{Cathodic}: \quad H^+ + e- \rightarrow H \rightarrow 1/2H_2$$
$$Zn(OH)_4^{2-} \quad Zn^{2+} + 4OH$$
$$Zn^{2-} + 2e - \quad Zn$$
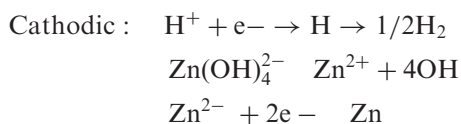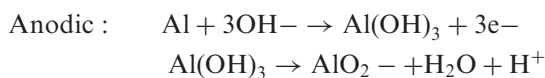
Zincate solutions are normally contain zinc oxide and caustic soda, while the ratio of these constituents, the time of immersion and temperature have large influence on the physical state of the zincate film and effect the adhesion of subsequent electroplating.

4.3.1 Types of Zincate Solutions

Concentration of the zincate solution have important effects on adhesion of deposited plate film. Especially, the higher alkali concentration result in finer deposit, and addition of small amounts of additives such as copper, nickel and other metal compounds have profound influence on the adhesion of plated film.

Table 20 Modified Alloy Zincate [59] (MAZ) and other Zincate Solutions

	Bondal process (MAZ) (g/l)	DSZ (g/l)	CSZ (g/l)
Viscosity	3.538	2.711	46.412
$Ni(SO_4)_2$ (nickel sulfate)	30	–	–
$ZnSO_4$ (zinc sulfate)	40	–	–
$NaOH$ (sodium hydroxide)	106	120	500
ZnO (zinc oxide)	–	10	100
KCN (potassium cyanide)	10	–	–
$KHC_4H_4O_6$ (potassium hydrogen tartrate)	40	–	–
$CuSO_4$ (copper sulfate)	5	–	–
$FeCl_2$ (ferric chloride)	2	–	–

DSZ: Diluted Simple Zincate, CSZ: Concentrated Simple Zincate.
Source: Ref. 60.

Careful selection of zincate solution and precise control of the solution and operating condition are the key to form thin, uniform and firmly adherent layer of zinc film and to attain high adhesion strength of plated deposits on aluminum.

Since Hewitson in 1927 [57] various zinc immersion processes with additives are investigated to improve plating adhesion by adding elements such as; copper in the form of potassium copper cyanide, ferric chloride tartrate, sodium nitrate, nickel in the form of nickel sulfate and zinc sulfate to get stable results even for the high magnesium alloys [56].

The composition of a few nominal solutions proposed by Zelley [58] contain sodium hydroxide (NaOH: 50–500 g/l), zinc oxide (ZnO: 50–100 g/l), ferric chloride (FeCl: 1–2 g/l), potassium sodium tartrate (10–50 g/l), and sodium nitrate (Na_2NO_3: 0–1 g/l) and many improved solutions have been developed. However, a double immersion processes by the use of any of such solution enables the improve-

Table 21 Relation Between Zinc Deposit and Pretreatments

		Zincate		Zinc Deposit	
Alloy	Pretreatment	time (min)	Temp (°C)	weight (g/m²)	thickness (μm)
1100, SlC	A,B,C	1	23	0.16	0.20
commercial		1/2	24	0.35	0.048
pure Al	A,C	3	24	0.59	0.084
3003, NS3	A,B,C	1	25	0.21	0.030
Al-1.25% Mn		1	24	0.51	0.074
	A,C	3	24	0.72	0.104
2024-T4	A.D,C	1	24	0.13	0.018
4.5 Cu-1.5 Mg-0.6 Mn		3	32	0.18	0.028

Weight and thickness of zinc deposit obtained on three aluminum alloys given various pretreated by various methods before zinc immersion.
Source: Ref. 6l.

Table 22   Constituents of Fluoborate and Sulfate Base Bath

Fluoborate zincating bath [65]	Sulfate bath [54]
$ZnBF_4$ (zinc fluoborate)	150 g/l, $ZnSO_4$ (zinc sulfate); 719 g/l
$NiCl_2$ (nickel fluoborate)	0.05 g/l HF/H_2O (hydrofluoric acid); 35 ml/l
NH_4Cl (ammonium chloride)	20 g/l
pH	3.0

Table 23   Types of Etching Acid Solutions

Composition	1	2	3	4	5 [67]
H_2SO_4	15%	–	–	–	150 g/l
HNO_3	50%	10 vol%	75 vol%	–	
HF/H_2O (40%)	–	–	10 vol%	20 vol%	–
H_2CrO_4 (SG: 1.42)	–	–	–	–	30 g/l
Temperature (°C)	90	–	R.T.→40	R.T.	65
Time (min)	2–5	0.5	1	3–5	3
Tank materials	lead lining	ceramics	PE or stone	PE	lead lining
Applications	castings		castings	←	
		wrought alloys	←	←	wrought alloys

Source: Ref. 57.

Table 24   Typical Copper Plating Solutions

Composition	1 (g/l)	2 (g/l)	3 (g/l)
$Cu(CN)_2$ (copper cyanide)	40	30	12–21
Total NaCN (sodium cyanide)	50	38	17–26
Free NaCN (sodium cyanide)	4 (max)	–	3–6
Na_2CO_3 (sodium carbonate)	30	38	12–25
Rochelle salt	–	50	–
Operating temperature (°C)	38–55	40	38–55

Source: Ref. 57.

Table 25   The Typical Constituents and Operating Condition of Pyrophosphate Process

Constituents		Conditions	
Pyrophosphate (as P_2O_7)	210 g/l	pH	7.5–8.5
Copper (as Cu)	30 g/l	operating temperature	40–60°C
Ammonia (as NH_3)	3 g/l	agitation	aeration

Source: Ref. 57.

Table 26 Types of Nickel Plating Solutions

Composition (g/l)	Watt	Bright nickel	Chlorine	Nickel sulfamate	Ammonium chlorine
$NiSO_4$(nickel sulfate)	240	300			150
$NiCl_2$(nickel chloride)	45	50	300	30	
$Ni(SO_3NH_2)_2 \cdot 4H_2O$ (nickel sulfamate)				300	
NH_4Cl (ammonium chloride)					15
H_3BO_3(boric acid)	30	45	30	30	15
Brightener		○			
Pit preventive agents				○	
pH	4–5	4–5	2	3.5–4.0	5.8–6.2
Temperature (°C)	40–55	40–55	55–65	30–60	20–30
Current density (A/dm)	1–8	2–8	2–15	2–15	0.65–1.0
Properties	soft film good elongation	brightness good leveling	good for strike film good cover	low film internal stress	for barrel brittle film good cover

Source: Ref. 68.

Table 27 The Constituents and Operating Condition of Nickel Acetate Bath

Constituents		Operating condition	
$CH_3COONi \cdot 4H_2O$ (nickel acetate)	65 g/l	temperature	Room temperature
H_3BO_3 (boric acid)	45 g/l	pH	6.0
Nickel glycolate	60 m l/l	Electrolytic deposition:	
Saccharin	1.5 g/l		C.D. 2.7 A/dm^2
$CH_3COONa \cdot 3H_2O$ (sodium acetate)	50 g/l	Ni or inert anode	(5 min)

Source: Ref. 57.

Table 28 Types and Properties of Composite Platings

Type of plating bath	Property of film	x: particles dispersed in solution
Watt bath; electrolytic		oxide (SiO_2, Al_2O_3, ZrO_2, WO_3, TiO_2 etc)
Sulfamate bath; electrolytic Ni-P-x,	wear resistance	carbide (SiC, CrC_3, WC, BC etc.)
Alloy composite platings	heat resistance	
Ni-B, Ni-Co-x,		solid lubricants (MoS_2, WS_2 Graphite,
Ni-W-x, etc.	self lubricant	hBN, $(CF)_n$, PTFE etc.)
Electroless plating		
nickel base or cobalt base Ni-P-x,	noncohesive	

Source: Ref. 68.

Table 29 The Typical Composition of Strike and Main Baths Conditions.

Constituents and condition	1st strike bath	2nd strike bath	Silver plate bath
AgCN (silver cyanide) (g/l)	1	5.3	30
NaCN (sodium cyanide) (g/l)	90	67.5	–
KCN (potassium cyanide (g/l)	–	–	56
free KCN	–	–	41
K_2CO_3 (potassium carbonate) (g/l)	–	–	45
Temperature (°C)	27	27	27
Current (A/dm^2)	1.5–2.5	1.5–2.5	1.5–2.4
Immersion time (sec)	10	10	(as spec.)

Source: Ref. 57.

ment of adhesion strength of casting alloys, magnesium-containing alloys as well as the Al-Zn-Cu-Mg and the Al-Cu-Mg-Mn alloys. Adhesion on high magnesium alloys is improved by increase of sodium potassium tartrate to 100 g/l in the concentrated solution [57].

4.3.2 Influence of Processing Variables

Influence of Pretreatment
The selection of pretreatment methods such as type of cleaning and etching procedures have significant influence on the state of deposit and adhesion. The morphology of zinc deposit and its adhesion results from the immersion process is influenced significantly by aluminum alloy response to the pretreatment procedure. The optimum adhesion is attainable through the set up of a particular sequence of pretreatment process adjusted for the alloy.

Influence of Alloy Elements and Secondary Phases
The influence of alloy composition and secondary phases are result from variation in

Table 30 Types of Iron Plating Solution and Operating Condition

Constituents (g/l)	Sulfuric bath	Chlorine bath	Mixed bath	Sulfamate bath
FeSO$_4$·7H$_2$O (iron sulfate)	250–300	–	250	–
FeCl$_2$·4H$_2$O (iron chloride)	–	300–550	42	–
CaCl$_2$ (calcium chloride)	–	150–335	–	–
NH$_4$Cl (ammonium chloride)	–	–	20	–
FeSO$_4$·(NH$_4$)$_2$SO$_4$·6H$_2$O (iron sulfamate)	–	–	–	250
(NH$_4$)$_2$SO$_4$ (ammonium sulfamate)	–	–	–	30
pH	< 2.5	1–2	3.5–5.5	3
Current density (A/dm^2)	4–10	2–8	2–10	2–15
Temperature (°C)	25–50	> 90	40–60	50–70

Source: Ref. 83.
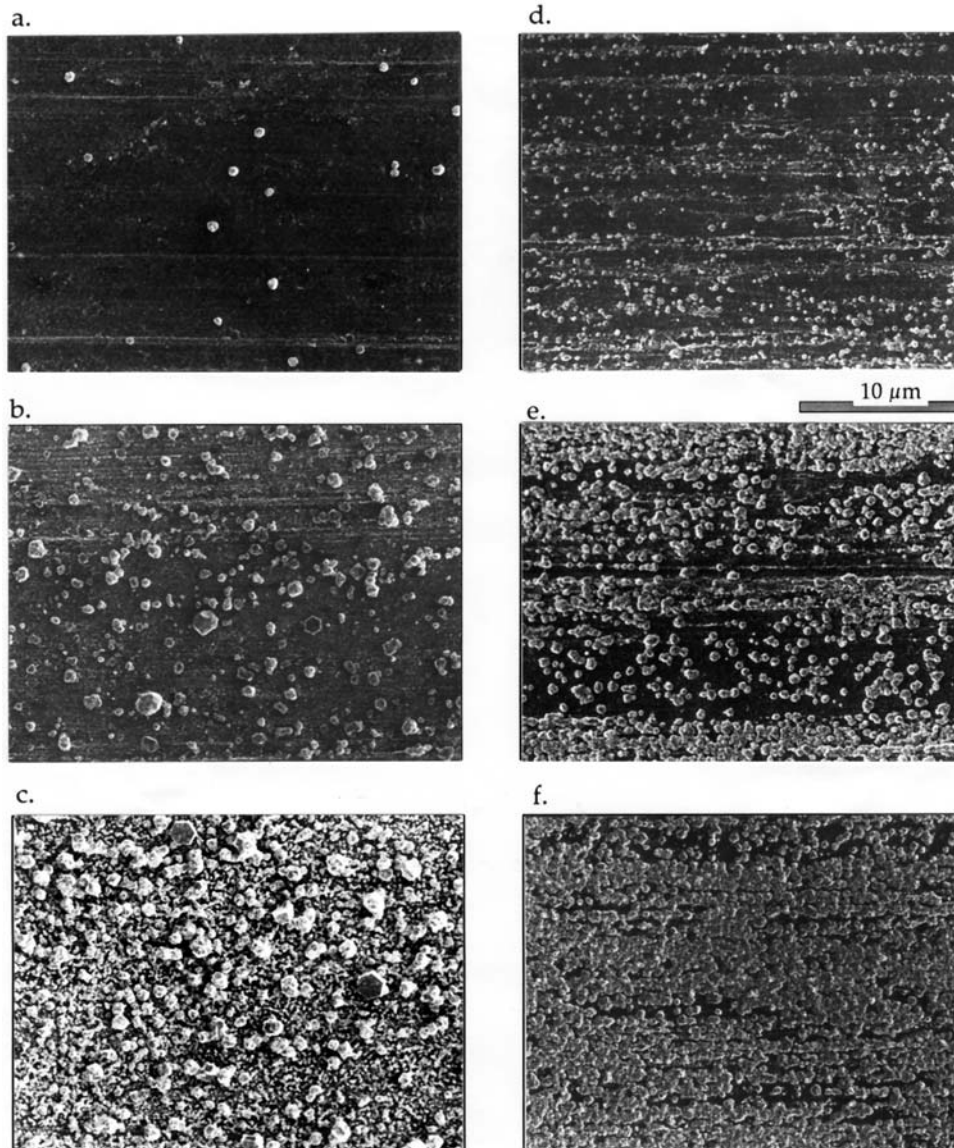
Figure 10 SEM pictures of Al during the 1st zincate process for (a) 4 s (b) 16 s (c) 30 s, and during the 2nd zincate process for (d) 1 s (e) 4 s (f) 16 s in zincate solution containing 170 g dm^{-3} NaOH, 25 g dm^{-3} ZnO and 2 g dm^{-3} $FeCl_3 \cdot 6H_2O$.

the potential differences (Table 14) between the substrate alloys and zinc in the alkaline solution – e.g. aluminum and the aluminum-magnesium alloys have a higher potential difference with respect to zinc than the heavy-metal-containing alloys – result in a thicker zinc deposit for the same conditions.
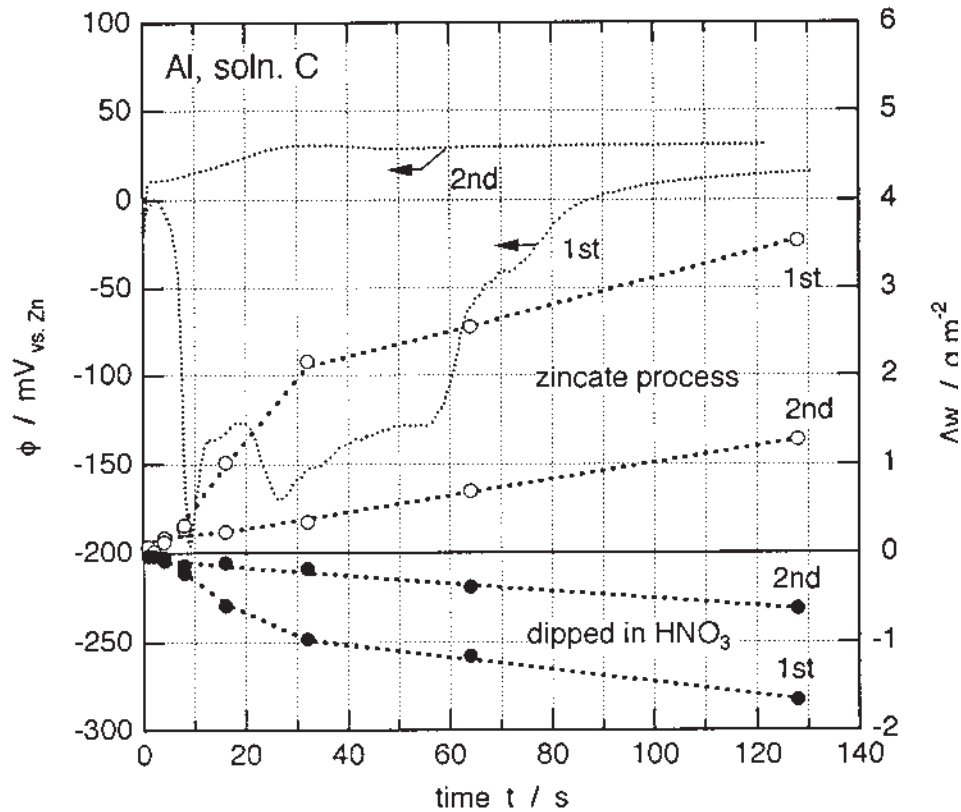
Figure 11   Change in electrode potential ϕ and mass Δw of Al during the 1st zincate process and the 2nd zincate process after the 1st zincate process for 30 s in zincate solution containing 170 g dm^{-3} NaOH, 25 g dm^{-3} ZnO, and additive of Fe^{3+}.

Influence of Temperature

The influence of operating temperature depends on the types of zincate solution and types of alloys. The operating temperature, immersion time should be optimized in conjunction with the selected types of alloys.

   The control of operating temperature is important and artificial cooling is recommended for the dissolution of aluminum in sodium hydroxide is an exothermic reaction and temperature rise progressively to reduce the bonding strength.

Influence of Contamination

Contamination from treating metal and alloys, from jigs and pretreatment bathes have profound effect on the quality of deposited zinc film, therefore, periodical chemical analysis and removal of such contamination is one of the important control measures to produce quality plate films on aluminum alloys.
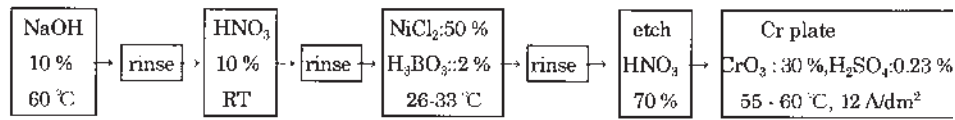
Figure 12 Typical chromium plating methods by nickel chloride process[69].

4.3.3 Double Zincate Process

The double zincating process is widely accepted for industrial plating on aluminum and its alloys to form uniform thin zinc film, in which the first nonuniform zinc deposit is stripped in nitric acid and a new thin and the more uniform zinc film deposit.

Types of Pretreatment: (A) 1 min carbonate-phosphate cleaning etch; (B) 1 min zinc immersion dip; (C) 0.5 min immersion in 1:1 nitric acid or, after zinc immersion, until zinc appears removed; (D) 5 min in 25% sulfuric acid at 80°C.

The modified zincate solution is quite effective to form thin and uniform film, which result from various additional elements contained in the MAZ as ferric chloride, Rochelle salts, where iron plays an important role in the adhesion process and a surface alloy formed in the first zincate is relatively stable even dipped into nitric acid. Figures 10 and 11 show the difference of deposits from single dip and double dip deposits from modified alloy zincate solution [28].

4.3.4 Other Zincate Processes

There are several other zincate processes, as ultrasonic zincate process, fluoride tetrahydrate type, sulfate type and fluoride tetrahydrate type bath [32] developed for special application.
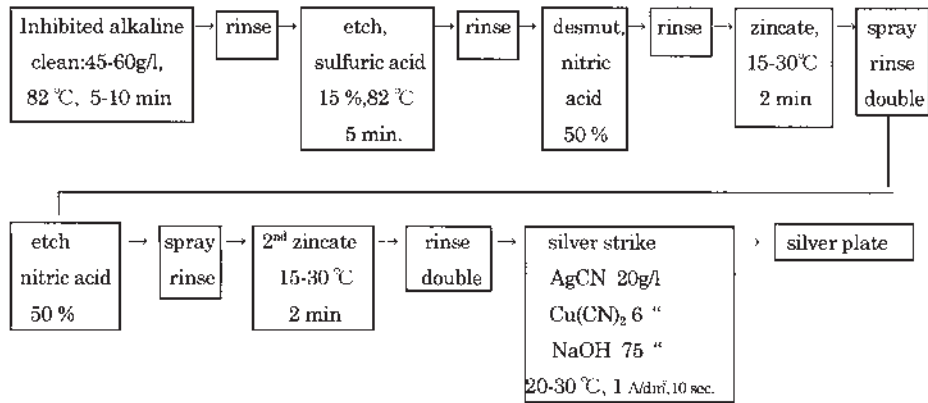
Figure 13 Processing sequence of silver plating[57].

Ultrasonic zincate process enabled the reduction of processing steps for electroplating aluminum bus bar with the zincate bath contains 125 g/l ZnO and 400 g/l NaOH and operated at 24–25°C [29,30]. Fluoborate bath is used for pretreatment of copper plating and sulfate bath is used to eliminate step for removal of oxide film.

4.4 Electroplating on the Zincate Film

4.4.1 Outline of Plating on Zincate Process

The typical processing flow sheet for plating by zincating method is as follow [23].

1. Machining or Mechanical polish
2. Degreasing by vapor degrease or solvent emulsion
3. Alkaline clean (e.g. sodium carbonate-phosphate)
4. Rinse
5. Acid etching or electrolytic or chemical polish
6. Double rinse
7. Zincate (single or double)
8. Double rinse
9. Electroplate (copper, brass, silver, chromium, nickel etc.)

Polishing

For decorative application part should be polished to the required level before plating.

Jigging

Jig design and handling are important for plating, especially the contacts should be positive enough to hold light weight aluminum. Spring contacts design and periodical maintenance to remove contaminant deposit by dipping into 50% nitric acid. For jigs made of brass, bronze or some other copper alloys, they must be stopped off adequately, not only due to the risk of corrosion but also because, if left bare, the aluminum may be anodic in the zincate solution.

Cleaning

To obtain a consistent and uniform deposit, parts surface should be clean and active for the zincate treatment. The necessity of cleaning operation depend on the degree of contamination, but in any case, it is normally essential to precede the remainder of the sequence with an alkaline cleaner. The details of an alkaline cleaner may vary as explained in Sec. 2.1.

Acid Etch

Acid etching have plural purposes to remove micro constituents from the surface and to remove surface oxide film.

Zincating

The zincate treatment can be carried out in solutions of various compositions, a number of which have been explained before in Sec. 4.3.1. Agitation is important to ensure uniform contact of the whole surface area with the solution.

For the high silicon casting alloys, a short zincate time of 5–10 sec may be suitable [57]. The components should be prevented from touching any of the steel tank or hunger as the aluminum may become anodic, impedes zinc deposition. The color of zinc deposit on every part and components should be uniformly gray.

Where the double zincate treatment is used, the components or parts immersed in 50 vol% nitric acid at room temperature to remove the excessive first deposit. Subsequently it is rinsed and then treated in the secondary zincate solution to get uniform surface chemical potential.

The formation of thin and uniform zincate deposit film is essential to plate other metals by normal methods suitable for plating on zinc. The metals that can be plated direct include nickel, copper, zinc brass, cadmium, silver and chromium.

4.4.2   Copper Plating

Copper is usually plated directly on zinc film and is frequently used prior to decorative nickel-chromium plating, but zincate solution containing nickel have been used before the direct nickel plating. For plating on zinc-based diecasting parts, a cyanide copper solution has been found the more useful. There are three types of copper plating solution such as copper cyanide, sulfuric acid and pyrophosphate solutions. The typical cyanide copper solutions have wide application such as strike plating.

The work should be protected before entering the plating bath by electrical contact and high strike current density of about 2.5 A/dm^2 to avoid dissolution of the zinc film. Care should be taken to prevent the rise of pH by addition of dilute tartaric acid and high free cyanide content and high temperature which may lead to blistering of the copper deposit. A copper plating applied for aluminum alloy memory drums prior to magnetic nickel coating used for mainframe computers, a periodical reversal process has been known as the useful method.

The copper pyrophosphate solution has also proved satisfactory for plating on zincate film deposited in zincate solution containing a chelating agent and about 6 g/l copper. The typical composition and plating condition is as follow [57].

4.4.3   Nickel Plating

Nickel film have white color, hard and corrosion resistant against various chemicals. Nickel film is easily deposited on zinc film, however, it is also deposited on copper strike film. Nickel plating is one of the very useful plating methods and have various applications such as strike plating for Nickel Composite platings, gloss, half gloss, dull, black nickel, double triple nickel layer and alloyed nickel plating. Electroless nickel plating processes also have wide applications. Simultaneously with the development of electroless composite plating methods, various nickel based composite plating methods have developed and used for mechanical pistons and engine cylinder components. Many types of plating solutions are used for nickel plating such as Watt bath, bright plating bath, chlorine base bath, nickel sulfamate bath and ammonium chloride bath. For nickel strike over the zincate film nickel modified acetate solution is also used adequately.

Watt Bath [68]

The Watt bath can deposit high purity nickel film with many good properties and applicable to bright, layered nickel and composite nickel platings. Property of nickel film deposited in Watt bath is influenced by pre-treatment and level of pH. Daily control of composition, pH and metal contamination are necessary to maintain good adhesion and film properties.

Nickel Acetate Bath

This bath can be used for immersion coating and also for electrolytic deposition of zinc film.

Chloride Bath [68]

There are several chloride baths such as straight chloride and ammonium chloride baths. Chloride baths have good coverage and deposit fine grains result in flat and strong films, however, the evolution of chlorine gas and hydrochloric acid fume, and have high internal stress which causes the difficulty to deposit thick film.

Sufamate Bath [68]

The sufamate bath has higher solubility of nickel in the solution and enables operation under high current density, thick film deposit for high speed plating processes. Also the film deposited in the solution have lower internal stress compared with Watt bath. On the other hand the concentration control is a little difficult and cost are higher than that of Watt bath.

Various composite plating methods are based on the sulfamate bath and used for plating of engine components.

Double and Tri-nickel Nickel Plating [68]

Corrosion resistance of nickel film can be improved by double or tri-nickel process which contains sulfur in secondary layer to prevent the progress of corrosion pit.

Composite Nickel Platings [68]

The composite plating or dispersion plating contains fine (3–8 μm) dispersion elements to produce dull surface or to improve wear and seizure properties. For wear resistant components, high hardness ceramics fine particles such as silicon carbide (SiC), silicon nitride (Si_3N_4) or silicon dioxide (SiO_2) are dispersed. And also, dispersion of solid lubricants particles such as hexagonal boron nitride (hBN), molybdenum disulfide (WS_2) are used for coatings of engine components (see Sec. 5 Other Plating Methods, and Sec. Electroless Plating).

4.4.4 Chromium Plating

Chromium deposited in plating bath have good outlook and gloss fit for decorative applications and also have high hardness, wear and corrosion resistant properties. The hard chromium plating and high stress cracked hard chromium coating are widely used for mechanical components. Thin chromium film (0.25–0.5 μm) is deposited on nickel film is used for decorative applications. Chromium is able to deposit directly on zincate film in low temperature chromium plating bath. Among several types of solutions for chromium plating, nickel chloride process is the most widely accepted method.

A chromic acid bath containing CrO_3: 250 g/l, H_2SO_4: 1.3–2.5 g/l, or $(Cr_2(SO_4)_2$: 1.5–3 g/l) is used for hard chromium plating and by the control of operating condition of the good adhesion strength [57].

4.4.5 Zinc Plating

Zinc is platable on zincate deposit film or zinc strike film deposited in the Vogt process, either from alkaline or acid solution. A suitable alkaline zinc plating solution used is almost same to that of zincate bath contains zinc cyanide ($Zn(CN)_2$: 60 g/l, sodium cyanide (NaCN): 40 g/l, and sodium hydroxide (NaOH): 80 g/l, operated at 25–35°C with live entry at current density of 0.5–5 A/dm^2 for a specified time [57].

4.4.6 Cadmium Plating

Cadmium plating is effective to protection mixed assemblies of aluminum with heavy metals which was widely used in aircraft components. For the best results dilute strike solution (CdO; cadmium oxide 7.5 g/l, NaO; sodium oxide 60 g/l) was used for 1 min at room temperature with current density of 2.5 A/dm^2, and then transferred to the main plating bath (CdO; 26 g/l, NaO: 100 g/l) operated at 1.5–5 A/dm^2. In many cases brighteners are added to the main bath and also copper-nickel undercoat is used [23]. Recently, almost all of the cadmium plating is converted to other corrosion resistant plating methods to eliminate the toxic cadmium processing methods [57].

4.4.7 Tin Plating

Tin plating is applied for soldering which is usually plated over copper and a flash nickel deposit may be interposed to prevent too-rapid diffusion of the tin into the copper. For tin plating normal stannate bath such as stannous sulfonate, fluoroboric or sodium or potassium alkaline solutions are used [57] (see Sec. 4.4).

4.4.8 Silver Plating [57]

Silver plating is mainly used for table wares, ornaments or electrical contacts. Silver is deposited by various processes such as chemical conversion methods, electrolytic or electroless plating methods.

The types of electrolytic plating solutions are a popular cyanide bath and metal sulfonate, and should be plated for 10 sec each in two strike baths or the second strike can be used by itself after applying a copper deposit from the copper plating solution.

After the strike deposition the silver is deposited in the main silver plate solution. A recommendation for silver or copper-silver coatings are as follows.

For extreme service conditions about or thicker than 50 μm, for intermediate conditions 20–25 μm, and for the mild dry conditions and for lighting applications film thickness of about 8–12 μm are recommended selectively [57].

Aluminum bus bars made of 1100, 3003, 2024, 5052 and 6063 are silver plated to reduce contact surface resistance by the process as indicated in Table 15.

There are few other silver plating method such as:

(a) electrolytic etch in sulfuric acid and silver strike by silver nitrate and sodium cyanide bath.

(b) ultrasonic silver plating used in conjunction with ultrasonic zincate process.

4.4.9 Precious Metal Plating [57]

Usually, precious metals are plated on top of heavy metal film and are not directly deposited over zinc film. The gold and rhodium plating is deposited after copper or brass and nickel undercoat.

4.5 Other Plating Methods on Aluminum

There are several other plating methods not using zinc immersion processes as explained previously and new plating processes such as Iron plating, Nickel base composite platings which are expanding their use in industrial products.

They are some methods investigated to prepare surface fit for plating such as mechanical treatment, chemical etching and anodic or chemical oxidation methods.

Mechanical methods such as sand or glass beads blasting are used to remove surface contaminants and produce rough surfaces, but the resultant surface properties are not sufficient to use for pretreatment for plating.

4.5.1 Chemical Etching Processes

Preparation methods such as acid mixture, Trichloroacetic acid mixtures, etching solution containing heavy metals.

(a) Nitric acid (S.G. 1.42:3 vol) and hydrofluoric acid (1 vol) mixture solution used at 20–30°C to roughen the surface of aluminum to obtain adhesion by mechanical keying of the deposited film.

The immersion time in the solution should be optimized depending on the state of products. The surface etched by acid is quite active and the time between removal from etching to plating solution is critical to prevent from forming oxide film [70].

Sulfuric acid etching method followed by rinsing and chromic acid dipping is possible to prepare surface for chromium plating [71].

(b) Trichloro-acetic acid (melted crystals, 1 l and dextrin (50 g) in 150 ml of water mixtures is used at 95°C is able to prepare aluminum surface for plating nickel. However, this process has many disadvantages such as high viscosity and drag-out, objectionable fume, high affection on operator and high solution cost [72].

(c) Hydrochloric acid containing heavy metals are investigated [39–43] as listed below. FeCl 3 g, HCl 3 ml/water 100 ml. 90–95°C for nickel chromium plating on iron layer [73] HCl (S.G. 1.18, 500 ml), water 1 l, $MnSO_4$ $2H_2O$ (4 g), solution at 37°C for 15–30 sec. [74]: for sheet and wrought alloys.

(d) Al-Cu-Mg alloys: First etch; Caustic etch by 10% hot NaOH.
Second etch; $NiCl_2$ (saturated), HCl: 2%, H_3BO_3: 4%
Remove deposit by nitric acid and plate.
Al-Mg alloys: Etch; $Cu(Cl)_2$: 15%, HCl: 0.5%
Remove deposit and plate.

(e) Etch and heat treat process: First etch; HF (45–52%) 10%, 10–15 sec.
Rinse

Second etch; HF (50 vol%), $MnSO_4$ $4H_2O$ (9–10%), 0–80 sec.

(f) Many other chemical etching methods are investigated and reported that the post heat treatment is quite effective to obtain good adhesion [75].

4.5.2 Preparation by Anodic or Chemical Oxidation [75]

There are many patented processes preparing aluminum surface for plating by the use of anodic oxide coatings deposited by a.c. or d.c. currents.

(a) Oxalic acid anodizing prior to plating. Anodizing; $C_2O_4H_2$ 2HO: 3–5%, 0.5 A/dm^2, 5–50 V, 10 min. Rinse and etch by cyanide (NaCN 6%, 5 sec to 15 min) or hydrofluoric acid (0.25–1.0 (vol%), solution to form pits and pores to obtain adhesion by anchorage the deposits.

(b) Phosphoric acid anodizing methods. H_3PO_4: 250–350 g/l, time: 10–15 min, temperature: 23–32°C, current density: 1.3–1.9 A/dm^2, voltage: 10–34 V. Operating condition have to be selected to fit to the composition of alloys being treated. As this process is very sensitive to composition use of this process for die cast products is limited.

(c) Phosphoric-sulfuric acid anodizing. Anodizing prior to plating Al-Cu-Mg alloys, Anodizing solution and operating condition are as follow; H_3PO_4 (85%): 15 vol%, HSO (96%): 15 vol%, temperature: 38 ± 3°C, current density 3 A/dm^2, 13 V, time: 5 min. Work should be rinsed and quickly transferred to nickel strike plating bath and processed to following plating operation. This mixed acid anodizing method is applicable for silver, nickel, nickel-chromium, and some other plating processes.

(d) Anodizing for plating of Al-Zn-Mg alloys. Anodizing processed by diluted phosphoric acid is applied for plating of aluminum bumper operated in following solution and conditions. H_3PO_4: 5–10%, temperature: 50–60°C,current density: 5–7 A/dm^2, voltage: 60–70 V, time: 3.5–10 min. Work is rinsed by cold water and dipped in nickel plate solution for 1 min and then electric current is gradually increased from 1.5 A/dm^2 to final 6.5 A/dm^2 to get good adhesion

(e) Sulfuric acid-copper sulfate anodizing and plating. Aluminum components are anodized in the following solution and then plated in reversed current flow. H_2SO_4 (96%) 2 vol%, $CuSO_4$ 5 H_2O: 20 wt%, temperature: 30–35°C, current density 1.5–2.5 A/dm^2. Copper plating is carried out at 4.0 A/dm^2 for 2–3 min. There are modified methods and similar other methods for nickel plating.

(f) There are many other methods using anodizing to prepare for plating. They are Dual nickel for plating of small parts (Micral process) [76,77], Hard anodizing and etching methods to obtain high bonding strength [78], Anodic cleaning method for plating of wire [79] and Alkaline anodizing method used for cadmium plating [80].

4.5.3 Iron Plating

Iron plating has been used to improve wear and seizure properties of pistons and similar components coupled with mating aluminum materials.

1. General Motors have applied iron plating for pistons used for Vega engine made of A390 alloys [81]. The iron plating process was as follow:

Before applying iron plating, skirt end of pistons were buff finished to improve surface roughness.

The four layers of metals films were plated layers are coated on piston skirt area. They are composed of;

(a) Zincate layer: thin 0.1 μm zinc,
(b) Copper strike plate: 2.5 μm copper strike, $Cu(CN)_2$: 60 $\pm$ 4 g/l, KCN (free): 7.5 $\pm$ 2 g/l, Rochelltex: 5 vol %, pH: 10.0 $\pm$ 0.2, temperature: 44 $\pm$ 4°C, current density: 2–3 A/dm^2.
(c) Hard Iron plating: 18 μm, Fe (metal): 205 $\pm$ 10 g/l, FeCl: 465 $\pm$ 23 g/l, H_3BO_3: 38.8 $\pm$ 8 g/l, pH: 0.2–0.4 (electrometric at 24°C), temperature: 70 2, current density: 2–3 A/dm^2.
(d) Tin coating 2.5 μm - Sn (metal): 42 $\pm$ 4 g/l, K_2SnO_3: 110 $\pm$ 10 g/l, temperature: 74 $\pm$ 2°C, current density: 2 A/dm^2.

The first two layers were to attain good adhesion strength and the latter two films are aimed to give resistance against seizure, wear and corrosion.

2. Similar coating methods are also applied to a few other engine pistons. Honda Fiber Reinforced Cylinder Bore (FRCB) engine [82] developed iron plating method directly deposit on aluminum pistons. But most of other pistons used in aluminum cylinder blocks made of high silicon materials the iron plating process applies method to deposits iron film on zincate film. Honda investigated a improved pulse plating technology to deposits highly oriented iron crystals on skirt area of their aluminum pistons to reduce friction coefficient [82]. The composition of solution and condition of iron plating are as follow:

   plating solution; $FeSO_4$ 7 H_2O: 400 g/l, pH: 6.0,
   plating condition; temperature 50°C, current density: pulsed 4 A/dm^2.

3. There are few other types of iron plating baths such as Sulfuric iron bath, Iron Chloride bath, Mixed solution bath and Iron sulfamate bath.

4.5.4 Electrolytic Composite Plating

There are various investigation works in types of electrolytic and electroless plating methods which co-deposit dispersed particles within deposit metals. The combination of matrix metal or alloy and disperser are numerous as listed in Table 31.

Nickel base alloy and composite plating methods are used for mechanical components such as engine cylinders, pistons, golf clubs and putters, dies, guide rails etc.

There are also composite plating layers, containing more than one type of disperser. When hard particle or disperser is co-deposited in the film, post etching methods are applied to expose the head of dispersed particles to increase wear and seizure resistance.

(a) Ni-SiC composite plating process.
Nickel silicon carbide coating processes such as Nikasil [85], ELNISIL [86], and CEM [87] were investigated since in 1960s.

Nikasil process was developed by Mahle Gmbh and used for various engine cylinders since 1970s. The thermal resistance of the composite coating is improved and widely used for coating of high-performance engine cylinder bores. Depending on application and processing plant pretreatment such as condition and solution for zincate processes and strike plating methods vary. Various modified composite plating methods are developed and used for cylinder coating [88].

(b) Ni-P-SiC or BN composite plating processes.

Nickel phosphorus based plating processes containing silicon carbide or boron nitride were developed by Suzuki [89,90] to improve the thermal properties of traditional Ni-SiC coating. Phosphorus contained in nickel matrix solid solution plays a very effective role and pre-stage of Ni_3P precipitation result in hardening by heating up to $375°C$ as observed in Fig. 14 [91].

The Ni-P-SiC or BN plating are widely used for various engine cylinder bores of motorcycle, snow boat, water jet ski, portable generator, moor cutters and hand saws. Also nickel phosphorous matrix with dispersed silicon carbide or boron nitride are also applied for engine pistons. Especially, the dispersion of hexagonal boron nitride is quite effective to eliminate piston ring welding troubles and to reduce piston/cylinder friction loss [92].

The basic solution for the plating developed by Ishimori et al., contains constituents and its basic condition are as follows [89].

> $Ni(SO_3 NH_2)_2$ $4H_2O$(Nickel Sulfamate): 500 g/l, $NiCl_2$ H_2O(Nickel Chloride); 15 g/l, H_3BO_3 (Boric Acid): 45 g/l, H_3PO_2(Hypophosphite): 0.1–4.2 g/l; Sodium Sacchalin: 3 g/l.
> Disperser: Silicon Carbide (1–4.5 μm); (80–150 g/l), or BN: hex boron nitride; (80–120 g/l); pH: 4; Current Density: 20 A/dm^2; temperature: 54–60°C.

The concentration of phosphorus, silicon carbide or boron nitride are optimized to the application needs. The deposition rate of electrolytic composite plating methods are much larger than electroless plating methods. Several advanced plating methods are investigated to increase plating rate and to reduce processing time.

Table 31 Combinations of Base Metal and Disperser for Composite Plating

Matrix	Disperser
Ni	Al_2O_3, Cr_2O_3, Fe_2O_3, TiO_2, ZrO_2, ThO_2, SiO_2, CeO_2, BeO_2, MgO, CdO, Diamond, SiC, TiC, WC, VC, ZrC, Cr_3C_2, B_4C, Graphite, MoS_2, WS, BN(α, β), ZrB_2, TiN, Si_3N_4, PTFE, CaF_2, Glass, Kaolin, Micro capsule, etc.
Ni-Co	Al_2O_3, SiC, Cr_3C_2, BN
Ni-Fe	Al_2O_3, SiC, Cr_3C_2, BN
Ni-P	Al_2O_3, Cr_2O_3, TiO_2, ZrO_2, SiC, Cr_3C_2, BC, Diamond, Graphite, PTFE, BN, CaF_2, Micro capsule,
Ni-B	Al_2O_3, Cr_2O_3, SiC Cr_3C_2, Diamond
Fe	Al_2O_3, Fe_2O_3, SiC, WC, PTFE, MoS_2, (B)
Cr	Al_2O_3, CeO_2, ZrO_2, TiO_2, SiO_2, UO_2, SiC, WC, ZaB_2, TiB_2

Source: Ref. 84.

(c) Jet electrolyte flow processes [93,94].

An electrolyte jet flow methods are investigated to improve efficiency of nickel ceramics composite plating processes. The deposition rate is increased up to 13 μm/min by high velocity electrolyte jet flow and high current density. Replacement of double zincate process by high velocity electrolyte jet containing coarse ceramics particles such as SiC or Al_2O_3 seems effective and improve adhesion [94], and also by flow rate control the content of dispersed particles is adjustable to form compositionally graded deposits [93].

(d) Rapid composite plating process [95,96].

Yamaha developed a rapid plating and plating piston rods and valve stems which replaced double zincate process by rapid anodizing method and rapid Ni-P-SiC plating (10–40 μm/min) process by high flow rate plating methods.

5 THERMAL SPRAYING

5.1 Equipment and Spray Materials

Thermal spraying has been mainly applied to make repairs on worn parts of mechanical systems, however, it is expanding the application fields to the fabrication of highly functional coatings. Thermal spraying, in which there are almost no limitations regarding the selection of substrate materials, is an extremely fast deposition process for surface coatings with a thickness of over 100 mm. The deposition rate is approximately 1–10 mm/sec or even more, which is at least 10 times higher than that in conventional electroplating. The thermal spraying technology has become a center of attraction as the surface modification for mass-production aluminum alloy parts in automotive industries, and it gradually increases the practical applications competing with other surface modification processes for the definite position. The coating technology is generally required the realization of high reliability, high adhesive strength to a substrate, high material efficiency and surface accuracy, because the common demands for inexpensive and fast processing, the usage of less strategic materials, simple and easily controllable equipment are being continuously searched for along with satisfying the required surface specification such as high hardness and wear resistance.

There are a variety of spray processes classified from a thermal source viewpoint: gas flame, arc, plasma including low pressure plasma spraying (LPPS) and radio frequency (RF) plasma spraying, laser and high velocity oxygen fuel (HVOF). The wide variation in thermal sources has led to the application of spray coatings to the surface modification of aluminum alloy parts. Plasma and HVOF are typical thermal sources for spraying, as shown in Fig. 15, which are differently characterized concerning the thermal and kinetic energy. The plasma spraying provides a high flame temperature (5000–10000 K) at relatively low flame velocity (250–400 m/sec), even in LPPS, whereas the HVOF does the reverse (2500–3500 K and > 600 m/sec).

Thermal spraying is also characterized by any wire or powder-typed spray materials, from ceramic to polymeric materials including most metallic materials can be utilized. Powders with a diameter of 10–100 mm are favorable to supply for the coating execution, and they are ready to melt down in the plasma and HVOF
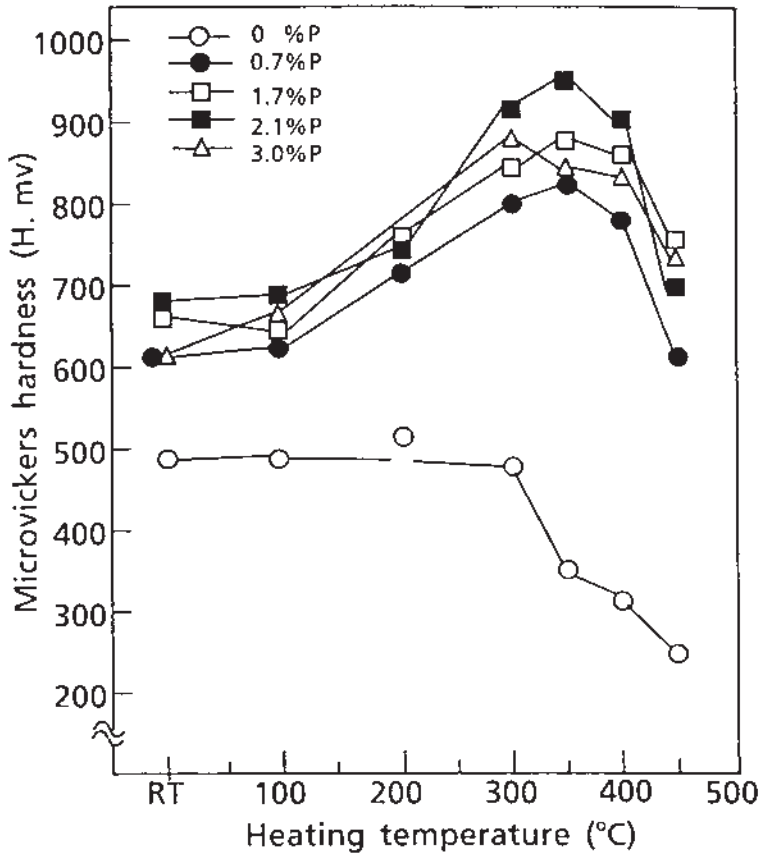
Figure 14 Changes in microvickers hardness of electroplated Ni-P films by heating.

flame within a short duration time. Molybdenum, nickel, chromium, Ni-Cr, Ni-Al powder are popular as a metallic spray material for hard-facing or undercoating. The most inexpensive powder, so a preferable spray material, is carbon steel, which

(a) plasma spray gun   (b) HVOF spray gun
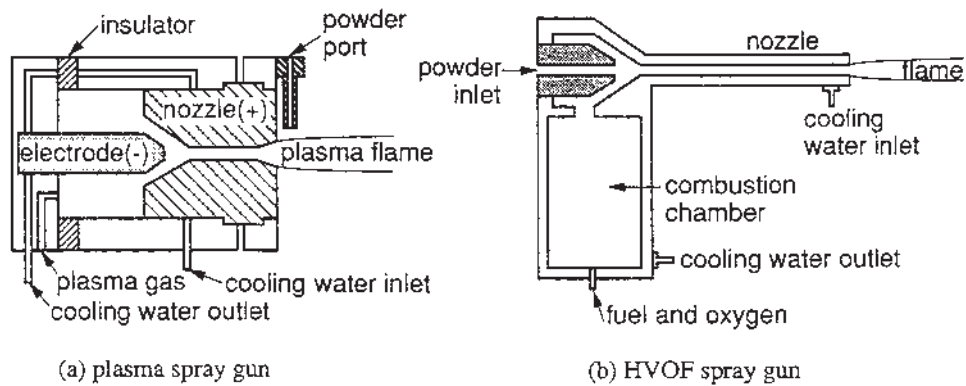
Figure 15 Schematic drawing of typical spray guns.

improves the wear resistance of aluminum alloy substrates. Ceramic powders of ZrO_2, Cr_2O_3, Cr_3C_2, TiO_2 and Al_2O_3 can be also utilized, besides WC/Co composite powder to improve the wear resistance. However, the original spray materials may change their chemical composition after spraying due to the exposure in a high temperature spray flame, for example, with WC/Co sprayed in an air atmosphere, most of the carbide grains are dissolved in cobalt and/or decomposed, which has led to porous coatings with reduced hardness [101]. It should also be noted that spray material of α-Al_2O_3 is changed to its crystalline phase to the mixture of α-Al_2O_3 and γ-Al_2O_3 in the coating sprayed in an air atmosphere.

5.2   Coating Procedure

All the coating procedures in thermal spraying should be segmented to an elemental step to understand well the whole process for the formation of highly functional coatings. The procedure of spray coating is explained as follows (Fig. 16).

(a) Preparation of spray materials, such as metal, ceramics and composite with a preferable shape and size distribution.
(b) Surface preparation of substrates, such as degreasing and roughening by blasting and substrate preheating, if necessary.
(c) Spraying: powder feeding into a high temperature and velocity zone with an appropriate feeding rate, continual impinging and depositing of molten droplets on a substrate after receiving the thermal and kinetic energy from the flame.
(d) Post-treatment of coatings, such as surface sealing and grinding.
(e) Quality control and inspection of coatings.

In order to prevent spray powder from their clogging and partial fusion, the powder size is strictly controlled within a certain range, which also affects the coating properties and spray efficiency. In addition to the direct powder formation by gas atomizing, there exist various methods to obtain the required powder size distribution, such as pelletizing, size reduction and sieving. We may also control the chemical composition of alloyed powders to obtain the required coating properties, besides the simple premixture of two or more powders, composite powders by mechanical alloying, or pelletizing and sintering from two or more kinds of powders, even metal coated ceramic powders can be prepared, although it makes the spray materials expensive.
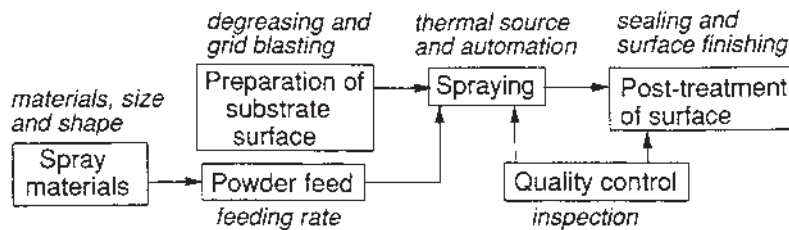
Figure 16   Processing steps in thermal spraying.

Blasting and preheating of substrates are performed to improve the poor adhesive strength between coatings and substrates, which is the principal disadvantage in thermal spraying. However, the preheating may decrease the adhesive strength due to thickening the oxide layer on a substrate surface. The adhesive behavior is understood through not only the mechanical adhesion due to an anchor effect of surface irregularity, but also the metallurgical bonding due to the mutual diffusion. Undercoating of Ni-Al, Ni-Cr or molybdenum is sometimes performed prior to the top coating to improve an adhesive strength, in spite of one of the expensive pre-treatments, which reduces the residual stress arisen at the top coating and also improve the anti-corrosion properties. The undercoating is rather effective for ceramic coatings especially. Typical adhesive strength of spray coatings is approximately 30 MPa with aluminum alloy substrates.

An analysis of splat flattening phenomena, which becomes a major research subject recently, is successfully proceeded theoretically and experimentally [102]. The main knowledge from the experimental studies is that metal splats are a regular pancake-shape without splash above the transition temperature depending on the combination of spray materials and substrates, and the adhesive strength of coatings becomes higher by preheating the substrate over the transition temperature [103,104].

The in-process structure control is also very important to form highly functional composite coatings. The most successful example is TiN/Ti in situ composite coatings fabricated by reactive LPPS [105]. The formation of TiN is occurred during the flight of molten titanium droplets in a nitrogen plasma flame along with on the substrate. The coatings strengthened with TiN exhibit much higher hardness and wear resistance than those without containing TiN [106].

Since as-sprayed coating surfaces are rough and may be porous depending on the degree of fusion and impinging velocity of droplets with a substrate, post-treatments of sealing and surface grinding after the coating process are carried out to achieve the required surface accuracy, especially in wear resistant applications. Sealing is carried out prior to the surface grinding, because of preventing spray coatings from the corrosion. It is recently shown that post-heating by laser is also effective to improve the mechanical properties of spray coatings [107].

5.3 Microstructure and Coating Properties

Microstructure of metal coatings sprayed in an air atmosphere is quite complicated even with an elemental powder, as shown in schematic cross-section of Fig. 17. Flight droplets are oxidized along with evaporation during their flight in the spray flame at their overheated molten states. Impinged droplets are immediately flattened and solidified on the substrate due to the dissipation of their thermal and kinetic energy. The shrinkage due to the solidification and also the partially molten droplets mainly cause an appearance of porosity in a coating layer. The pore ratio is usually 2 vol% or more, which worsens the mechanical properties and adhesive strength.

A typical cooling rate of splats experienced rapid solidification on the substrate is approximately 105–106 K/sec. Although the flattened splats on the substrate are annealed by the flame and following splats, the coating structures arising from splat deformation and thermal hysteresis are quite different from those in cast or wrought metals as follows:
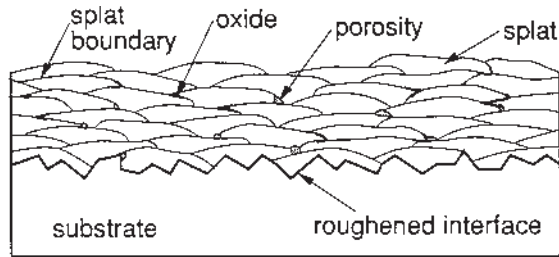
Figure 17 Schematic microstructure of cross-section of spray coating.

(a) The coating has a wavy lamellar structure composed of flattened splats, in which there exist splat boundaries.
(b) There are appreciable metal oxides and inclusions, although the oxides drastically decrease in the coatings fabricated by LPPS.
(c) There exists porosity in the coating, especially sprayed in an air atmosphere.
(d) There exists residual stress which may cause peeling off at the interface between coating and substrate, and also interlamellar of splat boundaries.
(e) The chemical composition and constituent phases of coatings may be different from those of original spray materials.

The main purpose for the application of spray coating to aluminum alloy parts is to improve their wear resistance, for example, the friction and wear properties of aluminum alloy substrates plasma-sprayed with Cr_2O_3 and WC/Co coatings were examined against the steel as a counterbody by Tronche et al. [101]. The WC/Co spray coatings by LPPS are very dense with low porosity of 2 vol% and hardness of Hv1050. The Cr_2O_3 coating sprayed in an air atmosphere is rather dense (porosity $= 2$ vol%) with hardness of Hv1500, which is lower than that of Hv1800 when sprayed at 40 kW instead of 22 kW. The hard and dense coatings, that is Cr_2O_3 by plasma spraying in an air atmosphere and WC/Co coatings by LPPS, exhibit the good frictional behavior against a steel counterbody. Although the friction coefficient is higher than 0.5, the wear rate of the two hard coatings is very low. Aluminum alloy substrates, owing to their high thermal conductivity, maintain a low surface temperature, which contribute to reduce the wear rate.

With rapidly solidified powders of Al-50 mass% Si, Al-17 Si-15 Fe and Al-50 Fe, the coatings are fabricated on aluminum alloy substrate by LPPS [108]. Dense coatings with high hardness are obtained with Al-50 Si and Al-17 Si-15 Fe powders, abrasive wear resistance of these coatings is improved. Although much higher hardness of Hv700–850 is obtained with Al-50 Fe powder, the coatings contain porosity and cracks which disturb an improvement in abrasive wear resistance.

5.4 Applications

The most attractive application of spray coatings to aluminum substrates is the surface modification of cylinder bores in automotive industries [109]. Wear resistant hypereutectic Al-Si alloy containing a small amount of copper was mainly used for aluminum cylinder blocks without surface modification. However, this alloy

exhibits poor castability and machinability, in addition to the difficulty in structure control of primary silicon phases. Even the Al-Si alloy requires its surface modification in severe wear conditions. If the sealing between the piston ring and bore is broken in the sliding pairs, the engine performance and the emission gas will be badly affected. The friction speed becomes zero at the upper point in a piston motion, at which the severest condition of the explosive pressure, temperature and lubricant thickness has led to boundary lubrication, whereas the rest friction remains in fluid lubrication. This is the reason why the wear is the worst at the upper point. In order to improve the wear resistance at the cylinder bores, the cast iron liners and also previously mentioned composite electrodeposit, composite liner have been developed, especially the latter is composed of self-lubricant graphite fibers and wear resistant Al_2O_3 fibers in aluminum alloy matrix.

For the wear resistant purpose of cylinder bores, carbon steel or low alloyed steel powder are supplied as a spray material. The optimum particle size distribution and spray parameters should be carefully selected. It has to be taken into consideration that the spray distance in HVOF is much longer than that of plasma. A short spray distance is required to coat the cylinder bores of cast Al-Si alloys, in which surface remelting is ready to occur in HVOF without an appropriate cooling system. Because of the surface coating, hypoeutectic Al-Si alloy, which has better castability and machinability compared with hypereutectic Al-Si alloy, can be used for a cylinder block itself. The process is briefly shown in Fig. 18.

After the machining of aluminum alloy block, blasting is performed to improve the adhesive strength. Then the appropriate coating sprayed beyond the final designed face on bores is performed, the final designed face is precisely machined out by grinding. It is necessary not only to limit the over trimming and over spraying as thin as possible, but also to the smooth surface roughness and good control of coating thickness.

In contrast, HVOF sprayed surface coatings have shown great promise for ferrous piston rings used for high power diesel engines, instead of coating for cylinder bores [110]. A highly dense HVOF cermet coating, $Cr_3C_2/NiCr$, has been developed with a low wearing rate and reasonable crack resistance during the service. The HVOF coating for piston rings is definitely less expensive compared with hard Cr electroplating.
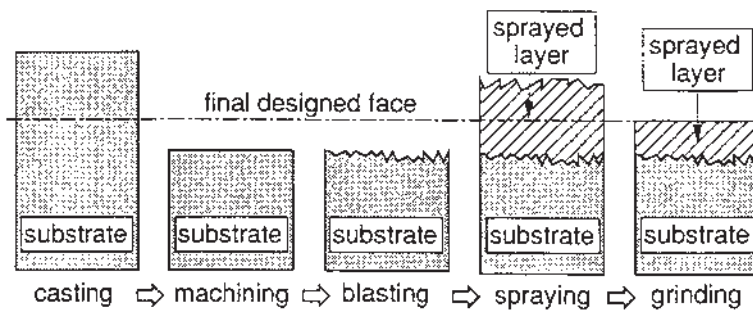
Figure 18 Schematic drawing of spray coating process on cylinder block bore.

6   SURFACE ALLOYING

6.1   Arc Welding

Thick modified layers are expected to support a high load, instead of thin but extremely hard coatings on aluminum alloy substrates. The simplest surface alloying method with a thick modified layer is remelting an aluminum alloy surface or overlaying of highly alloyed filler materials by using a conventional arc welding torch among rather thick surface modification processes. Since the thermal cracking, which is generated at the origins of porosity and/or eutectic silicon phases existed adjacent to the surface, occurs through thermal cycles in the practical uses, this process is carried out in such a way that aluminum alloy substrates are simply remelted by a TIG (Tungsten Inert Gas) arc traversed along the unidirection of certain presettled curves, and then the molten surface layer is rapidly solidified due to a self-cooling effect. The remelted and solidified layer with nearly constant thickness along the traverse direction shows a fine microstructure without containing any porosity previously existed. Then, the dendrite arm spacing, which is inversely proportional to the cooling rate, becomes from 80 mm to 10 mm after the TIG arc welding process. Although the alternating current (AC) TIG welding is known to break down the surface oxide effectively on aluminum alloys, however, the AC TIG is not as good as the DC TIG with positive polarity in electrode wear, remelted thickness and diminution in porosity, as shown in Table 32.

The mechanical properties, such as tensile strength and toughness, and also the resistance of thermal cracking in the surface modified layer can be significantly improved by the DC TIG welding process, because of the structural refinement and the diminution in porosity by the remelting and rapid solidification. Surface remelting has been introduced to the high power engine cylinder head of aluminum alloy for the partial strengthening between valve ports.

In addition to the simple remelting process, Al-Cu filler rods, which are produced by continuous casting and contain high copper contents of 25–50 mass%, can be overlaid on aluminum alloy substrates by conventional TIG welding [111]. The hardness of copper alloyed surface layers with hypoeutectic, eutectic and hypereutectic structure is Hv100–160, Hv180–220 and Hv210–330, respectively. The reason why the hardness increases with increasing copper content is explained to be due to the increase in an amount of eutectic in a hypoeutectic structure and the increase in an amount of primary $CuAl_2$ phase in a hypereutectic. The wear resistance of overlaid layer containing 30% copper increases with the increase in copper content, and it becomes saturated to approximately 3 times that of base

Table 32   Process Assessments for Surface Modification of Cast Aluminum Alloys by TIG Welding

	DC/AC	Electrode wear	Remelted thickness	Structure refinement	Porosity diminishing
DC	positive polarity	excellent	excellent	excellent	excellent
	negative polarity	poor	poor	excellent	excellent
AC		fair	fair	excellent	poor

metal. In order to produce the optimum overlaid layer, the appropriate welding condition is selected to reach the dilution ratio of about 10% by using highly concentrated Al-Cu filler rod with more than 34 mass% copper.

Arc weld-surfacing is characterized by a thick modified layer and an inexpensive process, although the improvement in hardness is low compared with other surface alloying processes. It is also noted that a layered structure containing AlN can be formed in aluminum alloy substrates by TIG welding in a nitrogen shielded atmosphere. Wear resistance of aluminum alloys are definitely improved by the introduction of in situ composite layers, such as the surface nitriding in molten state.

6.2 Plasma Transferred Arc Surfacing

6.2.1 Equipment and Alloying Materials

Plasma transferred arc (PTA) surfacing is a surface modification technology which produces a thick alloyed or composite layer composed of the elements from supplied metallic or ceramic powders. The PTA surfacing, which was introduced in the early 1960s [112], is characterized by a particular thermal source of transferred arc with a slow flame velocity, different from plasma spraying. Although there are many kinds of thermal sources applied to surface alloying, an attractive process is mainly performed using plasma and laser as a thermal source, because of the following reasons:

(a) The thermal sources are easy to realize an automated process for both large scale and selected surface modifications.
(b) The surfacing can be performed in an air atmosphere without any vacuum system.
(c) Powder materials such as ceramics and composite can be utilized, instead of wires which are sometimes difficult to produce.
(d) The traverse speed of torches is fast, since the energy density of thermal sources is high and powder materials are ready to melt down.
(e) In addition to the formation of thick and anti-peeling alloyed layers, the overlaid layers are prevented from the dilution and the oxidation.

The system of PTA surfacing consists of a plasma arc torch, a plasma power source, a pilot power source and a powder feeder, as shown schematically in Fig. 19. A direct current pilot power source connected between the tungsten electrode and the nozzle, controls the operating characteristics of the nontransferred arc. A main plasma power source connected between the electrode and the substrate, controls the transferred arc to heat and melt the supplied powder and substrate surface. The powder supplied from a feeder pass through the torch with carrier gas surrounded by shield gas, after entering the plasma flame, the powder are melted and deposited on a substrate. The high power equipment was developed to improve the deposition rate [113], where a low degree of dilution, a smooth modified surface and homogeneous microstructure were successfully achieved. On the contrary, the micro-plasma torch was also developed to modify a selected area precisely on a substrate [114].

If the supplied powder is too fine, it does not reach the molten aluminum pool on a substrate surface. In contrast, if the powder is too large, it is not fully fuse up to the core, but partially melted particles. Most metallic powders are possible
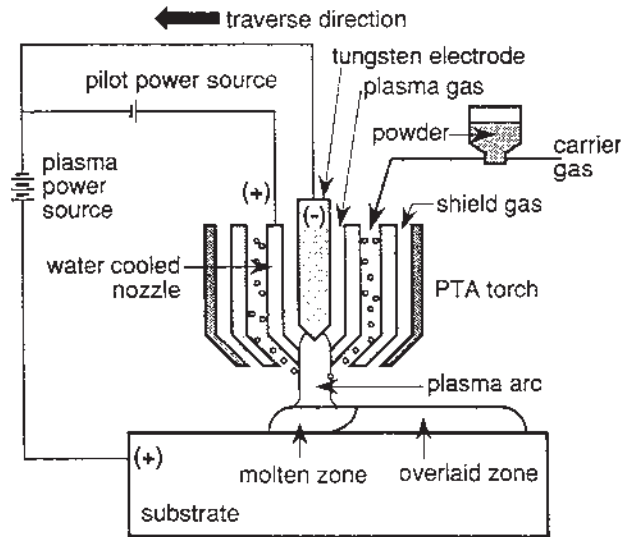
Figure 19 Schematic drawing of PTA surfacing system and overlaid layer formation.

to supply to this process, especially iron, cobalt and nickel base alloys, and these coatings produced by PTA surfacing exhibit good wear resistance especially at high temperatures. Ceramic powders of carbides, oxides and nitrides, are also utilized, including premixed powder to form a composite surface layer even on an aluminum alloy substrate. Fine ceramic particles have to be homogeneously dispersed in the metal matrix without the fusion and the decomposition during the process by adjusting the input thermal energy. Then, supplied powder may be a simple premix of ceramics and metal powder, but for the homogeneous distribution of fine ceramic reinforcements in the modified layer, we had better supply composite particles, as shown schematically in Fig. 20 of a WC/Co particle. Pelletized with cobalt and WC powder, or cobalt coated WC powders are experimentally utilized, although they become more expensive due to the additional powder preparation process. The good wettability between ceramic particles and molten metal is also required for the formation of composite layers with no defects, such as porosity, this is the reason why carbide powders are mainly utilized in this process.

6.2.2 Surface Alloying Procedure

The formation of modified layer by PTA surfacing was schematically shown in Fig. 19. The pilot arc is firstly initiated between the tungsten electrode with a negative polarity and the water cooled nozzle with a positive polarity by the pilot power source. An argon gas is introduced into the gap to produce the high temperature plasma, which is converged through a small hole at the nozzle tip. It reaches a substrate surface with high energy density due to a thermal pinch effect. Then, the arc current passes through the substrate by operating the main plasma power source. It is called a transferred plasma arc, by which the substrate is heated up to make a molten pool on the surface.
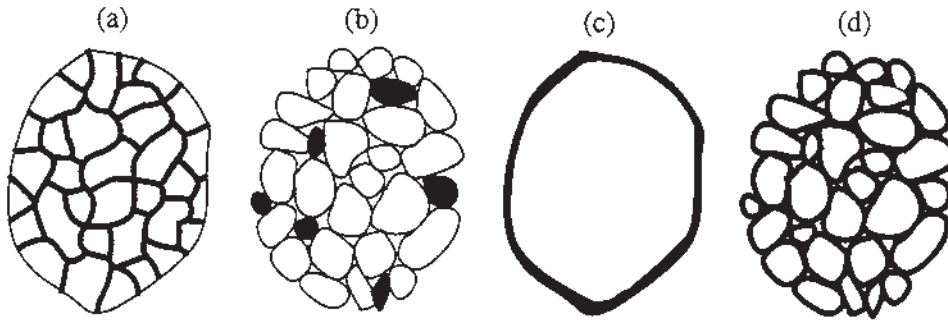
Figure 20 Different morphology of WC/Co composite particle, (a) sintered and crushed, (b) pelletized and sintered, (c) coated and (d) coated and pellitized (hollowed area: WC, solid area: cobalt).

Overlaid powder are fed into the plasma arc along with a carrier gas such as argon, and heated up in the arc. The molten droplets are mixed with molten surface on the substrate to form alloyed layer. The powder feeding rate is generally in the order of 1 g/sec depending on the plasma power. Because the plasma arc is converged by a hole at the water cooled nozzle tip, the energy density in PTA surfacing, that is 500–1000 MW/m^2, becomes much higher compared with that of TIG welding of approximately 200 MW/m^2.

The PTA technology is based on the fusion and solidification step of supplied powder, so that the modified layers may contain porosity, oxides, cracks, segregated ceramic powder and their decomposition. However, the modified layers of PTA surfacing are characterized by nearly pore-free even when the thickness of a modified layer is an order of mm, and also a good adhesive strength, which is an advantage in all the modified layers fabricated by surface alloying. The oxidation of overlaid layers are not so severe compared with that in coatings sprayed in an air atmosphere, because the distance between the electrode and substrate is shortened such as 10–20 mm, and the inert gas shielding over the molten surface prevent from the oxidation.

6.2.3 Microstructure and Properties of Modified Layers

Metallic powders are reacted with molten aluminum to form an alloyed layer composed of intermetallic compounds of aluminides or the added metal phases fundamentally according to their phase diagrams. The modified layers are hardened by the precipitated phase of aluminides. The structure of surface alloyed layers with metallic powder is classified into the following four categories:

(a) Copper and silicon which react eutectically with aluminum at relatively low temperatures produce a homogeneous alloyed layer.
(b) Chromium and titanium which react peritectically or monotectically with aluminum at rather high temperatures produce a nearly homogeneous compound layer at the surface.
(c) Nickel which reacts peritectically and eutectically with aluminum at a relatively high temperature produces an alloyed layer composed of compound and matrix, although a homogeneous alloyed layer with a small amount of nickel content in the matrix.

(d) Tantalum and tungsten which have very high melting points do not react with aluminum, but simply exist at lower part of alloyed matrix layer with the original shapes.

The hardness of modified layers by PTA surfacing is changed with addition of alloy elements depending on the distribution and kinds of intermetallic. However, the microcracking may occur in the modified layer when the precipitates become larger or segregated, and the cracking has been observed during the cooling stage in the layer with hardness of over Hv300, regardless of alloying elements.

Ceramic powders form a particle-dispersed modified layer on the aluminum substrate by the PTA surfacing. The cross-sections of the modified layers are schematically shown in Fig. 21, which was examined using various carbides, nitride, boride, oxide and silicide particles [115]. The fundamental factor classified into the five categories is a melting point of supplied ceramics: if it is higher than approximately 2773 K, the modified layer is a composite material (types 1–3), because the powder is not melted without any reactions with molten aluminum, but distributed in the aluminum alloy matrix. In contrast, the supplied powders with a low melting point are fused in the plasma arc, so that they form aluminides or simply slag-like deposits without mixing on the solidified layer. The distribution of non-melted ceramic particles in aluminum matrix is fundamentally governed by the characteristics of particle and processing parameters, that is density, particle size and wettability and reactivity with molten aluminum as the former factor, and the latter of high current density, low traverse speed of the torch and low amount of supplied powder tends to provide the modified layer of type 3. Ceramics powder with higher density compared with molten aluminum tend to flow down in the modified layer before the solidification of matrix, whereas lower density powders do the reverse.

The particle of SiC (ρSiC $= 3.12$ Mg/m^3) with the nearly same density as that of the molten matrix (ρAL$=2.39$ Mg/m^3) uniformly distribute in the thickness direction of the modified layer. The difference in density between ceramics particle and molten matrix metal fundamentally rules the powder distribution. However, even if there exists a large difference in density, such as NbC (ρNbC$=7.82$ Mg/m^3) and aluminum, the particles distribute approximately uniform with a small sedimentation. The particles of Cr$_3$C$_2$ (ρCr$_3$C$_2 = 6.68$ Mg/m^3) with the relatively low melting point are dissolved to matrix, instead of retaining the original shape, then the slag-like deposit is formed on the solidified layer.

Composite layer with highly strengthened matrix is an attractive approach to improve the wear resistance, a simultaneous addition of metallic and ceramic powder is the simplest method in the PTA surfacing to create it. Namely supplied metal powder reacts with aluminum to form an alloyed matrix layer, ceramic particles play a role of reinforcement. The distribution of ceramic particles depends on, not only the kinds of matrix alloy, but also, the segregation of reinforcement in the supplied powder, as previously mentioned. Composite powders are utilized to solve the particle segregation in PTA surfacing, in which the constituents become finer and ceramic particles with even a different density from that of molten aluminum are found to be uniformly distributed in the matrices of modified layer. For example, composite powders of Ni-Cr containing NbC, TaC and TiC are mixed, granulated and sintered to form alloyed layers on the aluminum substrates [116].
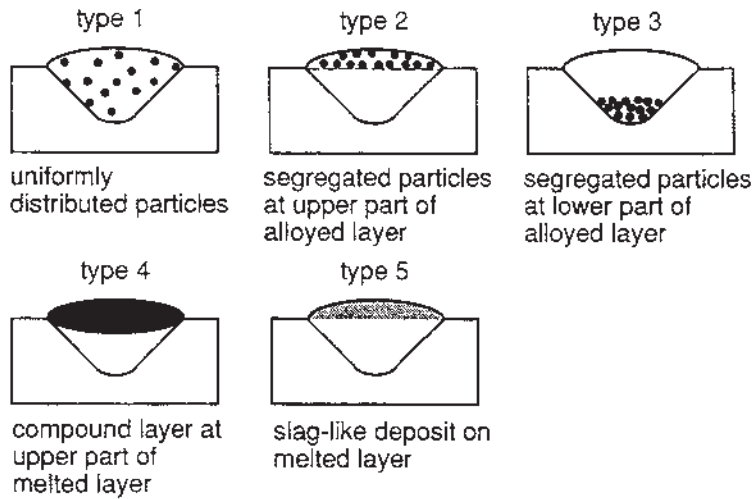
Figure 21 Different cross-section microstructure of modified surface processed by PTA surfacing.

Fine NbC particles are uniformly distributed in the matrix compared with TaC and TiC particles, instead of floating or sedimentation in the modified layer. It is found to show a superior wear resistance. In other examples, oxides of Cr_2O_3 and graphite powder also show good wear resistance, through preventing the particles from the decomposition and improvement in wettability by coating the particle surfaces.

An alloying of TiC is done by PTA process on the Al-Cu pre-overlaid layer as a matrix of a composite layer. The composite layer with Cu and TiC particles is made on the Al substrate. TiC particles were added by PTA process with 3 passes on Al-48 mass% Cu pre-alloyed layers. Surface morphology of the composite layer is smooth and cracking or porosity were rarely observed on composite layer. TiC particles are distributed uniformly from the top surface to the bottom modified layer. This uniform distribution of TiC particles is also obtained independent of the Cu content and the number of the pass for TiC addition.

Figure 22 shows the result in wear tests of base Al plate, Al-48 mass% Cu pre-alloyed layer, TiC composite layers on Cu free or Al-48 mass% Cu pre-alloyed matrices. Single addition of Cu or TiC decreases the wear volume to the nearly same level, about 1/10 of that of base Al plate and which is the same wear level as those of mild steel and austenite stainless steel. These decrease in wear volume caused by Cu or TiC single addition are due to the hardness increase of Al-48 mass% Cu hypereutectic structure and TiC particle dispersed composite structure, respectively. Moreover, TiC composite layer on Al-48 mass% Cu pre-alloyed matrix shows an extreme decrease in wear volume, about 1/100 of base Al plate and 1/10 of each single addition at an increased TiC volume fraction 35 vol% by 3 passes. In the case of the TiC composite layer on Cu free Al matrix, TiC particle moves easily by the abrasion of the counter disc due to a deformation of Al matrix itself, because the Al matrix is so soft. Therefore, it is difficult to fix the TiC particles there. On the contrary, at the Cu pre-alloyed layer the TiC particles are fixed tightly

to the matrix, because which is strengthened by a fine hypereutectic structure. These results clearly show that the strengthening and/or hardening of Al matrix are very effective to increase the wear resistance as same as the increasing the volume fraction of hard ceramic particles [113].

6.3 Laser Alloying

6.3.1 Equipment and Coating Materials

Surface modification by laser alloying is characterized by the process performed in an air atmosphere, supplying metal and ceramics with a high melting point, and a less thermally affected zone because laser is a thermal source with an extremely high energy density. With a laser alloying process, selective surface modification can also be performed by focusing a laser beam, in addition to receive a large scale modified surface, so that it has a high potential of the application to the surface modification of aluminum alloys.

Laser beams can be converged by an optical lens, so that the energy density is thinly and widely controlled with changing the degree of defocusing. The thinly focused beam remelts only the small selected areas on a substrate, which can be applied to a micro surfacing. The defocused laser beam, which is utilized to obtain a thin alloyed layer for the prevention from the dilution due to molten aluminum, is usually irradiated to perform large scale surfacing. Even the inner surface at a
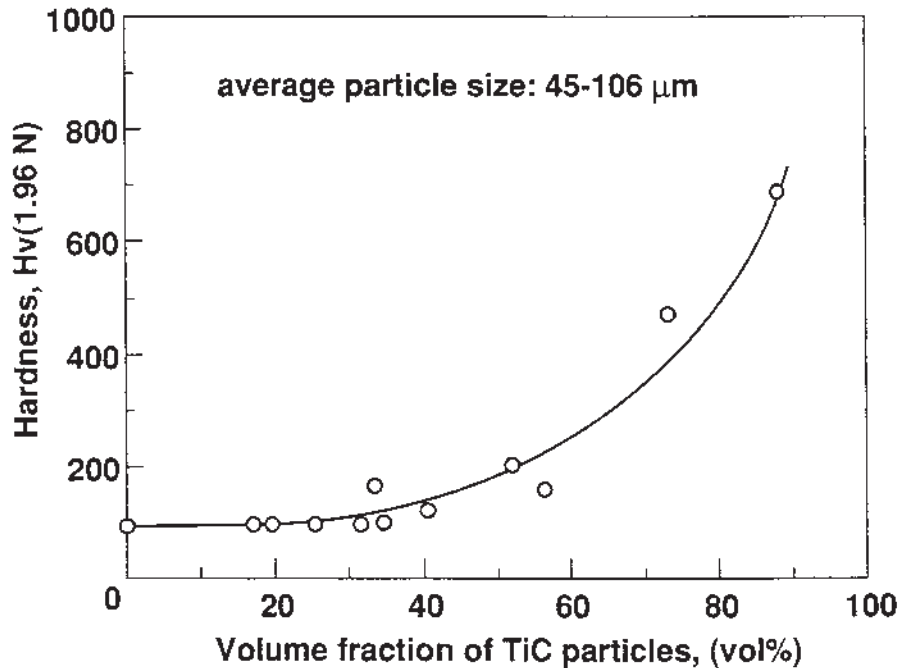
Figure 22 Change in microstructure as a function of particle volume fraction in modified layer by PTA surfacing.

small concave portion with a remote laser source can be modified using a mirror and/or optical fiber. Multiple surface areas can be simultaneously alloyed by splitting the laser beam.

There are mainly two kinds of laser sources to be utilized in surface modification of aluminum alloys: a CO_2 (wave length, $1 = 10.6$ mm) and YAG ($1 = 1.06$ 1.06 mm) laser. The absorption efficiency of the CO_2 laser beam is low because the one with a long wave length is reflected approximately 95% on an aluminum surface, so that the surface is coated by graphite paste as an absorber or roughened to improve the absorption efficiency. Since the molten surface of aluminum alloys absorbs the laser energy much more, the modified layer with a thickness of over 500 mm is easily obtained. The power of CO_2 laser is higher than that of a YAG laser, so that it is favorable to utilize for the thick surface modification. There are two different concepts regarding supplied powder utilized to the laser surface alloying,

(a) Forming intermetalic compounds or precipitates reacted with molten aluminum on the surface.
(b) No formation of any compounds or precipitates.

In the category of (a), there is an addition of semi-metallic elements: boron and silicon which harden the aluminum due to hard AlB_{12} (Hv3300) precipitates and silicon phase (Hv1000). Metallic elements are also supplied to form aluminide compounds such as $TiAl_3$ (Hv400–700), $NiAl$ (Hv300–700), $NbAl_3$ (Hv500) and $FeAl_3$ (Hv800) reacted with aluminum. Nitrogen and oxygen as a reactive gas which forms AlN (Hv1200) and Al_2O_3 (Hv2000), respectively, contribute the hardening the aluminum alloy substrates. Supplied hard ceramic particles are dispersed in the modified layer, as the example of (b), which is typical metal matrix composites. Then, the hardness increases with an increasing volume fraction of ceramic reinforcement. The supplied ceramic particles are shortly exposed in high temperature during a laser alloying process, so that they are required to be difficult to react with molten aluminum. According to this criterion, TiB_2, TiC, SiC, BN, Si_3N_4 and Al_2O_3 particles are frequently utilized as same as those of cast aluminum matrix composites.

6.3.2 Alloying Procedure

The simplest application of laser to surface modification processes is rapid remelting and solidification of aluminum alloy surface without adding any alloying elements. This proceeds remelting due to the condensed laser energy and rapid solidification due to self-cooling which causes a fine microstructure and over saturated solid solution depending on the degree of beam defocusing and the traverse speed of a substrate. The fast traverse speed is required to obtain the high cooling rate, although a modified layer becomes thinner.

The principal process with an addition of the alloying powders into molten substrate surface is called laser alloying, which forms a composite or alloyed microstructure. When the molten layer thickness is thin to prevent the alloyed surface from the dilution, it is generally called laser cladding. Hardness of modified layer depends on the volume fraction of hard ceramic particles or intermetallic distributed in aluminum alloy matrix.

The supply methods of alloying elements are schematically shown in Fig. 23. They are classified into two categories: the coating on the surface prior to the laser alloying and the direct powder injection to a molten pool along with a laser irradiation. In the former case, there are pasting, sputtering, electroplating or electroless plating, containing dry platings of spraying and evaporation, and cladding. Dense pasting with an uniform thickness is necessary to achieve an uniform modified layer thickness and a required alloy properties. In contrast, wire or powder is injected into the molten pool along with a laser irradiation, as the latter example, in which the powder feeding system including an injection nozzle is designed to feed with a required feeding rate. The powder injection method has an advantage of simultaneous treatment with laser irradiation. It is necessary to supply the powders with different densities, sizes and shapes, in which the injection angle and position of powder injection nozzle are controlled to decide the exact position to be modified with the required amount of powders. An argon shield gas is usually blown on the substrate surface to prevent the molten layer from oxidation. Even a reacting gas with molten aluminum, such as nitrogen and oxygen, can be supplied to form the nitride and oxide, respectively, instead of argon. However, the nitride and oxide formation are not so much, so that the reactive gas method needs to be used jointly with other methods such as a direct addition of ceramic powder.

6.3.3   Microstructure and Coating Properties

The effects of alloying elements on a microstructure and coating properties are examined in laser alloying with an addition of metallic and semi-metallic elements, titanium, nickel, niobium, iron, copper, boron and silicon powder or deposit. The alloying elements can form aluminides to improve the hardness and wear resistant property, except silicon powder.

The metallic elements of iron, cobalt, nickel, titanium, copper and chromium are expected to provide a good wear resistant modified layer with high hardness. The measurements of microhardness profiles were done on the laser alloyed
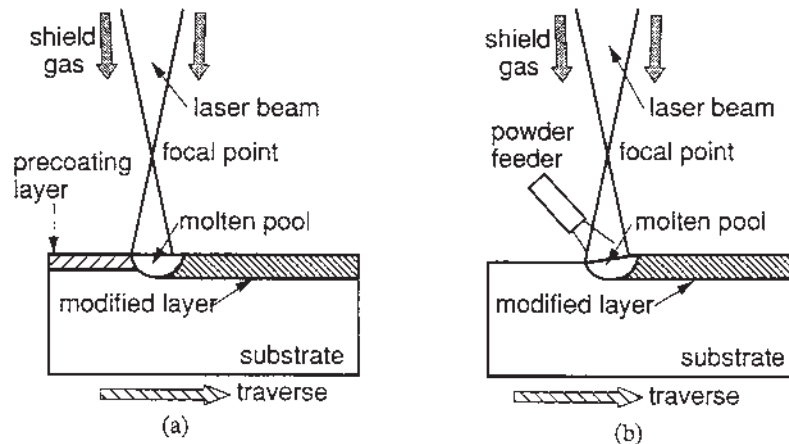
Figure 23   Schematic drawing of supply methods of alloying elements, (a) coating on the surface pior to laser alloying and (b) powder injection to the molten pool along with laser irradiation.

cross-section of aluminum alloy substrate which were electrodeposited nickel layer with 25 mm and 50 mm [117]. The alloyed layer is composed of aluminum, Al3Ni and small amount of Al_3Ni_2, because the substrate is mainly aluminum and silicon, The transition from the high hardness of melt zone (Hv550–600) to the low hardness of substrate (Hv100) is very sharp; this is due to the small heat affected zone. The laser alloyed surfaces exhibit a good wear resistance compared with that of conventionally precipitation hardened aluminum alloy (Hv200). The substrate with plasma sprayed nickel coating of approximately 100 mm thickness is expected to provide the same microstructure and hardness after the appropriate laser alloying. However, the absorption efficiency for a CO_2 laser becomes higher than that of electroplated substrate surface, because of the coarse surface roughness. The alloyed surface layer formed on a precoated aluminum substrate with titanium powder is composed of aluminum, TiAl and $TiAl_3$; it has a hardness of Hv400–600 [118]. The laser alloying with the injection of premixed powder of niobium and aluminum also exhibits a hardness of Hv500–650 due to fine precipitates of $NbAl_3$ [119].

The most interesting alloy element is iron in laser alloying on aluminum alloys. The alloyed microstructure is composed of aluminum and $FeAl_3$, in which the hardness increases up to Hv550 depending on an iron content in the modified layer, as shown in Fig. 24 [120]. The hardness increases with an iron content increases; it becomes over Hv600 with an iron content of over 40 mass%. However, when the microstructure contains blocky $Fe2A15$ precipitates, the microcracking is ready to occur. The most attractive microstructure of fine acicular $FeAl_3$, which provides a hardness of Hv300–550, is obtained in the range of 30–40 mass% Fe.

The basic operating conditions are reported for laser alloying of a cast Al-Si alloy substrate with an iron-base material [121]. The main difficulty is arisen from the interface cracking during the process, the factors affecting the interfacial cracking ratio, which is the ratio of cracking extent to the total interface region, are found to be preheating temperature and tempering temperature. Substrate temperature needs to be controlled between 583 and 653 K. The laser processing parameters, that is laser power, traverse speed, powder-feeding rate, are less important for the interfacial cracking. The material factors of the element content in the alloyed layer also had significant influence on the cracking ratio: the aluminides of $NiAl_3$ and NiAl are more brittle compared with iron aluminides, although iron content in the interface region is much higher than that of nickel.

Reactive laser alloying is also possible, but there are few reported researches. Typically, coated substrates with aluminum, chromium and titanium powder were irradiated by a CO_2 laser with an oxygen or nitrogen atmosphere. The alloyed layers including Al_2O_3, Cr_2O_3 and TiN are formed up to a thickness of 50 mm. They are hard and provide a good wear resistance [122].

As an example of metal matrix composite coatings, an injection of ceramic powder such as SiC, TiC, TiB_2 and TiN can be utilized in laser alloying [120]. The surface properties depend on the volume fraction of ceramic reinforcement, as mentioned before. When SiC powder and graphite coated substrates were irradiated by a CO_2 laser, the modified layer with a thickness of 0.8 mm and a hardness of Hv150–220 provides a good wear resistance, although it was not so hard. Laser alloying on TiB_2 powder coated substrate are also successfully performed with a thickness of 0.5 mm and a hardness of Hv500–600, without occurring any cracking and pores.
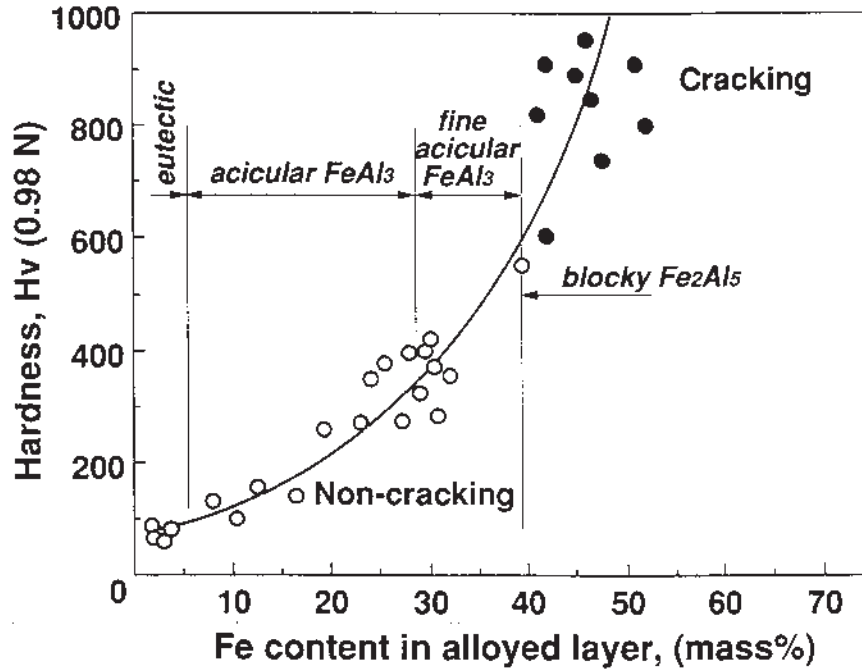
Figure 24 Change in microhardness as a function of iron content in alloyed layer.

Aluminum substrates were coated with a layer of premixed aluminum and SiC powder with different sizes, then the surface was scanned with a laser beam [123], only after using coarse SiC particles (20–40 mm), some SiC particles retained. During the laser irradiation most SiC particles react with oxygen, Al_2O_3 and SiO_2 are mainly formed as reaction products. A cellular network of aluminum and Al-Si eutectic is found in the bottom layer of the molten pool. In the top layer of the laser tracks, oxide plates are firstly formed, after which the remaining fluid is solidified as a cellar Al-Si network. It means that a shield gas stream is not enough to prohibit sufficiently contact of air with the molten metal. Originally designed composite layers are expected to form in the laser alloying technology, so that there must be new and modified processes regarding a laser irradiation in near future.

A new technology for manufacturing of engine valve seats has been developed and practically introduced using a laser alloying [124]. This process, different from the conventional process that valve seats made of sintered alloy were pressed into a cylinder head, directly forms valve seats onto a cylinder head. Because of an extremely small dilution with molten aluminum, it is called a laser cladding. In order to develop a cylinder head with laser cladded valve seats, the laser cladding has been established in such a way that copper based alloy is deposited onto aluminum alloy with an extremely small dilution and stable bead. The specially designed cladding alloy, which has two phases in the liquid state and in which iron and molybdenum-rich hard particles are dispersed in the matrix, has been developed for this purpose. Based on the newly designed material, the practical laser cladding technology has been successfully completed. It is revealed that the cylinder head

with laser cladded valve seats has many advantages that are the improvement of engine performance and temperature reduction at valves and their seats compared with conventional one.

6.4 Electrical Discharge Alloying

6.4.1 Equipment and Coating Materials

Electrical discharge machining (EDM) employs high frequency sparks between an electrode and substrate to form a required shape on a conductive substrate. The tool and substrate form a pair of electrode separated by approximately 20–100 mm and are completely immersed in a liquid dielectric, such as kerosene. A significant observation in EDM reveals the presence of traces of tool electrode material on the machined substrate surface. This phenomenon has been expanded into the surface modification, employing such process conditions which can be causes significant element transfer of selected tool materials onto the machined surface to obtain specific functional characteristics. One of the best examples is the surface modification of steel with hard WC/Co depositions using green compact electrode of WC/Co [125,126]. The new process named electrical discharge alloying (EDA) is proceeded in such a way that molten electrode elements are transferred and reacted on a substrate surface with a discharged arc between the electrode and substrate. With the use of green compact electrodes having a considerable wear rate, instead of solid copper electrodes, the modified layer thickness on steel substrate can be increased.

The EDA process is based on the electro-spark deposition [127], which is a pulsed-arc microwelding process using short-duration, high current electrical pulse to deposit an electrode material on a metallic substrate in an air atmosphere, instead of in a working fluid. The fused and then metallurgical bonded coating can be formed with a low total heat input that the bulk substrate material remains at or near ambient temperatures. The short duration of the electrical pulse allows an extremely rapid solidification of the deposited material and results in an extremely fine-grained, homogeneous coating. This microstructure is believed to contribute to the good tribological performance.

In order to apply the EDA process to the surface modification of aluminum substrates, the substrate was electrically discharged in the working fluid of kerosene using a titanium contained green compact electrode [128,129]. In the EDA process, other metallic powders can be used based on the chemical reaction with a working fluid and an aluminum substrate. That is also true in a working fluid, for example, silicone oil was utilized to transferred an alloying element of silicon to the substrate [130,131], in addition to a carbon element.

6.4.2 Alloying Procedure

A conventional electrical discharge machine can be utilized for the surface modification of aluminum, an electrode was then connected to give a negative polarity. A dielectric hydrocarbon oil, mainly composed of kerosene, was used as a working fluid. With the combination of titanium contained in electrodes and carbon decomposed from the hydrocarbon working fluid, the coatings of TiC/TiAl in situ composite layer are successfully formed on aluminum substrates.

According to the observation of discharged craters on the substrates, an element transfer from the electrode and the working fluid to a discharged crater is expected to proceed as shown in Fig. 25.

1. Generation of discharged arc due to the dielectric breakdown of working fluid.
2. Local fusion on the electrode and substrate surface.
3. Vaporization of working fluid due to a discharged arc.
4. Dripping of molten electrode droplets from the electrode.
5. Enrichment of decomposed carbon from the working fluid.
6. Element transfer to a crater.
7. Formation of a rapidly solidified discharged crater.

The actual discharge conditions, such as the discharge current, pulsed width, duty factor and gap voltage, were appropriately settled before the treatment. It should be noted that the current density on the electrode with negative polarity is higher than that on the substrate, which means the fusion is ready on the electrode compared with on the substrate. In addition to the discharged conditions, another process parameters of the working time and electrode forming pressure also affect the modified thickness and volume fraction of TiC precipitates, which decide the elemental distribution and the related hardness of modified layers. With the examination of an element transfer from the electrode and working fluid through a discharged arc, the average diameter and alloyed depth of separate discharged craters on the substrate surface, increase as the pulse width of the discharge conditions is increased.

Modified layers with a nearly uniform thickness can be formed by successive discharge arcs at different points on the substrate. The modified layers with an average thickness of 100 mm can be formed within a short processing time of a few minutes. The modified layer consists of non-reacted aluminum, TiC precipitates
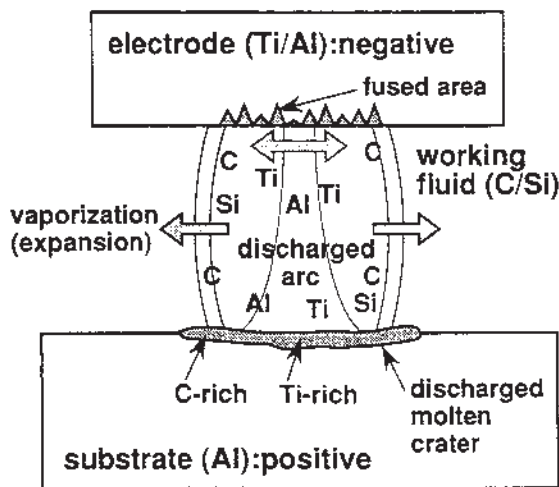
Figure 25   Element transfer and formation of discharged crater on substrate.

and TiAl matrix synthesized by the reactions between the electrode and working fluid, an electrode and a substrate. Hence, the modified layers formed by EDA are in situ composites of TiAl matrix with TiC precipitates.

It is possible to control the thickness and the volume fraction of TiC precipitates in the modified layers formed on aluminum substrates, by adjusting the parameters in the EDA process. For example, the effect of pulse width, which is one of the discharge conditions, on the thickness of modified layer and volume fraction of TiC is shown in Fig. 26. It can be recognized that the pulsed width leads to the increment in the modified thickness and also the volume fraction of TiC in the composite layers. The other discharge conditions, that is the peak current, duty factor and gap voltage, have an effect on the thickness and the volume fraction similar to that of the pulsed width.

6.4.3 Microstructure and Properties

Typical micrograph taken on the cross-section of modified layer is shown in Fig. 27. The matrix of modified layers consists of TiAl and aluminum, as identified by XRD. The TiC morphology can be formed from fine dendritic particulate and the size of TiC precipitates becomes smaller with increasing distance from the surface. The quantity of TiC varies depending upon mainly the pulsed width and working time.

Figure 28 shows the characteristic x-ray intensity profiles of electron probe microanalysis on the cross-sectional in situ composite layers with different volume fractions of TiC. The x-ray intensities of titanium and carbon become the highest close to the surface, and become gradually lower toward the inside. In contrast, the AlKα a intensity increases with an increasing distance from the surface and then it becomes constant. Hence, the fact of compositionally graded modified layers means a high resistance for a peeling off phenomenon resulting from thermal stresses generated at the interface.

The hardness profiles for cross-sections of modified layers are shown in Fig. 29. The maximum hardness is obtained close to the surface and the profiles decrease to the substrate hardness. The hardness close to the surface can be controlled to have any value between Hv350 and Hv1300 by adjusting the volume fraction of TiC. This surface modification technique is available to harden not only a varieties of aluminum alloys, but also aluminum matrix composite substrates [130,131].

7 DIFFUSION COATING

Diffusion coating on aluminum is not easy because of the limited temperature range up to the melting point of aluminum and its alloys. However, many investigation works are devote and several possible technologies are observed recently.

The easiest way to form diffusion layer on aluminum surface is to diffuse alloying elements by heating after coating aluminum surface by plating, spray coating or by vapor deposition methods.

7.1 Plating and Diffusion Coating

Copper, nickel and their alloys (Cu, Ni, Cu-In etc.) can be diffused on aluminum by heat treating below the melting point of aluminum after plating. By heat treating at 600°C for 25 h after 20 μm nickel plating, diffusion layer of 130 m with Vickers
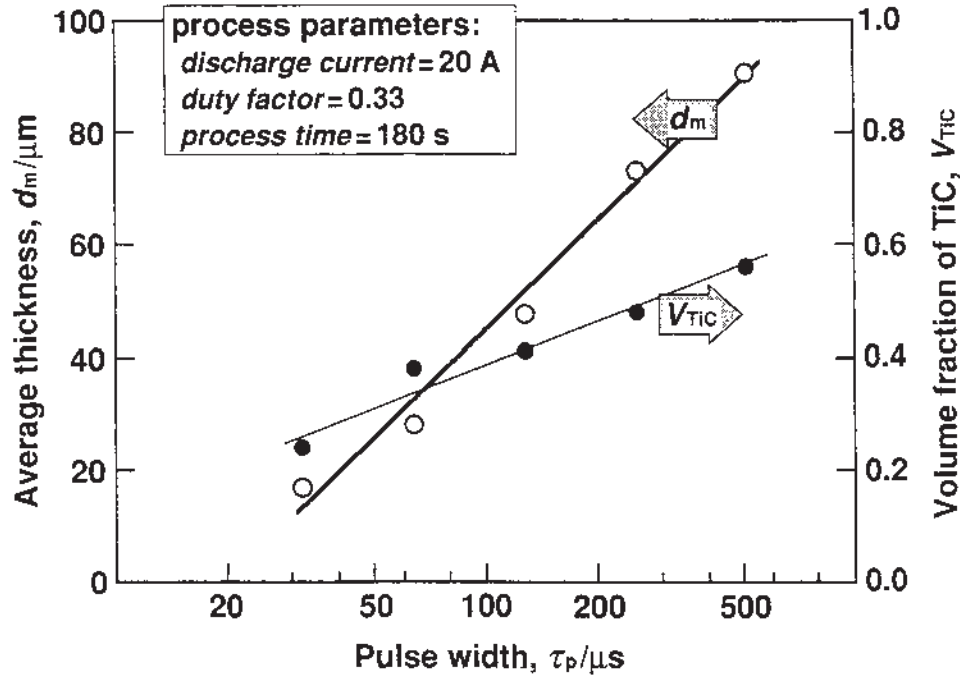
Figure 26 Variation of average thickness and volume fraction of TiC as a function of pulse width.
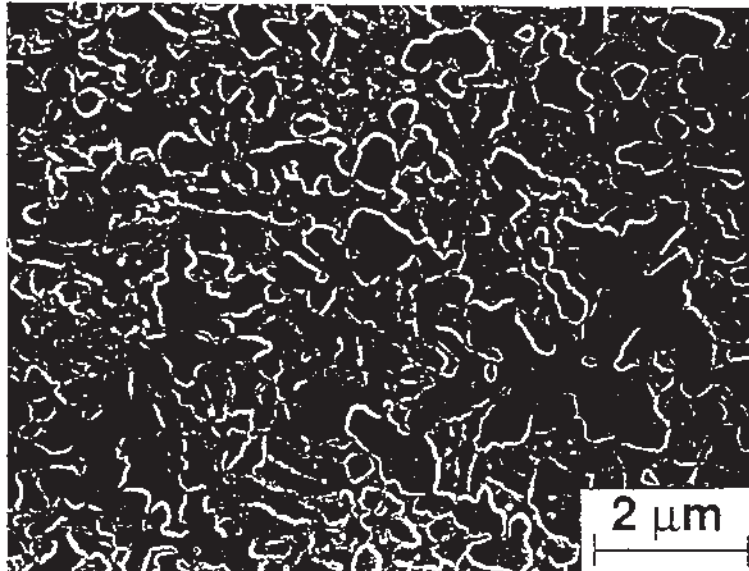
Figure 27 Scanning electron micrograph showing TiC precipitates on the cross-section of modified layer with $V_{TiC} = 0.41$.
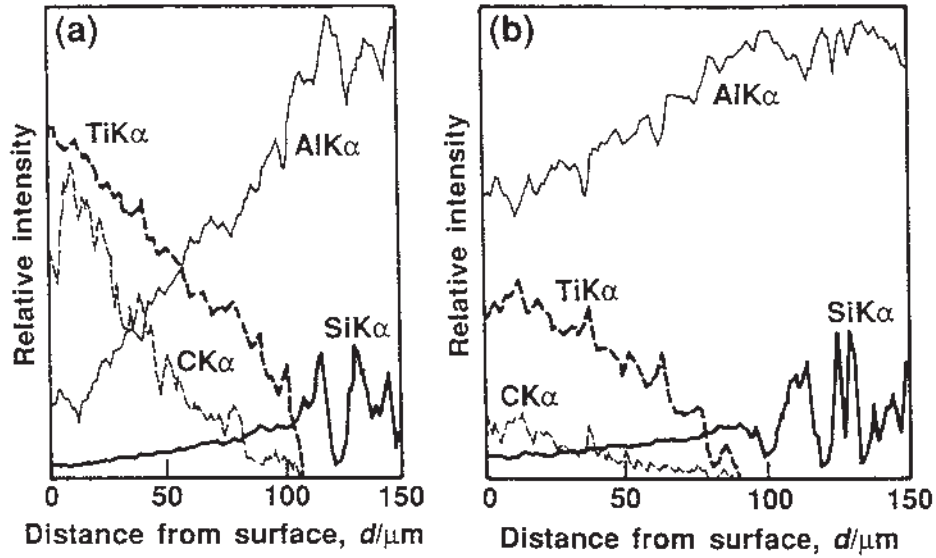
Figure 28 Characteristic X-ray intensity profiles of EPMA on the cross-section of modified layers with different TiC volume fractions, (a) $V_{TiC} = 0.66$ and (b) $V_{TiC} = 0.41$.
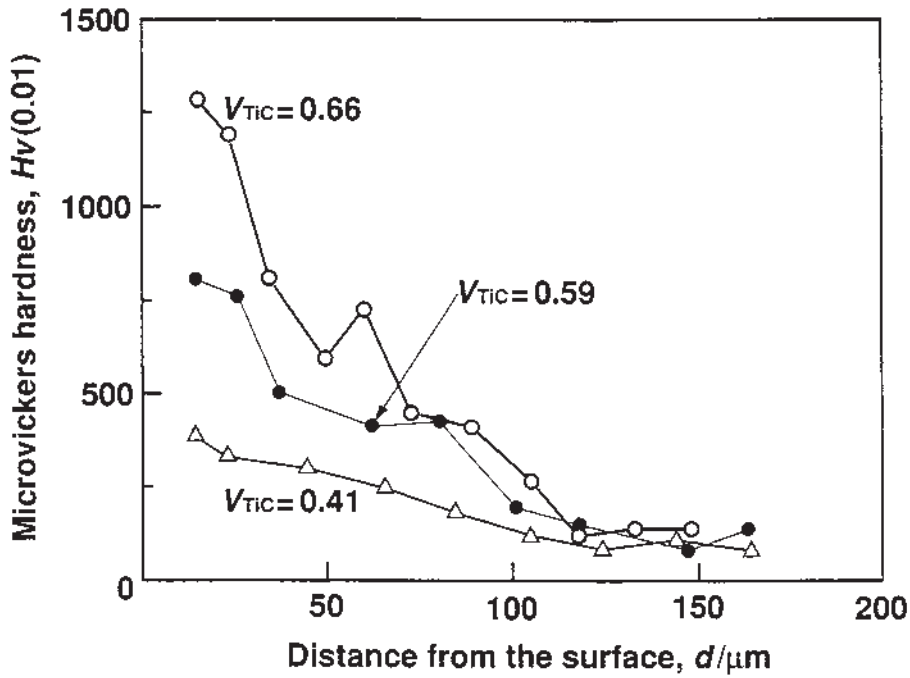
Figure 29 Hardness profiles on the cross-section of modified layers with different TiC volume fractions.

hardness 550–800 was formed, and by heat treating for 49 hr at 500°C after 20 m copper plating, 65 μm of diffused layer with Vickers hardness 450–600 was reported by Matsuda et al. [132]. Many similar plating and heat treating methods can be applicable to improve adhesion strength of aluminum surface.

7.2   Nitriding of Aluminum

Aluminum has not the solubility of nitrogen and it is difficult to get a diffusion layer on the aluminum surface. However, many research efforts are devoted to form aluminum nitride layer on their surfaces.

Tachikawa et al., investigated plasma nitriding methods after Argon sputtering at temperature of 400–500°C and formed a few microns of nitrided layer by heat treatment of 5 hr at 500°C [133]. Similar investigations are reported in recent international conferences [134,135], but as far as the nitriding methods simply depend on ionization or plasma energy have a limit to get thick nitride layer.

A unique nitriding technology by putting special paste on aluminum surface was proposed by Yamada et al. [136]. This method enables the formation of hundreds of microns of aluminum nitride layer by heat treatment in pure nitrogen atmosphere at temperature of 470–550°C and formed 20–500 μm layer which contain 20–90% of AlN in the formed nitride layer. This results suggest a methods to form aluminum nitride by some catalytic reaction of coated paste and shown a possibility to get thick hard layer on aluminum surface. Also, a recent investigation of activated nitriding method at 500–560°C reported by Okumiya et al., in fluidized a bed revealed the similar possibility to get thick (-200 μm) nitride layer [137].

The most recent technology to form low friction surface on aluminum is investigated by Miyasaka who reported resulting thin (1–2 μm) tin diffusion by fine particle impact peening of ceramics and tin powder on aluminum surface. Also fine ceramic particle impact peening under nitrogen atmosphere can form an aluminum nitride layer at the surface [138].

Although the surface hardening of aluminum has been understood as a difficult operation, the recent investigation results suggest the existence of a new technical seed to break through the traditional technologies.

8   VITREOUS ENAMELING

8.1   Introduction

Vitreous enameling on aluminum is difficult and not popular because of the low melting point in relation with the melting point of frit, and large thermal expansion rate of aluminum and its alloys. Recently, by the development of various low melting point frit (lower than 550°C) materials a lot of enameling of aluminum become possible. Basic materials used for enameling are borax, quartz, feldspar and fluorite and are mixed with nitric or carbonate compounds, and metal oxide of such as lead, nickel, cobalt and antimony. Selection of aluminum alloy is important to prevent troubles caused by softening of base aluminum. Accounting the tendency of age hardening effects of the formed enamel layer and its adhesion Al-Mg-Si type alloys are recommended. Adhesion is relative to the types of formed oxide film on the aluminum oxide layer.

8.2 Pretreatment [139]

The surface of base aluminum should be clean enough for good adhesion. Various pretreatment methods are applied depend on state of treating materials, type of aluminum alloys, frit and application needs. Shot blasting and pre-firing (400–500°C) are effective for rough and or badly contaminated surface. General cleaning processes such as solvent or alkaline degreasing and rinsing are usable for most of low contamination materials. Alkaline chromate processes are also used for pretreatment of alloys contain magnesium.

8.3 Preparation of Vitreous Enamel [139]

Enamel slip are generally processed by ball milling of mixed frit in water. Milling techniques to get fine mixed frit for enameling on aluminum are a little different from that of used for steel enameling. The slip should be free from impurities of milling agents and kept cold and should prevent from agglomerating and aging.

8.4 Types of Vitreous Enamel for Aluminum

Various types of enamel have been developed by different countries and developer. The enamels used for aluminum should be designed to have melting point lower than that of aluminum alloys by addition of lead or alkaline metal compounds.

They are roughly divided into three types, as lead-bearing type and lead-free types such as barium oxide type and phosphorus compound type as listed in Table 33.

8.5 Firing [139]

Firing temperature is selected depending on types of alloy and composition of frit especially in relation to the concentration of lead oxide or alkaline oxide. For lead-free type frit firing process is operated at temperature of 480–560°C and for leaded frit 450–580°C.

9 COMPOSITE POLYMER COATING

9.1 Introduction

Composite Polymer Coating such as Bonded Solid Film Lubricant [145] is an effective method to improve surface properties of aluminum alloy components such as corrosion resistance, seizure or scuffing properties during wear in period and reduction of coefficient of friction. Not only a coating like low friction PTFE, but also various coating films containing solid lubricants are the more widely accepted for aluminum parts such as engine pistons to reduce friction and enable the improvement of fuel mileage. Especially, the future engine technology of light all aluminum engine needs appropriate surface coating to prevent scuffing during wear in period and reduce friction loss.

Funatani et al.

Table 33 Composition of Frit for Enameling on Aluminum Alloy

Composition	Lead base (%)			Barium base (%)		Phosphate base (%)	
	A [144]	B [140]	C [140]	D [144]	E [141]	F [144]	G [142,143]
SiO_2	25–38.4	38–56	30–40	15–30	18–36	–	–
B_2O_3	4–8.7	–	10–15	15–30	8–13	7–10.5	2–15
Na_2O	5–7.5	–	–	10	–	10–12.3	15–33
R_2O	–	0–42	9–12	K_2O:10	21–36	K_2O:6	–
PbO	37.2–50	10–18	25–35	5	–	–	–
BaO	–	–	–	BaF_2:0–5	17–30	–	(CuO, CaO, NiO, SdO):3–12
(Zn,Ca,Sr)	–	–	–	–	–	–	
P_2O_5	–	–	–	–	–	44.4–50	20–32
Al_2O_3	0–2.0	–	1–3	–	0–4	15–18.7	15–23
TiO_2	–	0–11	–	–	0–15	–	–
Li_2O	0.4–7	5–12	0–3	10	–	6.8–7	5–15
Zr_2O	–	–	–	–	0–5	–	–
SnO	–	–	–	–	0–5	AlF:0–4.2	–
F	–	–	–	CaF_2:15	–	NaF:5–6.3	2–15

B: U.S. frits (du pont), C: British frits (Alugas B), D: Alugas D.

9.2 Composition of Solid Lubricant Coating

The base of bonded solid lubricant films composed of Polymer (Polyethylene) or Resin (Epoxy, Poly-Imide, Poly-Amide-Imide and etc.) as carrier of solid lubricants such as PTFE (Poly-Tetra-Fluoro-Ethylene), MoS_2, Graphite, Boron Nitride and etc.

Fluoro-polymer is known as a common low friction material and used for light load applications. For highly loaded applications selected solid lubricants are mixed within polymer or resin carrier and sprayed or coated on surface of parts.

Coating process is generally, as indicated below starting from degreasing to final baking operation. Depend on the surface condition mechanical roughening operation or chemical pretreatments are selected to attain sufficient adhesion strength and wear properties.

9.3 Pretreatment

Figure 30 shows the example of pretreatment and coating methods applied for this coating processes.

9.3.1 Degreasing

Degreasing is the first and very important pretreatment to achieve optimum application of organic coatings containing solid lubricants, and the parts should be carefully degreased. Solvent degreasing in a steam bath has proved to be the most effective, however, the regulation for exhaust emission reduction washing treatment by alkaline solution is becoming popular as the more environmentally friendly degreasing method.

Example 1. Common coating methods.
Degreasing   Anodizing   Washing   Rinsing   Drying   Coating   Baking
         Sand Blasting   Chromate

Example 2. Screen printing method.
Degreasing   Washing   Rinse   Drying   Coating   Setting   Baking
              (hot water)   (pure water)   (screen print)

9.3.2 Removal of Rusts

Common acid or alkaline treatments are usually sufficient to remove corrosion products. The bath should be adjusted to prevent unnecessary reaction with the base materials. Methods to provide roughened surface improves the adhesion of the coated films, by sand blasting with fine alumina or some other hard powder are also proved their worth. All the residue as chemicals, solution or particles used in cleaning and fingerprints should be carefully removed.

9.3.3 Anodic Oxidation

Anodic oxidation treatment provide favorable adhesion strength and wear properties. The types and concentration of alloying elements have profound influence on anodic oxidation processes. Depend on alloy types sulfuric acid or chromic acid baths are selected to attain optimum results. In relation with types of alloy and oxidation baths thickness or weight of surface film should be carefully selected as below.

Example 1. common coating methods.

Degreasing → Anodizing → Washing → Rinsing ⇢ Drying → Coating → Baking
↓   ↑
Sand Blasting → Chromate

Example 2. Screen printing method.

Degreasing → Washing → Rinse → Drying → Coating → Setting ⇢ Baking
(hot water) (pure water)   (screen print)
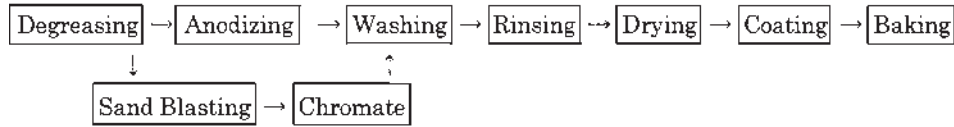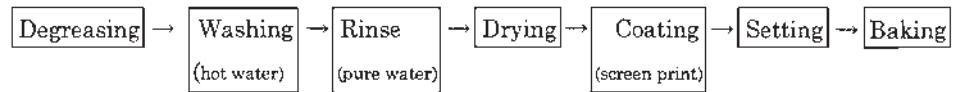
Figure 30   Coating processes for composite polymer coatings.

Sulfuric acid bath: 5.0 μm; 5.5–6.5 g/m^2
Chromic acid bath: 2.5 μm; 2.0–2.5 g/m^2

9.3.4 Washing and Drying

For chromic acid method the rinsing by hot water of about 65–70°C is recommended to dry well. For sulfuric acid method thorough water washing and sealing by dipping into 5% sodium or potassium dichromate solution, and rinse and dry under about 100°C are recommended. With both cases the finger printing of work surface should be prevented.

9.4 Coating Methods

9.4.1 Spray Method

In any coating methods, the selected solid lubricants solution should be well mixed by adequate mixing machine or tools to prevent agglomeration and segregation of solid lubricants.

9.4.2 Brush Paint Method

Brush painting methods can also be applicable, but care should be taken to get uniform thickness as far as possible.

9.4.3 Dip Paint Method

For the economic use of solid lubricant solution dip coating methods can be applicable by preheating of parts before dipping, By preheating up to 80–100°C, solid lubricant film cures quickly and helps the formation of a uniform film and better adhesion. The parts may be dipped for a very brief period only to prevent cooling. Vigorous shaking of the parts during dipping is necessary to get uniform distribution of the bonded lubricant, and prevents sticking of parts. Care must be taken to prevent ignition or explosion by overheating.

9.4.4   Screen Print Method

Screen printing method is the more economical process compared with spraying for yield rate of spray solution and also to maintain uniform thickness on the parts surface. Solid lubricant solution must be prepared for screen printing methods to prevent moistening and thickening by optimizing chemical constituents and concentration control. Solution viscosity, printing speed, preheat temperature, and printing conditions must be precisely controlled to attain uniform thickness and resulting adhesion strength.

9.5   Preheating and Baking

The adhesion of solid lubricant organic coating film is largely dependant on baking performance. Thorough removal of water or organic solvents is the key for utmost adhesion and durability, therefore, the preheating and or good baking methods should be designed to get best fit for the selected types of polymers and solvents.

9.6   Applications

Organic coatings containing solid lubricant are applied to automotive pistons to reduce friction and wear for economy and durability.

Piston skirt surface is usually coated by spraying or screen printing methods with polymer base paint containing solid lubricants. Base resin is selected to meet high temperature operating condition and Polyamide, or Polyamideimide resins are used for piston coating. Solid lubricants such as MoS_2, Graphite and PTFE are mixed with above listed polymers, but their concentration and the details of each lubricants are kept as know how.

The thickness of bonded solid lubricant film is in the range of 5–20 μm. Highest durability is obtained by careful control of every factor, especially the perfect baking condition is quite important to eliminate residual solvents completely [146,147].

Pistons with bonded solid lubricant coatings polyamid-imid base coating containing molybdenum disulfide and graphite are proved to reduce friction energy and contribute to the improvement of fuel consumption. By the application of solid lubricant film enabled the more reduction of friction than traditionally used tin plating. The performance proved with fired engine shows the reduction of friction force by about 11% and fuel mileage improvement of 0.4% [148]. Similar applications are becoming more popular especially for high performance racing engines of motorcycles and automobiles [149–153].

REFERENCES

1. ''Cleaning and Finishing of Nonferrous Metals,'' *Metals Handbook*, Desk Edition, ASM 2nd print, 1985, pp. 29–43,44.
2. M. Yoshida, et al., The 3rd Asian Coating Forum, Proceedings, 1992, p. 280.
3. M. Yoshida, et al., The 74th Conference of Japan Institute of Light Metals, Proceedings, 1988, p. 55.
4. D. B. Freeman, *Phosphating and Metal Pre-Treatment*, 1986, Woodhead-Faulkner Ltd p. 111.
5. J. A. Treverton and N. C. Davis, Surface and Interface Analysis 1981, 3, p. 480.
6. J. A. Treverton and N. C. Davis, Metal Technology, 1977, 10, p. 480.

7. K. Asami, et al., Electrochemica Acta, 1987, 32, p. 337.
8. H. A. Katzman and G. M. Malouf, Applied Surface Science, 1979, 2, p. 416.
9. M. Yoshida, et al., Nihon Parkerizing Technical Report (Japan), 1991, 14, p. 26.
10. K. Nakada, et al., The 4th Asian Coating Forum, Proceedings, 1992, p. 280.
11. U. S. Patent 2,438,877.
12. N. J. Newhart, Metal Finishing, 1972, 70, pp. 49 and 66.
13. L. A. Nimon and G. K. Korpi, Plating, 1972, p. 5942.
14. M. Yoshida, et al., Nihon Parkerizing Technical Report (Japan), 1988, 1, p. 68.
15. K. Hatanaka, et al., J. Japan Inst. Light Metals, 1989, 39, p. 210.
16. U. S. Patent 4,148,670.
17. D. B. Freeman, *Phosphating and Metal Pre-Treatment*, 1986, Woodhead-Faukner Ltd, p. 6.
18. H. Ishi, et al., Metal Finishing, 1993, 91(4), p. 7.
19. U. S. Patent 5,259,937.
20. J. Kawaguchi, et al., The Surface Finishing Society of Japan, 93rd Conference, Proceedings 1996, p. 42.
21. U. S. Patent 5,538,078.
22. *Metals Handbook*, Desk Edn, 1985, American Society for Metals, pp. 29–44,45.
23. "Anodizing" *The Surface Treatment and Finishing of Aluminium and Its alloys*, 5th Edn. Finishing Publication Ltd, and ASM International, pp. 292–293.
24. *Aluminum Handbook*, 5th Edn. 1994, Japan Light Metal Inst., p. 157.
25. Ministry of Supply, Specification D. T. 910 (now DEF-151).
26. G. V. Akimov, N. S. Tomsashov, and M. V. Tyukina, J. of Gen. Chem. (USSR), 1942, p. 12568; Light Metals, 1944, 7(78), pp. 211–219.
27. R. W. Buzzard, J. Res. Natl. Bur. Standards, 1937, 18, pp. 251–257.
28. O. F. Tarr, M. Darwin, and L. G. Tubbs, Ind. Eng. Chem., 1941, 33(1), pp. 575–580.
29. Reynolds Metal Co., "Finishes of Aluminum", 1946, pp. 51–52.
30. M. Schenk, *Werkschtoff Aluminum u. seine Anodische Oxydation*, A. Franke A. G. Bern, 1948.
31. G. D. Bengouh and J. M. Stuart, British Patent 223, 994, 1923.
32. K. Huber, J. Colloid Sci., 1948, 3, pp. 197–200.
33. S. Setoh and A. Miyata, *Proc. World Engineering Congress*, Tokyo, 1929.
34. 1994 *SAE Handbook*, Vol. 1, pp. 10–39.
35. W. W. Binger, E. H. Hollingsworth, and D. O. Sprowls, *Aluminum*, (K. R. Van Horn, Ed.), 1967, ASM, Ohio, Vol. 1, pp. 204–276.
36. E. G. West, J. Electrodepositors Technical Soc., 1946, 21, p. 211.
37. B. C. Schwarz and J. B. Newkirk, Plating, 1972, 59, p. 431.
38. J. T. N. Atkinson, J. Electrochem. Soc., 1958, 105, p. 24.
39. "The Surface Treatment and Finishing of Aluminium and its Alloys," *Plating on Aluminium*, Chap. 13, 5th Edn, Direct Plating, pp. 989–1019.
40. M. F. MacFadden, *Metals Handbook*, 8th Edn, 1964, Vol. 2, ASM, Ohio, pp. 424–431.
41. N. A. Solov'ev, USSR Pat. 252,806.
42. M. B. I. Janjua, J. Yernaux, P. J. Nicoll, and N. R. Bharucha, Plating, 1973, 60, pp. 1124.
43. Y. Lukomski, S. Lamina, and I. Gricoreva, Zash. Met., 1968, 4, 326.
44. J. T. N. Atkinson, U. S. Patent 3,003,933, 1961.
45. J. T. N. Atkinson, Plating, 1960, 47, p. 509.
46. C. L. Faust, G. R. Acha, and D. E. Semones, Plating, 1961, 48, p. 605.
47. G. R. Schaer, W. D. MacLean, C. L. Faust, K. C. Graham, and J. G. Beach, Brit. Pats. 877 386, 1961.
48. *Electroless plating* (*Basic and use*), Edited by Electrolytic Plating Research Group. June 1998, 5th Edn. Nikkan Kogyo, NDC566.78, Sn on Al, p. 223.

49. "Immersion Tin Processes," *The Surface Treatment and Finishing of Aluminum and its Alloys*, 5th Edn. pp. 1054–1071.
50. J. M. Bryan, *Metal Industry*, 1953, Dec. 4th, pp. 461–463; Dec. 18th, pp. 502–504.
51. Y. Y. Matulis, Russian Patent 265 652, 1970.
52. C. A. Hunt and H. Webb, (A. E. R. (1938) Ltd.), British Patent 579,443, 1948.
53. H. R. Watson and A. T. Ward, British Patent 626 693, 1949.
54. S. Heiman, *J. Electrochem. Soc.*, 1949, 95, p. 205.
55. "Tin Plating," *Surface Engineering Handbook*, 1999, K.K. Industrial Technology Service Center (Japan). Vol. 2, pp. 334–341.
56. "Zinc immersion processes," *The Surface Treatment and Finishing of Aluminum and its Alloys*, 5th Edn. 1985, Finishing Publications Ltd and ASM International, pp. 1023–1053.
57. E. H. Hewitson, (Eastman Kodak CO.) U.S. Patent 1,627,900, 1927.
58. W. G. Zelley, (Aluminum Company of America). U.S. Patents. 2,650,865, 1953; 2,676,916, 1954.
59. R. W. Leloup, Caning and Co. Ltd.). British Patent. 1,007,252, 1965.
60. J. W. Golby, J. K. Dennis, and A. E. Wyszynski, Trans. Surf, Technol., 1981, 67(1), p. 37.
61. W. G. Zelley, J. Electrochem. Soc., 1953, 100, p. 328.
62. K. Azumi, Y. Fujishige, M. Seo, L. Nanis, H. Nakao, and K. J. Tashiro, Surface Finishing Soc. Japan, 1996, 46(9), pp. 68–73.
63. J. W. Golby and J. K. Dennis, Surf. Technol., 1981, 12, p. 141.
64. J. W. Golby, J. K. Dennis, and A. E. Wyszyski, Trans. Inst. Metal Finishing, 1981, 59, p. 17.
65. H. J. Wittrock, Kaiser Aluminum and Chemical Corp., Technical Service Report No. 52–31, "Recommended Practice for Silver Plating Aluminum for Electrical Contacts," 1955.
66. W. P. Kampert, (Aluminum Company of America). U.S. Patent 655,737, 1951.
67. M. A. Bengston and J. M. Guthrie, Mon. Rev. American Electroplates' Society, 1947, 34, p. 421.
68. K. Maruyama, *Plating Guide Book*, Nikkan Kogyo N. P. Co., p. 49.
69. *Aluminum Handbook*, 5th Edn, Japan Institute for Light Metals. Chap. 14, "Surface Finishing," p. 166–167.
70. H. K. Work, J. Electrodepositors Techn. Soc., 1933, 8(7), p. 1.
71. General Development Corp. D.A.S. 1 226 392, 1966.
72. J. M. Bryan, Metal Industries, 1953, 83, pp. 461–463, and 502–504.
73. D. S. Tandem and J. Banjerjee, Indian Inst. Metals, Dec. 1957., Indian Patents 51,524 and 58,358, 1956.
74. J. Nat. Metall. Lab. Jamshepur, 1961, 1(3), p. 37.
75. "Chemical Etching Processes," *The Surface Treatment and Finishing of Aluminum and its Alloys*, 5th Edn., Finishing Publications Ltd. and ASM International, pp. 1001–1022.
76. A. L. Peach, (Micral Industries Inc.). U.S. Patent 3 531 379, 1970.
77. A. Brenner and G. Riddle, U.S. Patent 2 532 283, 1950.
78. D. J. Arrowsmith, and A. W. Clifford, Int. J. Adhesion and Adhesives, 1983, 3, p. 193, and 1985, 5, p. 40.
79. W. E. Cooke and J. Hodgeson, (Alcan Research and Development Ltd). British Patent 4,169,770, 1979.
80. Sperry Gyroscope Co. British Patent 636 880, 1950.
81. O. J. Klingenmaier, Plating Aug. 1974, pp. 741–746.
82. T. Gunji, T. Suzuki, N. Fujita, K. Dosaka, and Y. Toyoda, Honda R&D Technical Review. 1996, 8, pp. 194–203.

83. "Electrolytic Plating." *Surface Engineering Handbook*, 1999, K.K. Industrial Technology Service Center (Japan), pp. 334–341.
84. "Plating Technologies," *Surface Engineering Handbook*, 1992, K.K. Industrial Technology Service Center (Japan), 2, pp. 275–283.
85. E. Meyer, Raessler, and H. Huebnr, MTZ, 1972, 33(2), p. 48.
86. G. Stephan, H. Harst, G. Laux, and A. Kappes, U.S. Patent 3,514,389; 3,061,799.
87. A. E. Gazen, U.S. Patent 3,061,525.
88. K. Mier, VDI 1990 866, pp. 99–122.
89. S. Ishimori, M. Shimizu, S. Honda, S. Ohtsuka, and M. Toyoda, J. The Metal Finishing Soc. Japan. 1977, 28(10), pp. 508–512.
90. S. Ishimori, et al., Japan Patent. 1,076,729 (s. 56–18080).
91. K. Hasegawa, H. Takada, S. Yamaoka, T. Ito, and K. Yoshida, J. Surface Finishing Soc. Japan. 1989, 40(3), pp. 419–424.
92. H. Terada, M. Hirose, M. Takama, N. Konagai, J. JSAE, 1989, 43(5), pp. 81–86.
93. H. Takeuchi, Y. Tsunekawa, and M. Okumiya, Mat. Trans. JIM, 1997, 38(1), pp. 43–48.
94. H. Takeuchi, Y. Tsunekawa, M. Okumiya, and M. Tooyama, J. JIM, 1997, 61(7), pp. 595–601.
95. M. Isobe and H. Ikegaya, J. SAE-9439645.
96. M. Isobe and H. Ikegai, J. JSAE 9439645; J. JSAE, 1994, 48(11), pp. 89–96.
97. A. Brenner and A. Riddle, J. Res. N.B.S., 1946, 37, p. 1; 1947, 38, p. 385, U.S. Patent 2,532,283, 1950.
98. T. Ootaka, Uemura Journal 38, pp. 70–71.
99. Surface Technology 29/28, 29, *ASM Metals Handbook*, Desk Edn. 1985.
100. *Electroless Plating-Fundamental and Applications*. 3rd Edn. 1998, Nikkan Kogyo N.P.
101. A. Tronche and P. Fauchais, Materials Science and Engineering, 1988, A102 pp. 1–12.
102. R. C. Dykhuizen, J. Thermal Spray Tech., 1994, 3, pp. 351–361.
103. M. Fukumoto, S. Katoh, and I. Okane, *Proceedings of 14th International Thermal Spray Conference*, Kobe, 1995, pp. 353–358.
104. M. Fukumoto, T. Yokoyama, K. Oku, and Y. Tanaka, J. High Temp. Soc. 1997, 23, pp. 240–246.
105. N. Akashi and Y. Kojima, *Proc. 7th Int. Conf. Vacuum Metallurgy*, Tokyo, 1982, pp. 305–312.
106. Y. Tsunekawa, H. Hirai, M. Okumiya, and I. Niimi, J. Japan Thermal Spray Soc. 1989, 26, pp. 1150–1155.
107. G. Y. Liang and T. T. Wong, Surface Coating Tech. 1997, 89, pp. 121–126.
108. K. Nakata and M. Ushiroda, Surface Eng. 1997, 13, pp. 45–49.
109. G. Wuest, G. Barbezat, and S. Keller, SAE Tech. Papers, Detroit, No. 970016, 1997.
110. F. Rastegar and D. E. Richardson, Surface Coatings Tech. 1997, 90, pp. 156–163.
111. S. Tomida, K. Nakata, and F. Matsuda, J. Japan Inst. Light Metals. 1994, 44.
112. R. S. Zuchowski and R. P. Culberston, Welding Journal, 1962, 41, pp. 548–555.
113. H. Hallen, H. Mathews, A. Ait-Mekideche, F. Hettiger, U. Morkramer, and Lugscheider. *Proc. 13th Int. Thermal Spray Conf.* 1992, pp. 899–902.
114. A. Saltzmann, Surface Modified Technologies III, TMS, 1990, pp. 777–792.
115. K. Nakata, T. Hashimoto, K. Lee, and F. Matsuda, Materials Science Forum, 1996, pp. 217–222, 1655–1660.
116. T. Tomita, Y. Tanioka, G. Hashizume, and Y. Harada, Mater. Trans. JIM, 1993, 34, pp. 69–75.
117. J. M. Pelletier, S. Bonnet-Jobez, A. B. Vannes, and P. F. Gobin, IITT Int. Laser 4: High power lasers in metal processing, 1989, pp. 143–152.
118. A. Matsuda and K. Nakata, *Proc. 2nd Int. Conf. Power Beam*, TWI. 1990, pp. 297–306.

119. P. Petrov, R. Vilar, A. Almeida, *8th Int. Conf. Surface Modification Tech.* The Institute Materials, 1995, pp. 345–353.
120. K. Nakata and Tomida. J. Japan Inst. Light Metals, 1995, 45, pp. 578–584.
121. A. Wang, C. Fan, W. Huang, and K. Cui, J. Mater. Engin. Performance, 1996, 5, pp. 775–783.
122. N. Yasunaga, A. Obara, S. Minata, and M. Ikeda,: *Proc. LAMP'87*, Osaka, 1987, pp. 485–490.
123. H. J. Hegge, J. Boetje, and J. M. D. Hosson, J. Mater. Sci. 1990, 25, pp. 2335–2338.
124. M. Kawasaki, K. Takase, S. Kato, M. Nakagawa, K. Mori, M. Nemoto, S. Takagi, H. Sugimoto, SAE Tech Papers 920571, 1992.
125. N. Mohri, H. Momiyama, N. Saito, and Y. Tsunekawa, *Proc. 10th Int. Symp. Electromachinging*, Magdeberg, 1992, pp. 587–593.
126. M. S. Shunmugam, P. K. Philip, and Gangagbar. Wear, 1994, 171, pp. 1–5.
127. R. N. Johnson and G. L. Sheldon, J. Vac. Sci. Tech., 1986, A4, pp. 2740–2746.
128. Y. Tsunekawa, M. Okumiya, N. Mohri, T. Kosugi, and I. Niimi, J. Japan Foundry Soc. 1993, 65, pp. 783–788.
129. Y. Tsunekawa, M. Okumiya, N. Mohri, and I. Takahashi, Mater. Sci. Eng. 1994, A174, pp. 193–198.
130. Y. Tsunekawa, M. Okumiya, N. Mohri, and E. Kuribe, Mater. Sci. Forum, 1996, pp. 217–222, 1665–1660.
131. Y. Tsunekawa, M. Okumiya, N. Mohri, and E. Kuribe, Mater. Trans. JIM, 1997, 38, pp. 630–635.
132. A. Matsuda and K. Nakata, *Proc. 36th Welding Society Meeting of WSJ* 1985, p. 86.
133. H. Tachikawa, H. Fujita, K. Oguri, T. Suzuki, M. Nakamura, and T. Arai, *Proc. 21th Conf. of JSHT*, 1985, p. 17.
134. B. Reinhold, H-J. Spies, *Proceedings 1st Int. Automotive Heat Treating Conf.*, 1998, ASM International, pp. 213–220.
135. B. Reinhold, J. Naumann, and H.-J. Spies, *Proc. 11th IFHT Congress.*, Oct. 1998, pp. 341–349.
136. Y. Yamada and H. Miura, *Proc. 90th Conf. of Japan Inst. of Light Metals* 1996, pp. 181–182.
137. M. Okumiya, Y. Tsunekawa, K. Wakida, and H. Abe, *Proc. 46th Conf. of JSHT*, 1998, pp. 49–50.
138. Y. Miyasaka, *Proceedings of "State of the Art In Surface Modification" Application of Fine Particle Peening for Surface Modification*, 1999, pp. 10–11, Nov. FPP SM Research Group, Japan.
139. "Vitreous Enameling," Chap. 18, *The Surface Treatment and Finishing of Aluminium and Its Alloys* VII, 1985, Finishing Publication Ltd. and ASM International, pp. 1151–1777.
140. J. C. Baily, Light Metals, 1951, 13, pp. 647–656.
141. Amalgamated Research and Development Co. Ltd. British Patent 655, 881, 1952.
142. N. H. Stradley, *Sym. on Low Temp. Glasses*, Metrop. New York Sec. of American Ceramic Soc., 25.1.57.
143. Minnesota Mining and Manufacturing Co., U.S. Patent 3,383,225, 1957.
144. *Handbook of Surface Engineering of Aluminum*, Sec. 10, Porcelain Enameling, pp. 285–303. Pub. Light Metal Publication, Japan.
145. M. Kawamura, K. Yoshida, and I. Aoki, Interfinish '80, pp. 313–317.
146. K. Ando, Y. Kawauchi, and K. Murayama, J. JSAE, 1992, 46(5), pp. 82–87.
147. T. Hashimoto, Toyota Technical Review, 1993, 43, p. 144.
148. K. Shibata, K. Kimura, and A. Suzuki, J. JSAE 1988, 42(9), pp. 1211–1215.
149. M. Murakami, T. Ibi, H. Nohira, Y. Maeda, and M. Nakata, JSAE 902214, Proc. 902, 1990–10, pp. 2.5–2.8.

150. K. Ando, Y. Kawachi, and K. Murakami, J. JSAE, 1992, 46(5), pp. 82–87.
151. A. Tanke, T. Suguyama, and A. Kotani, SAE Tech Paper 930881.
152. V. D. N. Rao, Ford Motors. U.S. Patent 5,239,955, 1993.
153. K. Saito, Y. Fuwa, M. Sugiyama, and M. Murakami, JSAE-9539716; J. JSAE, 1995–9. pp. 101–104.

14

Aluminum Nitriding

HEINZ-JOACHIM SPIES

Freiberg University of Mining and Technology, Freiberg, Germany

BERT REINHOLD

ALD Vacuum Technologies AG, Hanau, Germany

1 INTRODUCTION

The poor hardness characteristics of aluminum and its alloys lead to less satisfactory tribological performance. This shortcoming can be overcome by means of load-specific surface engineering. An objective of research work is among others the application of nitriding treatment already proven on ferrous materials. Until now for aluminum alloys this method is still under investigation and not yet introduced in industrial practice.

The behavior of the light metals aluminum and titanium, and of iron, with regard to nitrogen and oxygen reveals marked differences. In order to use the potential of nitriding for improving tribological performance, care has to taken in relation to the effect on the resulting nitriding behavior (Table 1). Nitrogen is virtually insoluble in aluminum alloys. In nitriding these alloys, therefore only a compound layer of AIN is formed which is not supported, as is the case with ferrous alloys, by a hard precipitation layer. Nitriding temperatures, at 0.80–0.90 T_s[1] for aluminium alloys are clearly higher than those for ferrous alloys. They are clearly above the annealing temperatures (Table 2). Taking note of the constitutional diagrams it is evident that there is a more severe effect on structure and properties of the basic material. In order to support the hard nitrided layer and to regenerate the core structure, duplex surface treatment and post-heat treatment processes are of particular significance in the case of aluminum materials.

Given the high oxygen affinity of the light metals, which is evident in the extremely low oxidizing limit (Table 1), the demands for very pure reaction gases and low plant leakage are correspondingly high.

---

[1]The reference magnitude is the solidus temperature of the alloys.

Table 1 Properties of Aluminum, Titanium and Iron with Technological Relevance and Effect on the Plasma Nitriding Behavior

	Aluminum	Titanium	Iron
Max. solubility			
• of nitrogen [at%]	0	α-Ti: 27	α-Fe: 0.4
• temperature [K]		1376	865
Diffusion coefficient			
• of nitrogen [m^2 s^{-1}]	0	α-Ti: $2.6\cdot10^{-12}$	α-Fe: $7.4\cdot10^{-12}$
• temperature [K]		1173	823
Oxidation limit[1]			
• p(O$_2$) [bar]	$1.6\cdot10^{-70}$	$2.3\cdot10^{-33}$	$1.9\cdot10^{-27}$
• p(H$_2$O)/p(H$_2$)	$1.0\cdot10^{-20}$	$6.0\cdot10^{-9}$	0.3
• temperature [K]	723	1173	823
Nitriding temperature, T$_N$/T$_s$	0.75–0.90	0.55–0.65	0.4–0.5
Discharge parameters	Decrease of the current density by the high electr. Resistivity of AIN	No significant changes by the formation of the nitrid layer	
Case structure	Nitrid layer, no diffusion zone	Nitrid layer, diffusion zone	Nitrid layer, precipitation zone
Hardness			
• Nitrid layer [HV0.005]	>1200	>1500	750...1250
• Diffusion resp. Precipitation zone [HV0.05]	–	700...900	500...1500
• Core [HV1]	50...100	150...400	200...480
Influence of the core structure at T$_N$	Age hardening loss, recrystallization	Age hardening loss, recrystallization	Recrystallization, tempering

[1]Calculated with thermodynamic values. (From Ref. 13.)

Table 2 Solidification Range and Treatment Temperatures of Selected Al-Alloys

Alloy	Solidification range (°C)	Treatment temperatures (°C)			
		Annealing	Solution treatment	Nitriding	T_N/T_S
AIMg3 5754	610–640	330–360	—	400–460	0.76–0.83
AIMg4.5Mn 5083	574–638	380–420	—	400–460	0.79–0.86
AlCuMgl 2017	512–650	350–380	495–505	400	0.86
AIMgSil 6082	585–650	340–380	525–545	400–460	0.78–0.85
AlZnMgCu1.5 7075	480–640	420–460	470–480	400	0.89
AlSil0Mg 360.0	575–620	—	520–535	460	0.86

2 NITRIDING TECHNOLOGY

Plasma nitriding of aluminium alloys has been the subject of intensive investigation since the mid-80s [1–6]. It has emerged that also these materials can be successfully plasma nitrided. Equal success has been obtained with nitriding trials using simple d.c. [2], thermionic-assisted d.c. [3], and pulsed d.c. discharge plasma [4], as well as high-frequency [5] and microwave discharge [6] plasma. Also plasma immersion ion implantation (PIII) has been applied and has lead to the formation of a nitrid layer as well [7] (Table 3).

A significant problem in the nitriding of aluminum alloys is the naturally occurring oxide layer which prevents the diffusion of nitrogen into the metal matrix, and has therefore to be removed by sputtering. Work to date has been concerned with the need for a sputter process carried out directly before the nitriding treatment. In order to avoid reoxidation of the free aluminum surface, the facilities used have been evacuated to very low ultimate pressure. In practice, it has been shown that even under extreme treatment conditions, complete removal of the oxide layer and avoidance of reoxidation is not possible [11]. In view of the very low formation pressure of aluminum oxide this is understandable (Table 1).

Additional problems arise in relation to the physical properties of aluminum nitride. As can be seen from Table 4, the coefficient of thermal expansion of aluminum and aluminum nitride differs considerably. From this difference, very large residual compressive stresses (> 2 GPa) result which, in thick layers, can lead to cracking and spalling. For that reason, treatment with the objective of forming the thickest possible layer is not sensible. Much more important is the establishment of treatment conditions for achieving crack-free layers of homogenous thickness on structural components of complex geometry. With layer growth, the high electrical resistance of aluminum nitride requires adaption of the plasma parameters. The

Table 3 Plasma Nitriding of Al Alloys – Equipment and Process Parameters

Process	Conventional d.c.	Pulsed d.c.	Pulsed d.c. + r.f.	Thermionically assisted d.c. triode	Thermionic arc	Microwave + ECR	Plasma immersion ion implantation
Base pressure	10^{-3}	10^0–10^{-1}	10^{-3}–10^{-4}	2×10^{-3}	6×10^{-4}	2×10^{-4}	?
Sputter cleaning							
Time (h)	1	—	?	0.5	1.5	0.5	—
Pressure (Pa)	90	—	200	3.3	0.4	1.0	—
Bias voltage (V)	600	—	400–800	2000	80	1000	—
Gas	Ar	—	Ar–H$_2$	Ar	Ar–H$_2$	Ar	—
Nitriding							
Time (h)	≤20	4–12	d.c.?/r.f. ≤2	2–10	2–18	0.5	3
Pressure (Pa)	505	200–300	350/350	2.7	0.8	1	0.14
Bias voltage (V)	?	460–600	600/200	2000	50	1000	40000
Temperatur (°C)	450	400–460	480	300	340–460	500	300–500
Gas	N$_2$	N$_2$	N$_2$	N$_2$	N$_2$–Ar	N$_2$	N$_2$
AlN layer thickness (μm)	≤5	2–8	≤18	ca. 0.1	≤10	very thin	≤0.05
Reference	[1]	[8]	[5]	[3]	[9]	[10]	[7]

Table 4 Physical Properties of Al and AlN

	Al	AlN
Melting point [K]	933	3273
Density, ρ [g/cm^{-3}]	2.7	3.24
Young's modulus, E [GPa]	72	320
Hardness HV at load [N]	20–30/100	1530/1
Thermal expansion coefficient, α [10^{-6} K^{-1}]	23	4.7
Thermal conductivity, η (max) [Wm^{-1} K^{-1}]	226	320
Electrical resistivity, ρ [10^{-8} Ωm]	2.5	$\approx 10^{19}$

Source: Refs. 14 and 15.

nitriding temperature (Table 2) reaches 83–89% of the solidus temperature of the alloys investigated. Even under conditions of minor local overheating, therefore, melting can occur. This makes great demands on temperature control.

Starting from the assumption that reoxidation of an aluminum surface cleaned by a sputtering treatment is in practical terms unavoidable, the possibility of forming aluminum nitride on a surface still covered with oxide have been examined. For these tests, a commercial plasma nitriding plant designed for d.c. pulsed plasma nitriding of ferrous alloys was used (Fig. 1). Nitriding was carried out at temperatures between 400°C and 460°C and at pressures of 200–300 Pa and a discharge voltage of 460–600 V in nitrogen or nitrogen argon mixtures.
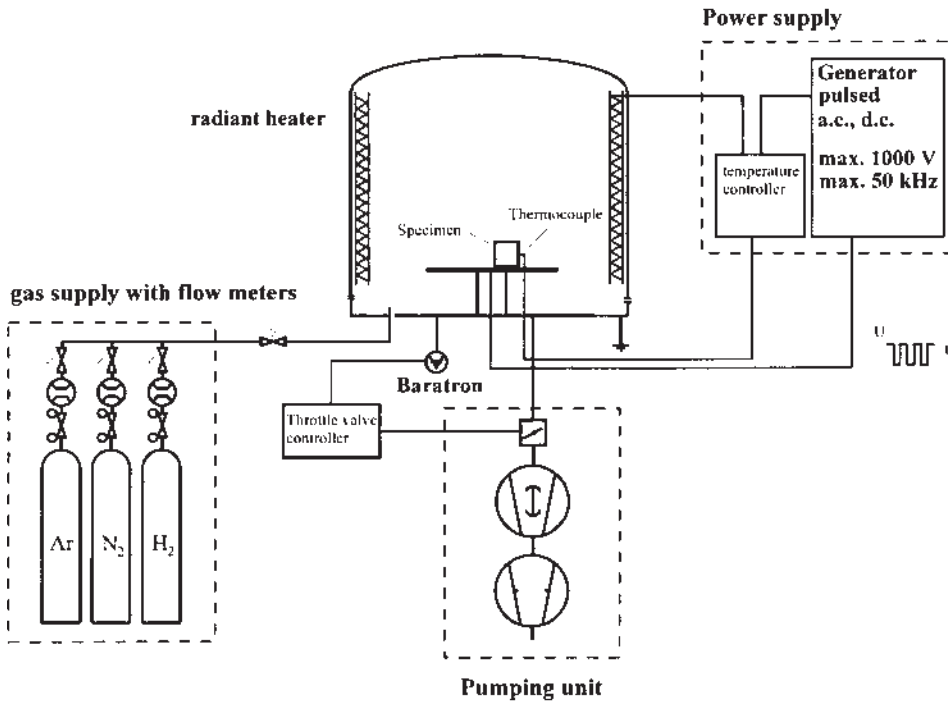
Figure 1 Scheme of the commercial plasma nitriding facility.

The experiments showed that the formation of aluminum nitride is possible even without prior sputtering. An example of an aluminum nitride layer thus produced is shown in Fig. 2. The mid zone of this layer is characterized by a nitrogen concentration of over 50 at%; the oxygen content is low, being 2% at maximum within the layer. AES investigation showed that the oxygen is present mainly in the form of magnesium oxide. Analysis of the chemical composition of the nitrided layers was carried out, after the testpieces had been covered with a thin layer of titanium by means of GDOS. This is the reason for the high titanium content at the very surface of the layer.

According to the results obtained, the energy of the nitrogen ions under the treatment conditions selected is sufficiently high to remove the oxide. The formation of the nitride takes place at the locations where the metallic surface has been sputter-cleaned, by physi- and subsequently chemisorption of atomic nitrogen. Despite the lower formation enthalpy in comparison with the oxide this occurres; since oxygen is only present as residual gas or as leakage, nitrogen available for reaction is, however, formed at high concentration in the plasma. These data lead to the conclusion that the AlN, once formed, is to a significantly lesser extent sputtered than Al_2O_3, probably due to the greater interatomic forces (covalent bonding). That would explain the formation of a closed nitride layer as the treatment
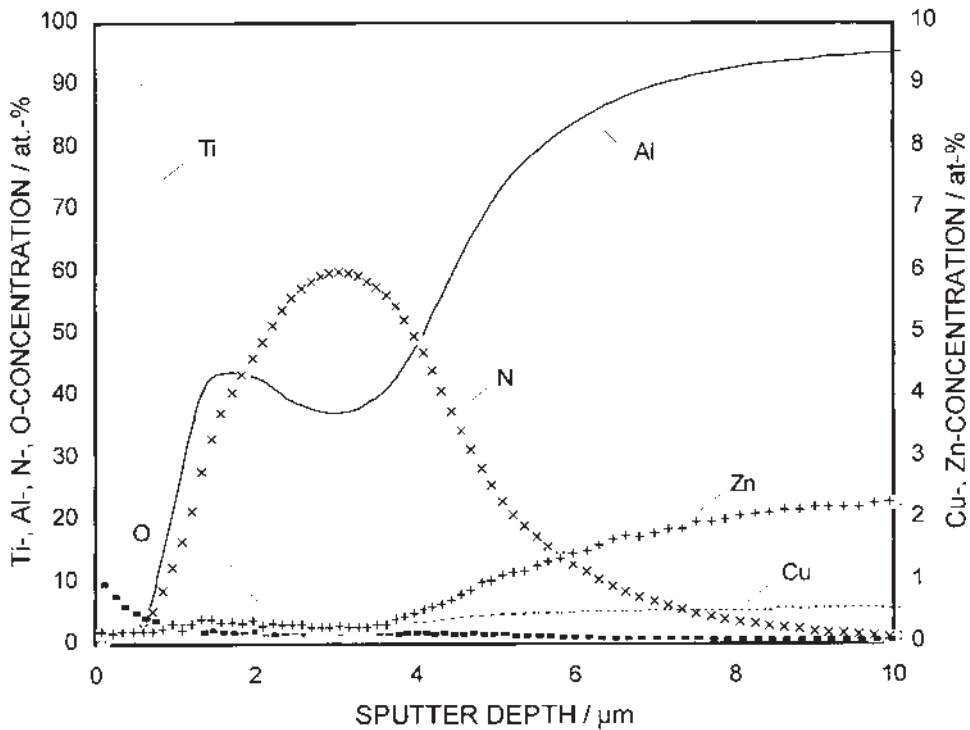
Figure 2   Concentration depth profile (GDOS) of a nitrided case; alloy 7075 $T_N = 400°C$; $t_N = 8$ h.
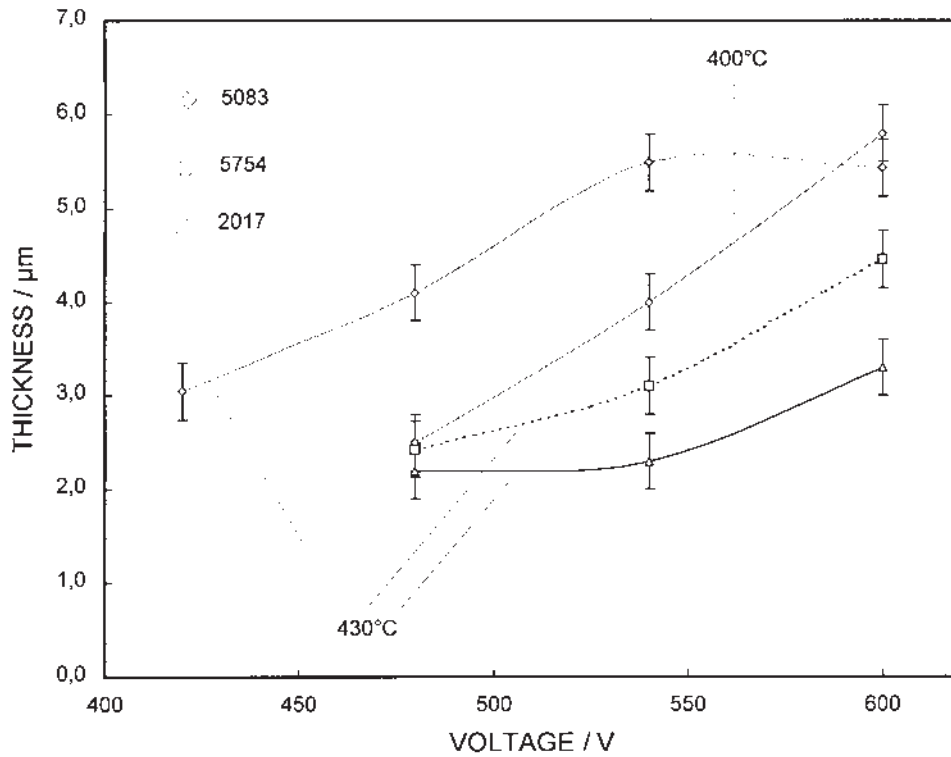
Figure 3 Thickness of the AlN layer versus discharge voltage for different alloys.

time moves on. A further advantage of omitting the sputter pretreatment is a lower surface roughness by avoiding the formation of Al_2O_3-cones that is attributed to the great difference in the sputter yield between aluminum and aluminum oxide [4].

An increase in the sputter yield and hence in the growth rate is also possible by increasing the ion energy and/or the mass. The ion energy can be increased by increasing the discharge voltage at constant pressure. That leads, in association with the simultaneous increase in the pulse current density, to a significantly thicker layer (Fig. 3). A similar effect is observed when the ion mass is increased by adding argon to the process gas (Fig. 4). The decrease in the layer thickness of alloy 5083 at a voltage of 600 V and a nitriding temperature of 430°C can be attributed to resputtering of AlN.

3 STRUCTURE AND PROPERTIES OF NITRIDED LAYERS

In almost all nitriding experiments the formation of a hexagonal close-packed (hcp) phase has been observed. This structure can be related to AlN in its equilibrium state (Wurtzit) (Fig. 5). But also a face centered cubic (fcc) phase could be detected [3].

The nitride layers are characterised by high hardness. On a cross-section of a sufficiently thick layer, a plastic universal hardness of HU_{plas} 0.05 $=$ 16 GPa is measured. As a first approximation, this corresponds to a Vickers hardness of 1600
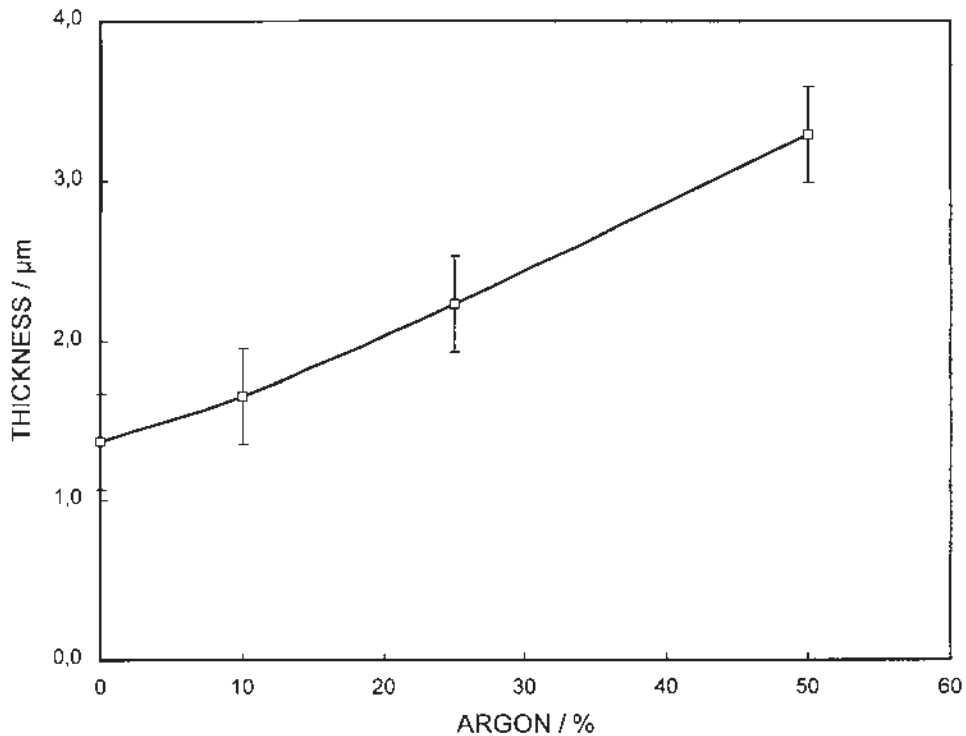
Figure 4   Relation between layer thickness and argon concentration; 6082; 460°C/4h.

HV 0.005. An indication of the load-bearing capability of the layer compound can be obtained from a measurement of the universal hardness penetration depth curves. As can be seen from Fig. 6, the compound hardness increases significantly with layer thickness and the increase in core hardness due to a heat treatment. Nitriding leads to a significant reduction in adhesive wear, characteristic of most of the aluminum alloys, as long as the contact pressure is limited to avoid layer breakthrough. As has been already shown by Arai et al. nitrided aluminum alloys exhibit a outstanding wear resistance that is even comparable with those of conventionally treated as for example hardanodized material [1]. Chen et al. have reported an improvement in the wear resistance by nitriding of about a factor of ten [4].

4   NITRIDING BEHAVIOUR OF ALUMINUM ALLOYS

The growth behavior of the nitride layers is illustrated in Fig. 7 and 8. It is clearly visible, that the growth rate of the nitride layers increases with increasing magnesium content of the alloys. As can be seen from Fig. 9, the oxide layer of the magnesium-containing alloys consists to a great extent of MgO as a result of the high oxygen affinity of the magnesium. The magnesium oxide portion increases, as would be expected, in proportion to the magnesium content of the alloy (Fig. 9). Magnesium oxide has a much lower surface binding energy in comparison with
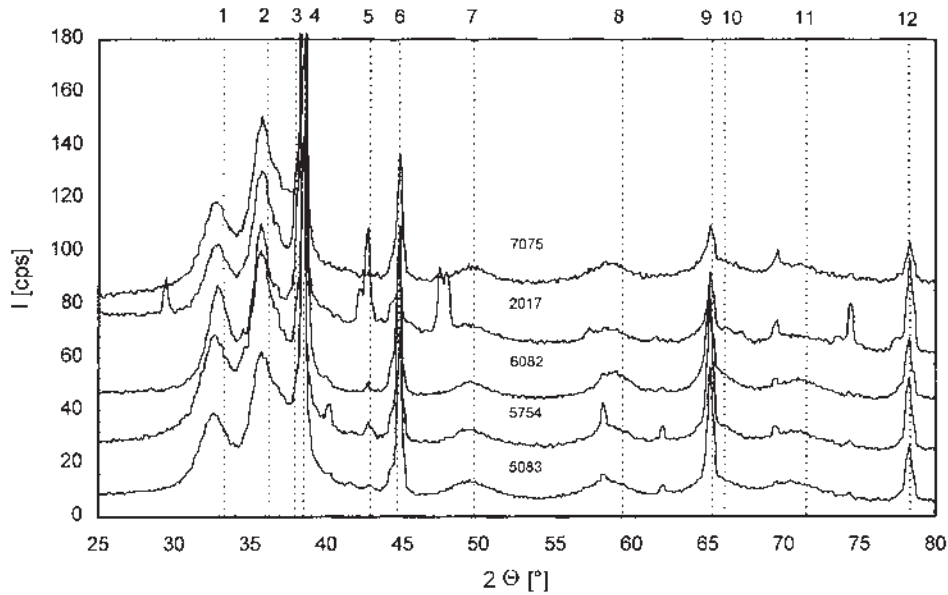
	1	2	3	4	5	6	7	8	9	10	11	12
AlN	(1,0,0)	(0,0,2)	(1,0,1)				(1,0,2)	(1,1,0)		(1,0,3)	(1,1,2)	
Al				(1,1,1)		(2,2,0)			(2,2,0)			(3,1,1)
MgO					(2,0,0)							

Figure 5 X-ray diffraction pattern of different nitrided alloys.

Al_2O_3 and is therefore easier to sputter [8]. This explains the effect of the magnesium content on the AlN layer growth. The thickness of the nitride layer depends, in a linear manner on the nitriding time as has been observed also by others [1,9,12]. Renevier et al. believe that, in spite of the high affinity between aluminum and nitrogen, the formation reaction of AlN is the rate determining step. They estimated the activation energy for the chemical reaction to be ca. 1 eV [9]. Blawert and Mordike suppose that aluminum nitride can only be formed if a certain minimal energy is delivered by the incoming ions [7]. This is a further indication that the AlN formation is to a great extent governed by the interaction with energetic particles.

Nitriding is, as would be expected, associated with a reduction in core hardness. Quantitative data on this are given in Table 5. The hardness loss is especially pronounced in the age-hardening alloys. Solution treatment and aging should be carried out on components made from these alloys, therefore, after the nitriding treatment. Many investigations have confirmed that the layer adhesion will not be impaired by subsequent age-hardening cycle. The core strength of the naturally hard alloys cannot be raised further after nitriding. Fortunately these alloys exhibit, along with a comparably small reduction in strength, the highest layer growth rates as a result of their higher magnesium content.
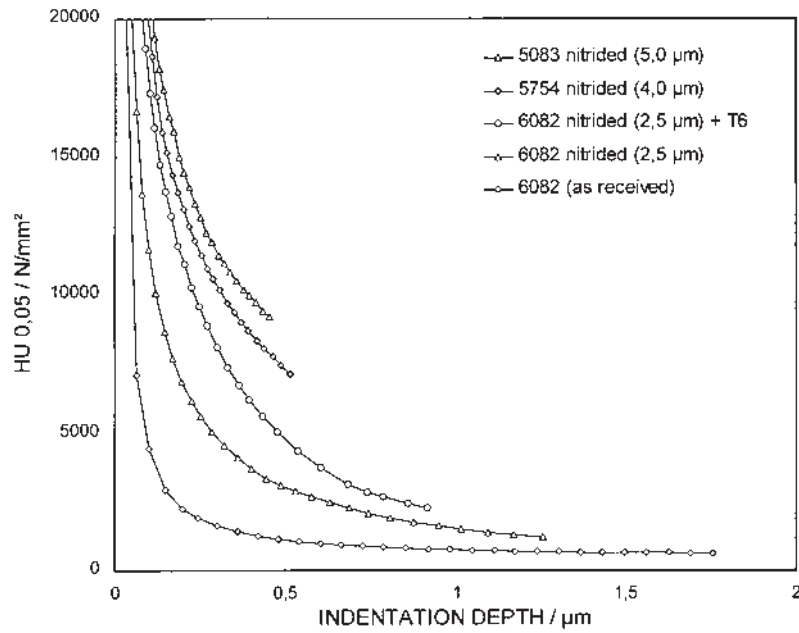
Figure 6   Hardness-indentation depth-profiles measured on various alloys in different states of treatment.
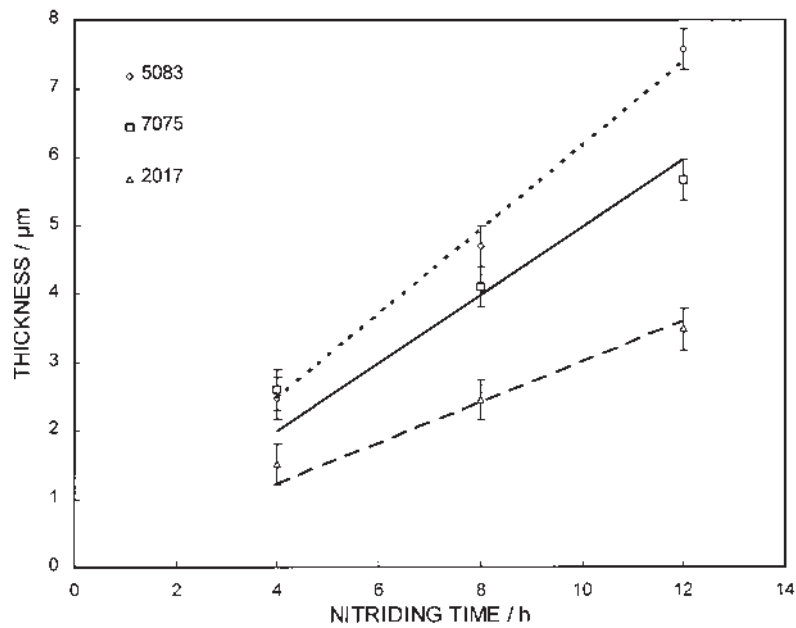
Figure 7   Growth of the AlN layers as a function of the effective nitriding time for different alloys, $T_N = 400°C$.
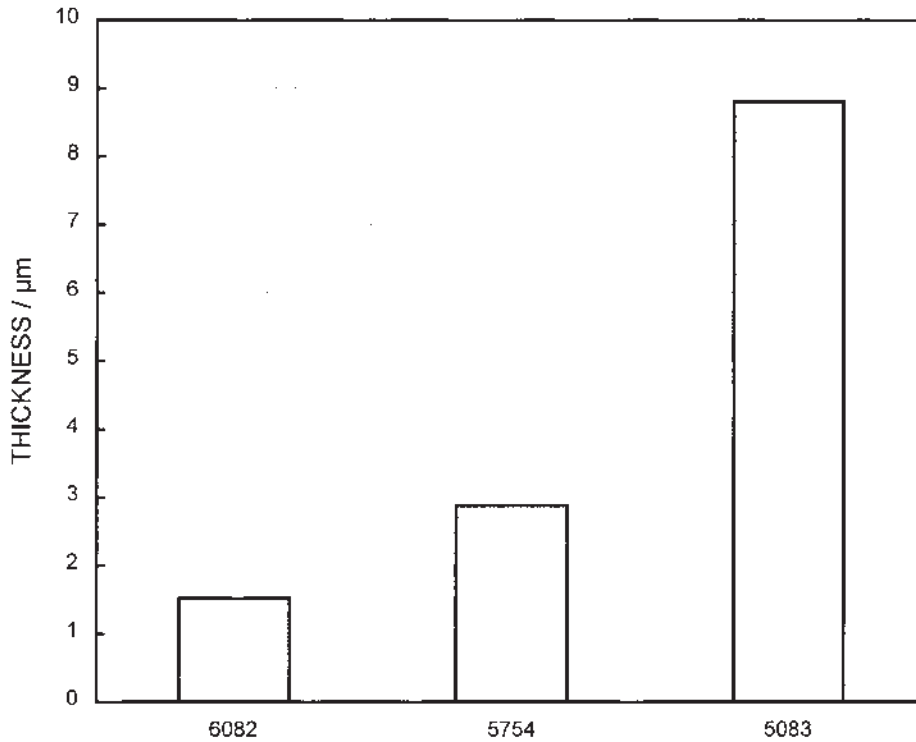
Figure 8 Thickness of the AlN layers on different alloys; $T_N = 460°C$; $t_N = 4$ h.

As is well known from plasma nitriding of steel, the uniformity of compound and diffusion layer depends on the uniformity of the cathode glow. Due to the fact that sputter removal of oxide layer is absolutely required in the case of aluminum, the uniform plasma coverage is even of greater importance. On cylindrical specimens it was observed that the lateral area had been uniform covered with nitride first. On the top surface, AlN was formed later and with a gradient in thickness. Generally, on edges nitride is formed fastest, because of intensive ion bombardment

5 SUMMARY

It can be asserted that considerable progress has been made in the field of plasma nitriding of aluminum alloys in recent years. An essential part of this is the evidence that plasma nitriding of aluminum and its alloys can be carried out in commercial in plants of the type used for the plasma nitriding of ferrous and titanium alloys. As in the case of other groups of material, the aluminum alloys exhibit significant material-specific differences in nitridability. Successful nitriding processes requires knowledge and taking into account these differences. Nitriding temperatures are, in comparison from to the solidus temperatures of aluminum alloys, very high. To avoid localized melting, therefore, a very precise, flexible and responsive tem-
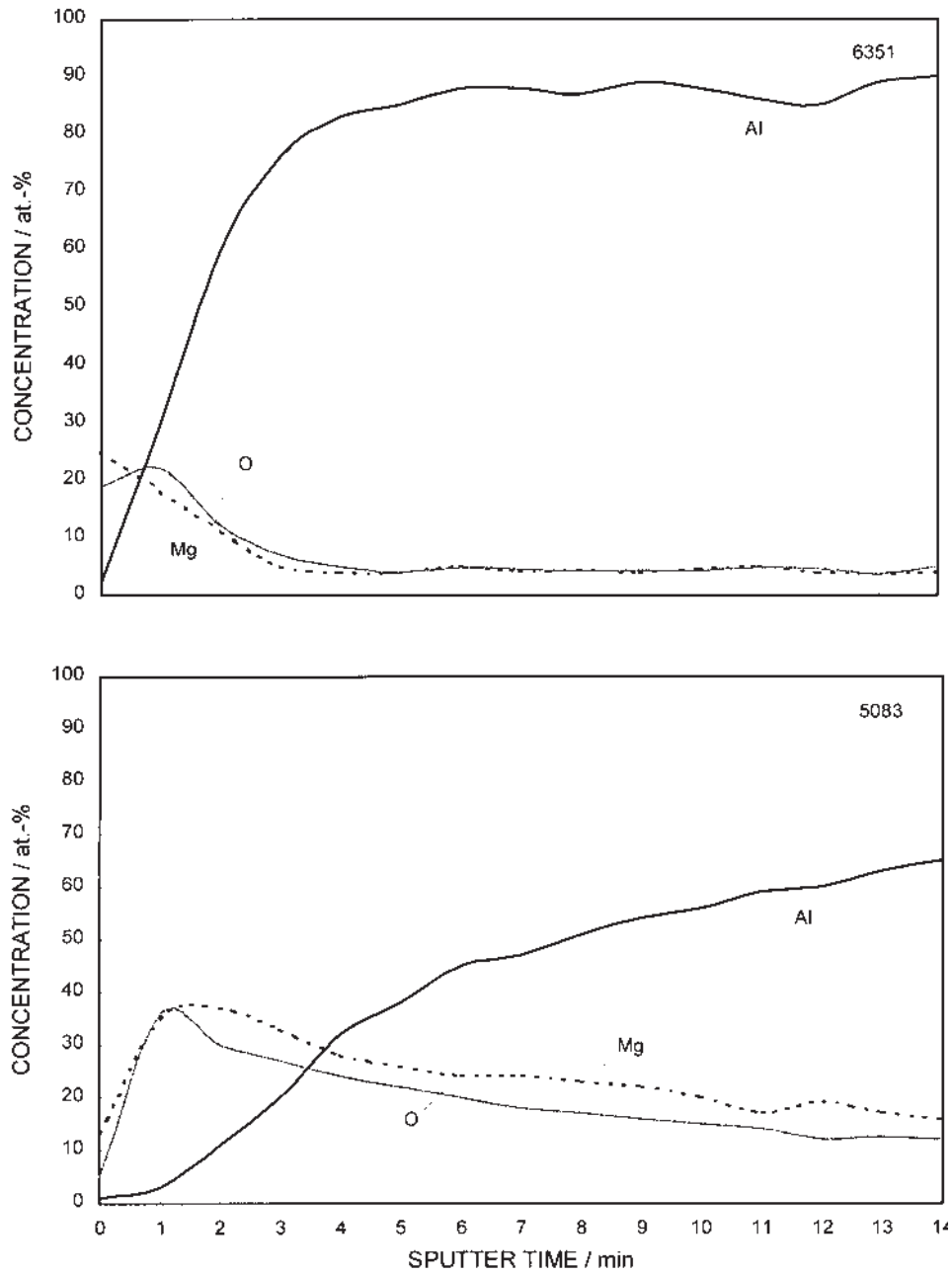
Figure 9 Concentration depth profiles (AES) in cases of the alloys 6351 and 5083 in as received condition.

perature management is a necessary prerequisite. The application of nitrided aluminum alloys under tribologically loads with high contact stresses requires the thin hard nitride layer to be supported by a duplex treated case.

Table 5 Nitriding Influence on the Core Hardness of Al Alloys

Alloy	Hardness [HV0,2]		Decrease in hardness %
	as received	nitrided	
AlMg3 5754	79	56–64	29
AlMg4.5Mn 5083	92	76–82	17
AlCuMg1 2017	148	63–69	57
AlMgSi1 6082	78	26–41	67
AlZnMgCul.5 7075	97	66–70	32
AlSi10Mg 360.0	72	48–52	33

REFERENCES

1. T. Arai, H. Fujita, and H. Tachikawa, ''Ion Nitriding of Aluminum and Aluminum Alloys,'' in *Proc. 1st Intern. Conf. Ion Nitriding*, 1986 Cleveland, OH, pp. 37–41.
2. I. Kanno, K. Nomoto, P. Nishijima, T. Nishiura, T. Okada, K. Katagiri, H. Mori, and K. Iwamoto, ''Tribological Properties of Aluminum Modified with Nitrogen Ion Implantation and Plasma Treatment,'' Nucl. Instr. Meth. Phys. Res. B 1991, 59/60, pp. 920–924.
3. E. I. Meletis and S. Yan, ''Formation of Aluminum Nitride by Intensified Plasma Ion Nitriding,'' J. Vac. Sci. Technol. 1991, A 9, pp. 2279–2284.
4. H.-Y. Chen, H.-R. Stock, and P. Mayr, ''Plasma-Assisted Nitriding of Aluminum,'' Surf. Coat. Technol. 1994, 64, pp. 139–147.
5. C. Jarms, H.-R. Stock, and P. Mayr, ''Nitriding of Aluminum in hf Plasma,'' HTM, 1996, 51, pp. 113–118.
6. T. Ebisawa and R. Saikudo, ''Formation of Aluminum Nitride on Aluminum Surfaces by ECR Nitrogen Plasmas,'' Surf. Coat. Technol., 1996, 86/87, pp. 622–627.
7. C. Blawert and B. L. Mordike, ''Plasma Immersion Ion Implantation of Pure Aluminum at Elevated Temperatures,'' Nucl. Instrum. Methods Phys. Res., B, 1997, 127–128, pp. 873–878.
8. B. Reinhold, J. Naumann, and H.-J. Spies, ''Effect of Composition and Component Geometry on the Nitriding Behavior of Aluminum alloys,'' HTM, 1998, 53, pp. 329–336.
9. N. Renevier, T. Czerwiec, A. Billard, J. von Stebut, and H. Michel, ''A Way to Decrease the Aluminum Nitriding Temperature: The Low-Pressure Arc-Assisted Nitriding Process,'' Surf. Coat. Technol., 1999, 116–119, pp. 380–385.
10. T. Hino, I. Fujita, and M. Nishikawa, ''Nitriding of Zirconium and Aluminum by using ECR Nitrogen Plasmas,'' Plasma Sources Sci. Technol., 1996, 5, pp. 424–428.
11. H.-R. Stock, H.-Y. Chen, and P. Mayr, ''Plasma Nitriding of Aluminum Materials – Potential and Limitations of a New Method Part 1: Surface Cleaning,'' Aluminium, 1994, 70, pp. 220–228.
12. H. Tachikawa, H. Fujita, and T. Arai, ''Growth and Properties of Nitride Layers Produced by Ion Nitriding'' Surf. Eng. Int. Conf. Tokyo, October 18.–22, 1988, Japan Thermal Spraying Society, Osaka, 1988, pp. 347–352.

13.   O. Knacke, O. Kubaschewski, and K. Hesselmann, *Thermochemical Properties of Inorganic Substances*, 1991, Springer Verlag Berlin/Verlag Stahleisen m.b.H. Düsseldorf, 2. Aufl.

14.   O. Madelung, (Ed.), *Landolt-Börnstein*, Vol. 17, 1982, Springer Verlag, pp. 158–161.

15.   G. Subhash and G. Ravichandran, ''Mechanical Behavior of a Hot Pressed Aluminum Nitride under Uniaxial Compression,'' J. Mat. Sci., 1998, 33, pp. 1933–1939.

15

Friction Stir Welding of Aluminum Alloys

ANTHONY P. REYNOLDS

University of South Carolina, Columbia, South Carolina, U.S.A.

1 INVENTION AND CURRENT USE

Friction Stir Welding (FSW) was developed at and patented by The Welding Institute (Cambridge, UK) in 1991 [1]. Since the time of its invention, the process has been continually improved and its scope of application expanded. Friction stir welding is a solid-state joining process combining deformation heating and mechanical work to obtain high quality, defect-free joints. Friction stir welding is especially well suited for joining aluminum alloys in a large range of plate thickness and has particular advantages over fusion welding when joining of highly alloyed aluminum is considered: high quality joints may even be made in discontinuously reinforced aluminum alloys (aluminum MMCs) [2–4].

Because of the many demonstrated advantages of FSW over fusion welding techniques, the commercialization of friction stir welding is proceeding at a rapid pace. Current production applications include both large- and small-scale products. For example, Boeing is using FSW to assemble oxidizer and fuel tanks for Delta Rockets [5], Marine Aluminum in Norway joins large extrusions for fast ferry decks using FSW [6] and Lockheed-Martin Corporation is in the process of implementing FSW for production of the external tank of the space shuttle [7]. Perhaps the most ambitious project is the development of a friction stir welded business jet (fuselage and wings) by Eclipse Aviation [8].

Much of the work done to bring FSW to production applications has been of a very practical nature, driven primarily by the pressing industrial need and the information generated has often been shielded from general dissemination by proprietary considerations. Little information of a scientific nature has been

published in the open literature, although this trend is now beginning to change and a substantial number of papers are being published. The proliferation of conference sessions devoted to FSW is further evidence of the increasing importance of the technology. The proceedings of the TWI sponsored International Symposia on FSW (1999, 2000, and 2001) are excellent sources of information on the emerging aspects of friction stir welding of aluminum and other alloys. The purpose of the following is to introduce the reader to friction stir welding and provide an understanding of the underlying process physics as well as a flavor for the capabilities of the process.

2 FSW PROCESS BASICS

In principle, friction stir welding is a very simple process. The two plates to be welded are butted together (lap and other configurations are also possible) and clamped to a rigid backing plate. The rotating FSW tool is plunged into the plates at the joint line and traversed along the line, forming the joint. At its simplest, an FSW tool consists of two concentric cylinders as shown in Fig. 1. The pin is the smaller diameter cylinder and may be threaded, fluted or have a variety of other detail features machined into it. The length of the pin is the primary determinant of the depth of penetration of the friction stir weld. In order to make a full penetration butt weld, the pin is typically several thousandths of an inch shorter than the plate thickness. The larger diameter cylinder is the shoulder. The tool is plunged into the workpiece until the shoulder makes contact with the crown of the weld with sufficient pressure to confine material within the weld zone and produce substantial deformation heating in the workpiece as the shoulder rotates. The shoulder diameter is typically two to three times that of the pin; however, this is a rule of thumb and not cast in stone.

Friction stir welding tools for the welding of aluminum are often made of oil hardening tool steels although other materials (e.g. nickel-based or ferrous-based superalloys) are applicable as well. In ''conventional'' FSW, the tool is set at some small lead angle, φ, from the workpiece normal and the shoulder is slightly concave. The lead angle is typically from $1°$ to $3°$ and is used to prevent the tool from ''plowing'' into the workpiece on the front or leading edge and provide some forging force at the rear or trailing edge of the tool. The concavity is designed to provide a reservoir of material above the original crown surface of the weld, facilitate transport of material around the tool and reduce plate thinning in the weld zone.

In all friction stir welds, the tool rotation rate and the tool travel rate (the welding speed) must be controlled. In addition, either the vertical position of the tool or the downward (z-axis) force on the tool must be controlled. If the vertical position of the tool is controlled directly, the weld is considered to be performed in position control. The controlled position may be referenced either to the crown side (typical for a lap or partial penetration welds) or the root side (necessary for the production of a full penetration butt weld). If the vertical position of the tool is indirectly controlled by the applied vertical or z-axis force, then the weld is a force or load controlled weld. Friction stir welding was originally performed on machine tools, typically vertical mills, and as such, most early welding was performed in a nominally position controlled mode. With the development of purpose-built FSW equipment, a trend toward primarily z-axis force control welding may be observed in current practice. Once proper force levels are established, force control is more
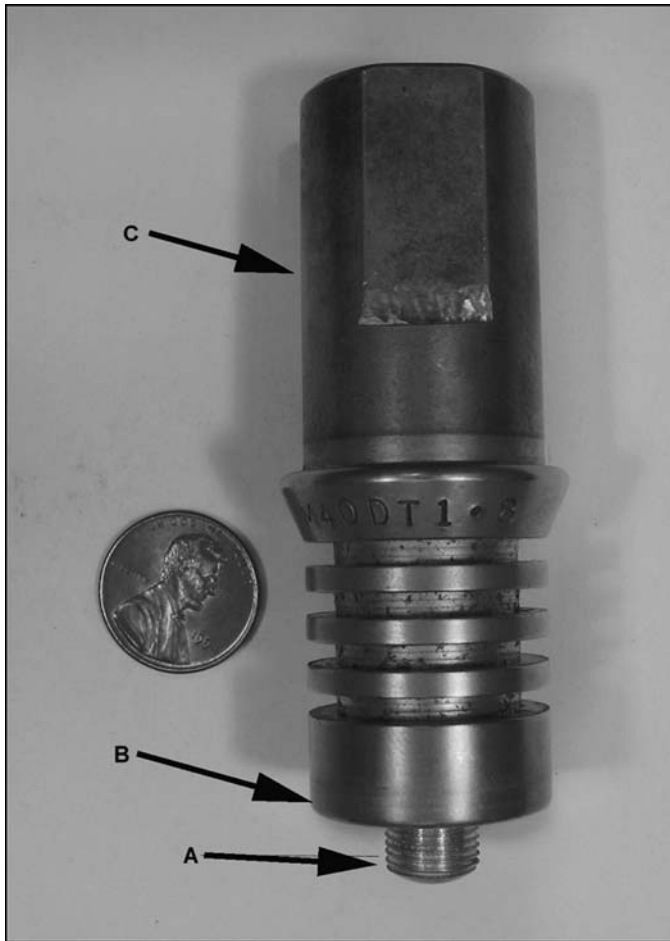
Figure 1 An FSW tool for welding aluminum plates of approximately 8 mm thickness. The pin, A, the shoulder, B, and the shank, C are illustrated. The tool is made of H13 tool steel and is coated with TiN (the pin has a coating of aluminum left over from previous welding operations).

accommodating of slight variations in process variables such as workpiece thickness. It should be noted that even welds made using position control act, to some extent, as force controlled welds due to the finite stiffness of the machines that are used to produce them.

The production of high quality, defect-free joints by FSW is dependent on proper choice of the three control variables: tool rpm, welding speed, and z-axis force (or position). Appropriate values for these variables depend on the alloy being welded, the thickness of the plate or sheet, and the particular tool geometry used. Of course, values for z-axis force, tool rotation rate and welding speed cannot be chosen independent of each other. The factors that must be considered when choosing appropriate FSW process parameters include not only the ability to make a defect-free weld but also the machine capabilities and the strength of the welding

tool. For example, it may be possible to create an excellent weld using a very large z-axis force and a low tool rotation rate. However, if this combination of parameters is outside the capabilities of the welding machine, either due to excessive force or torque requirement, then the parameters must be adjusted to accommodate the machine's deficiencies. In addition, the welding process exerts a force on the tool in the direction of welding. This so-called x-axis force is generally observed to increase with increasing welding speed [9]. At some point, the bending moment on the tool will cause it to break. Again, the welding parameters must be adjusted to account for the limitations of the equipment.

Equipment for friction stir welding is sized primarily on the basis of the desired thickness of plate to be welded. z-Axis forces may range from a few hundred pounds for 1—2 mm thick sheet to greater than 10,000 lb for a plate of 0.5″ or more. Required torque may also vary widely depending on the alloy and the tool geometry. Most FSW of aluminum is performed using tool rpm of between 100 and 1500. For some high speed welding of thin sheet, higher rpm may be used. Welding speeds as high as 60 in./min have been obtained for thin sheet [10], but all other things being equal, the welding speed is limited by the plate thickness: the thicker the plate, the lower the maximum possible welding speed.

As stated previously, standard machine tools (milling machines) may be used for FSW but purpose-built machines allowing z-axis force control welding allow more flexibility and generally give more consistent results. Within the last few years, several companies have begun producing equipment designed specifically for FSW.

3   FSW STRUCTURE

While details of friction stir weld microstructure depend on the particular process parameters and alloy properties, there are several features common to essentially all aluminum alloy friction stir welds. In order to describe these features, it will be useful to first discuss some terminology. An important feature of friction stir welding is that the process is not symmetric about the weld centerline: the advancing side of the weld is defined as the side on which the rotational velocity vector of the welding tool has the same sense as the translational velocity vector of the tool relative to the workpiece. The retreating side is where the two vectors are of opposite sense. The leading side is the front of the tool and the trailing side is the back of the tool. The crown is the top surface of the weld (in contact with the tool shoulder during welding) and the root is the bottom surface.

The crown of a friction stir weld exhibits a very characteristic pattern of semicircular ridges. The ridges are concave in the direction of tool travel relative to the plate. On the advancing side of the crown of the weld, a lip forms where weld metal is rolled over onto the base metal. This lip is often called "flash" and can be minimized, but not eliminated, by a proper choice of welding parameters and tool design. At the termination of a friction stir weld is a keyhole that remains after the tool is extracted from the welded plate. The keyhole will have the approximate dimensions of the pin of the welding tool. The features described above are illustrated in Fig. 2.

Figure 3 shows the microstructure resulting from friction stir welding of AA2024-T3. The structure observed in this figure is representative of that seen in most friction stir welds. The weld features can be separated into two broad

(a)

(b)
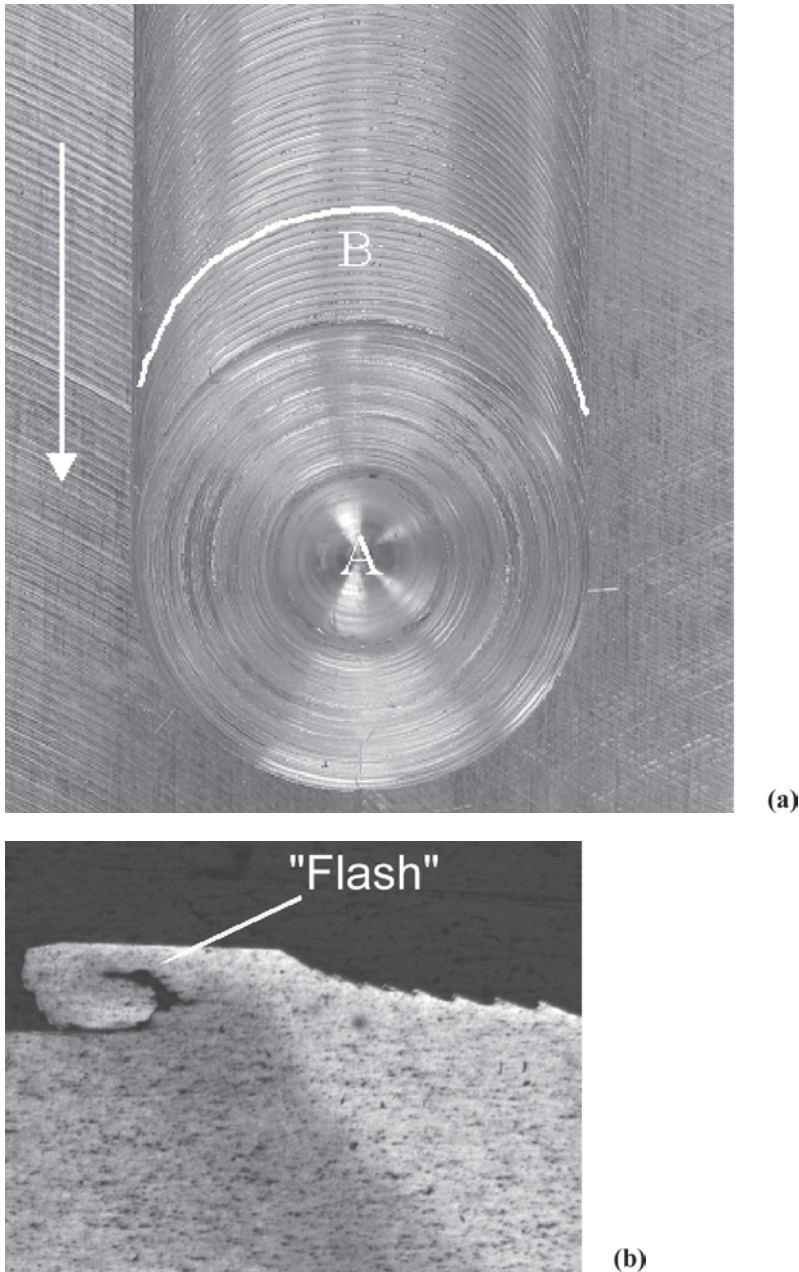
Figure 2 (a) Shows the termination of a friction stir weld. "A" is the location of the keyhole that remains at the end of the weld: the keyhole diameter is the same as the pin diameter. The curved line labeled "B" indicates a ridge left by the passage of the FSW tool shoulder. The spacing of the ridges is the same as the tool advance per revolution. The vertical arrow indicates the welding direction. The tools markings on the plate outside of the tool diameter are artifacts of the plate preparation and are not FSW related. In (b), the advancing side "flash" at the edge of the tool shoulder is shown in cross-section. The serrations to the right of the flash on the top surface of the weld are cross-sections of the ridges shown in (a).
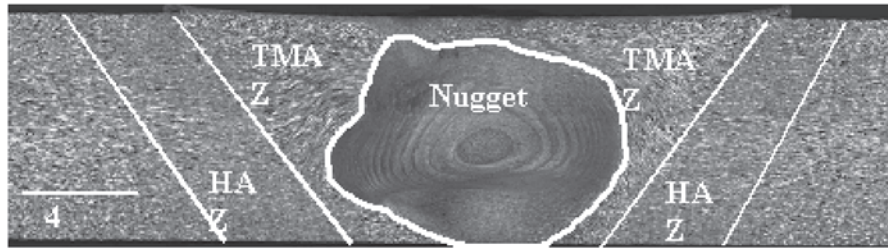
Figure 3 A macroscopic cross-section view of a typical FSW. Note the "onion-skin" pattern in the nugget region. The advancing side of the weld is on the left, the retreating on the right.

categories: the thermo-mechanically affected zone, or TMAZ, and the heat affected zone, or HAZ. The HAZ is similar to heat affected zones resulting from conventional, fusion welding processes. The extent of the HAZ will depend on the time–temperature profile experienced by the material as a function of the distance from the weld centerline: this is, in turn, a function of the spindle power, the welding speed, and the heat lost to the clamping fixtures.

The TMAZ of a friction stir weld might be considered analogous to the fusion zone of a conventional weld except that, instead of being melted, the material in the TMAZ has been mechanically worked. Within the TMAZ, there are three somewhat distinct regions. The most obvious is the "weld nugget". The weld nugget is the region which has undergone the most severe plastic deformation and is characterized by a fine, relatively equiaxed, recrystallized, grain structure. Typical grain size observed in the nugget ranges from 2 to 10 μm depending on the alloy and the welding conditions. The width of the nugget is normally similar to but slightly greater than the diameter of the pin. Outside of the nugget on either side, is a region which has been deformed to a lesser extent and which, depending on the alloy, may or may not show signs of recrystallization. In Fig. 3, the delineation between the nugget and the rest of the TMAZ is relatively sharp because of the recrystallization resistance of the base metal. The deformation of the base metal grains manifests itself as bending in the plane of the metallographic section (they are also bent in the horizontal plane perpendicular to the section). The third region of the TMAZ is the "flow arm". This is the region of material above the nugget. The flow arm is formed when the rotating tool shoulder passes over the weld. Again, it is worth noting that the friction stir welding process is not symmetric about the centerline: the advancing and retreating sides exhibit significant differences. In general, the transition from nugget to base metal is very sharp on the advancing side and more gradual on the retreating side. The region of lightly deformed material, outside of the weld nugget, is very narrow on the advancing side of the weld.

Another FSW feature that has been noted by almost all FSW practitioners is the presence of the so-called "onion rings" or the "bulls-eye" pattern which may be observed in optical cross-sections of most FSW weld nuggets. These rings may be more or less distinct depending on the alloy and process parameters used. However, the spacing of the rings in the longitudinal (welding) direction is always found to correspond to the tool advance per tool revolution (welding speed divided by the rotational rate). The mechanism of formation and significance of these rings are currently subjects of debate.

4 PHYSICAL PROCESSES OCCURRING DURING FSW

4.1 Material Transport

The transport of material around the tool during FSW has been visualized experimentally [11–15] and simulated mathematically [15–20]. In the flow visualization experiments, primary and secondary components of material flow have been identified. The primary flow, which transports the bulk of the material around the pin, is similar to laminar flow of a viscous fluid around a rotating cylinder. In order to produce a defect-free weld, sufficient material must be carried by the rotating pin, around the retreating side, to fill in behind the tool on the advancing side of the weld. Under most FSW conditions, it appears that all of the material flows around the retreating side of the pin along flow lines similar to those shown in Fig. 4. Figure 4 is of course only a two-dimensional representation of the flow and ignores three-dimensional effects that arise from the action of the tool shoulder and the threads on the pin.

Pin threads are normally designed in such a way as to push material down toward the root of the weld, thereby causing a vertical component to the material flow. In addition, the dragging of retreating side material over to the advancing side at the weld crown by the shoulder rotation sets up a circulation of material about the longitudinal axis of the weld. The rotation about the longitudinal axis is illustrated in Fig. 5, which shows the positions of markers set in the welding path before and after welding. It can be seen in the figure that the material originally on the retreating side
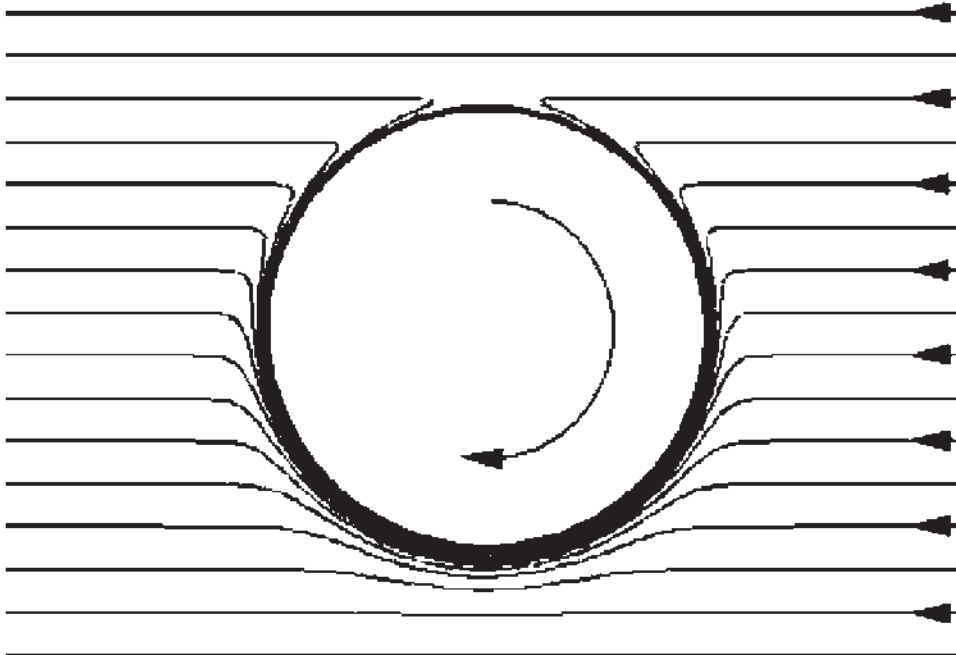
Figure 4 A simplified, two-dimensional, schematic representation of the primary material flow around the pin during FSW. The advancing side of the tool is at the top of central circle and tool rotation is indicated by the arrow.

(a)

(b)
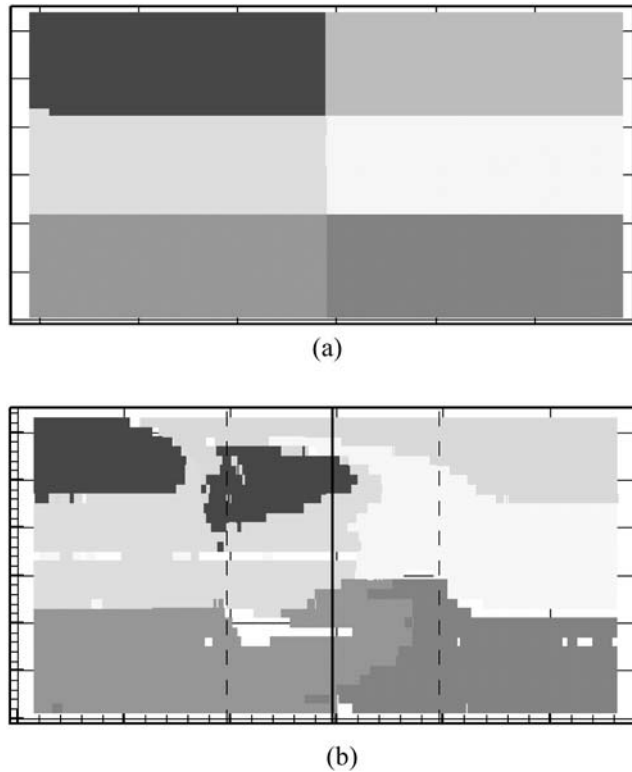
Figure 5 The position of 6 marker inserts (a) prior to welding and (b) after welding is shown. Marker position was determined by serial sectioning and etching.

near the crown of the weld ends up on the crown of the advancing side, while the material originally near the crown on the advancing side is pushed down toward the midplane of the welded plate. The importance of this secondary flow on the overall material transport will diminish as the plate thickness increases. In very thick welds, it is likely that the transport of material is accomplished almost entirely by the pin except for a thin region near the crown.

There are two types of common FSW defects, both resulting from insufficient material transport or deformation. The first is the root side lack of penetration or LOP. This defect normally results from excessive stand-off distance between the backing plate and the bottom of the FSW tool. When the stand-off distance is too large, material at the root side interface is not sufficiently deformed to consume the bond line. An LOP is often identified in a metallographic section when the recrystallized nugget material does not extend to the bottom of the plate. Lack of penetration is not detectable via radiography because it is not a volumetric defect. Typically, the unbonded interface at the root appears to be tightly closed and, in some cases, a weak diffusion bond is present, making dye penetrant inspection ineffective. Phased array ultrasound or high sensitivity eddy current techniques may be capable of finding these defects. Lack of penetration typically results in poor weld transverse tensile strength and fracture through the weld nugget. The

required minimum distance between the bottom of the tool and the root side of the weld depends on the alloy, the tool geometry (e.g. round or flat tipped pin) and the welding parameters.

The other typical FSW defect, the advancing side wormhole or tunnel, is a result of insufficient material transport around the tool pin to the advancing side. The wormhole most commonly occurs when the tool advance per revolution is too high [21]. In severe cases, the defect may be crown side surface breaking, but normally it is not visible except in a metallographic cross-section. Typically, the wormhole is a volumetric defect that can be observed by radiography. In addition, because of the very repeatable nature of FSW, the wormhole or tunnel, if present, is continuous (hence the name). In friction stir welding of aluminum, the proper choice of the welding parameters can usually prevent the formation of an advancing side wormhole.

4.2   Process Simulations

It is quite obvious that FSW is a coupled thermo-mechanical process, i.e. power is dissipated by the intense deformation of the weld metal causing an increase in the weld metal temperature which, in turn, changes the flow characteristics of the weld metal. Friction stir welding process simulations can be divided into two types: those which consider material transport and those which are concerned only with heat transfer. In those models which consider only heat transfer, the tool is normally considered to be a traveling heat source and the temperature histories of various locations in the weld are calculated either by the use of an analytical technique or by the use of the finite element method: models may be either two- or three-dimensional. Variation in the heat transfer models is derived from the differing heat source geometries used and the schemes used for distributing heat flux along the boundaries of the source. Boundary conditions (e.g. transfer between the backing plate and the workpiece) are also of significance. The heat transfer models are certainly valuable when experimental data (temperature measurements, weld power input) are available to verify their accuracy. They can be used to predict changing thermal histories associated with some process variable changes and the thermo-physical properties of the weld metal (e.g. conductivity). However, the heat transfer models do not provide insight regarding the effects of material flow stress and process parameters on important issues such as defect formation, process force requirements, and the mechanisms of heat generation during FSW [22,23].

The simulations that include material flow generally make use of the finite element method. This category of simulation may also be further subdivided by the level of complexity. Friction stir welding simulations of material flow may be either thermo-mechanically coupled or uncoupled. In the coupled models, heat is generated by either viscous dissipation (fluid mechanics-based models) or by work of deformation (solid models). In the coupled models, the temperature field and the associated material properties in the weld metal are allowed to develop naturally in response to the applied deformation. In uncoupled models, a measured temperature field is imposed on the computational domain and the material flow is allowed to occur in response to the imposed tool/workpiece boundary conditions. The fully coupled models are far more flexible in their ability to simulate the process; however, the uncoupled models require fewer computational resources.

Of critical importance for high fidelity process simulation and for accurate prediction of trends resulting from varying process parameters, is the selection of an appropriate constitutive law for the weld material. For best results, the time and temperature-dependent material flow properties should be accurately known to calculate the highest temperature encountered during actual FSW. Constitutive law development is a critical area for advancement of FSW process modeling; particularly, for temperatures approaching the alloy solidus.

Although flow simulation models developed to date may differ in detail, the common viewpoint is that nearly all of the material transport around the tool takes place on the retreating side as described in a preceding section (see Fig. 4). It is also well established that the material that experiences the highest strain and strain rate is on the advancing side of the weld. Actual predictions of maximum strain and strain rate values are not common in the literature; however, strains of the order of 10^2 and strain rates of the order of 10^3–10^4 s^{-1} have been predicted by both fluid- and solid-based models. In general, the process simulations are in good agreement with the flow visualization experiments. Another important point brought out by process simulation is that the strain, strain rate, and temperature gradients within the FSW process zone are very steep; i.e. *most* of the deformation associated with FSW occurs within a millimeter of the tool surface.

4.3   Metallurgical Transformations

From a practical standpoint, the metallurgical transformations that take place in the vicinity of an aluminum friction stir weld are of utmost importance, as these will govern the properties of the weld. Detailed transmission electron microscope studies of FSW microstructures are somewhat limited [24–28]. Much of the information regarding the metallurgy of the welds has been deduced from light optical investigations and, indirectly, from hardness mapping of the welds; however, general trends are fairly well established. The transformations occurring in nonheat-treatable alloys are more simple than those that take place in precipitation hardening systems and so will be discussed first.

The 5XXX series aluminum–magnesium alloys are good examples of friction stir weldable, nonheat-treatable material. These alloys may be obtained either in a fully annealed (O-temper) or in strain hardened tempers of varying hardness (HXXX). When the O-temper material is friction stir welded, the hardness of the weld region may be either identical to that in the base metal or may exhibit a very slight overmatching [29]. As the alloy is single phase, it is likely that the over-matching is a result of the high degree of grain refinement that occurs in the weld nugget or, possibly, due to some retained cold work. When 5XXX alloys are welded in a strain hardened condition (e.g. H113), the weld region is substantially undermatched relative to the base metal (the weld hardness is low compared to the base metal). The friction stir welding process acts as a local annealing treatment, greatly reducing the dislocation density and the strength in the weld region. Typically, in strain hardened single phase alloys such as 5454-HXX, the HAZ consists of both a recrystallized region (closest to the weld nugget) and a recovered region. Figure 6 illustrates schematically the typical hardness profiles resulting from FSW of annealed and of strain hardened 5XXX series alloys as well as precipitation hardening alloys (discussed in the following). Typically, the region
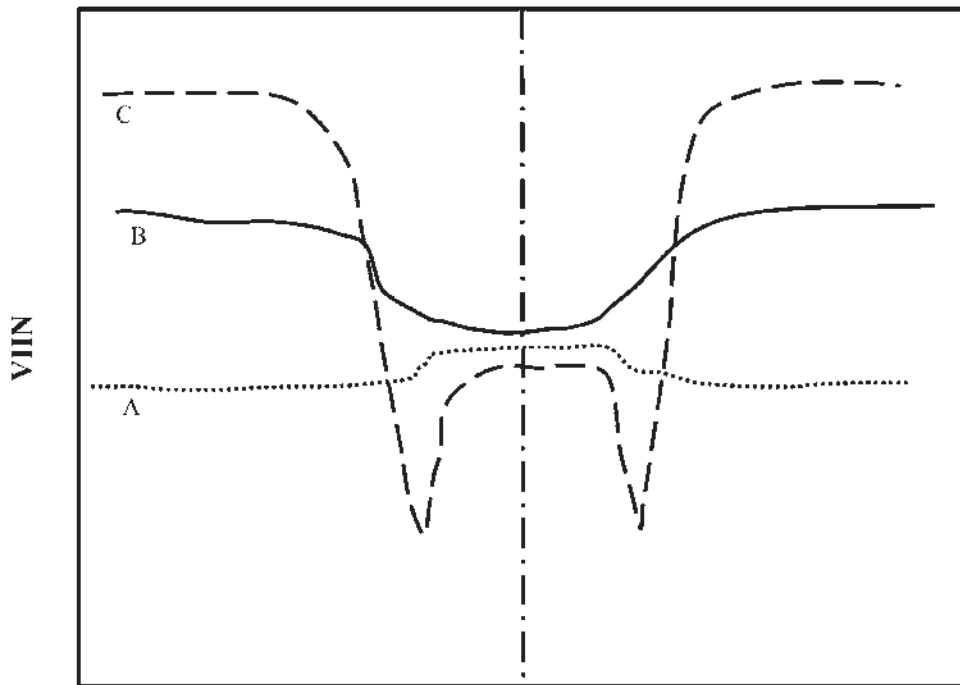
Position Relative to the Weld Centerline

Figure 6 Schematic representation of typical FSW hardness profiles corresponding to various alloy types. "A" is typical of pure materials or solution strengthened alloys in an "O" temper: note slight overmatching in the nugget region. "B" is typical of strain hardened materials. "C" illustrates the typical behavior of precipitation hardened alloys: note the characteristic W-shaped curve with hardness minima just outside the nugget region.

of softened material adjacent to the weld will be narrower as the welding speed is increased.

When precipitation hardening aluminum alloys are friction stir welded, the metallurgical transformations and the distribution of properties in the weld and adjacent HAZ are more complex. In the nugget region of the weld, strengthening precipitates are normally not present. This has been observed in 7XXX, 2XXX, and 6XXX alloys [30–32]. The absence of the precipitates is indicative of a solution heat treated condition in the weld nugget. Although second phase strengthening precipitates are not observed in the nugget, the dislocation density may be quite variable with adjacent grains having widely differing dislocation densities.

Outside of the nugget, in the HAZ close to the nugget, severe precipitate coarsening is observed: the number density of precipitates is greatly reduced and the average diameter increases. At greater distance from the nugget, the coarsening is less severe and after some distance, the base metal microstructure is again observed. The observed progression from a solution heat treated weld center, through varying

severity of overaged structures to the base metal structure leads to a "W"-shaped distribution of hardness that is characteristic of friction stir welds in the precipitation of hardening aluminum alloys. This hardness distribution is illustrated schematically in Fig. 6. The minimum hardness region normally corresponds to the severely overaged material outside of the weld nugget. The nugget itself is at a hardness level intermediate between the overaged material and the base metal. Effects of post-weld aging have been examined in 6063 by Sato et al. [31] and in 7050 by Jata et al. [30]. In both the cases it was observed that the nugget hardness could approach the base metal value with appropriate post-weld aging treatment. The hardness in the severely overaged region could not be returned to the base metal level. In the 7050 alloy, essentially no change in the hardness of the minimum hardness region was observed due to post-weld aging, while in the 6063, *near* base metal hardness was observed throughout the weld region after aging. Work by Reynolds, Lindner, and Tang has shown that the difference between the hardness of the minimum hardness region and the weld nugget in 7050 can be manipulated by choice of welding parameters. The depth of the hardness "well" on either side of the nugget is increased by increasing the welding speed. Higher welding speeds tend to increase the hardness of the weld nugget while maintaining a relatively low hardness in the overaged region. The hardness difference between the minimum and the nugget in friction stir welds in 6.4 mm thick 7050 plate is shown as a function of welding speed in Fig. 7 [33].
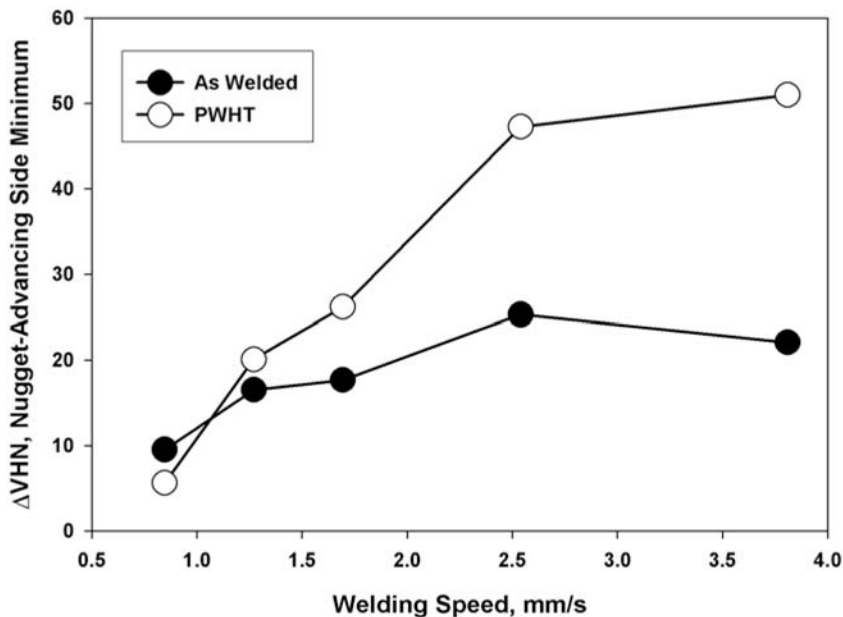
Figure 7 The effects of welding speed and post-weld heat treatment (PWHT) on the difference between nugget and minimum hardness (DVHN) in friction stir welded, 6.4 mm thick 7050-T7.

5 PROPERTIES OF FRICTION STIR WELDS

5.1 Mechanical Properties of Friction Stir Welds

Because the FSW process has only recently become a subject of wide study, there are currently no large databases of weld properties and, in fact, no specifications on how to make or test friction stir welds currently exist. In general, the process is robust and a wide range of processing parameters and tool designs can be used to make metallurgically sound welds in a given alloy and plate thickness. While welds free of LOP or wormhole defects may be made using a wide range of processing parameters, the chosen process parameters may significantly affect the mechanical properties of the weld either through direct modification of the weld microstructure or by indirect influence (e.g. by modification of the residual stress state).

5.1.1 Tensile Properties

The most common measure of FSW quality after visual inspection for surface breaking defects may be the transverse tensile test (loading direction perpendicular to the welding direction). Friction stir welds are heterogeneous materials; therefore, offset yield strengths and percent elongation to fracture are not very meaningful measures of weld properties. Because most FSW in aluminum alloys are undermatched (the weld is softer than the base metal), deformation is nonuniform as soon as the weakest part of the weld begins to yield. As such, a 0.2% offset yield strength will be calculated for a gage section in which the strain is always nonuniform and will be highly dependent on the gage length chosen. The same applies to the calculation of elongation or strain to fracture. However, techniques have been developed for measuring local strains and even for local yield strengths of nonhomogeneous material like welds and it has been observed that the weld nuggets and HAZs of friction stir welds are generally quite ductile [34]. Figure 8 is an example of a strain map of a 2024 friction stir weld loaded in transverse tension. The strain was determined by digital image correlation. The map shows that in this precipitation hardened alloy, severe strain localizations occur in the HAZs on both
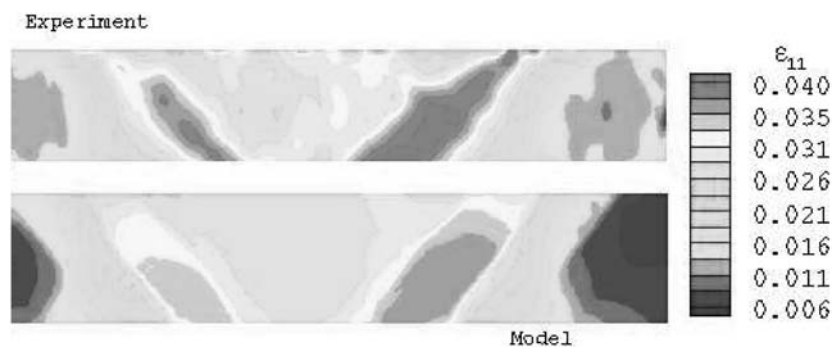
Figure 8 Contour maps showing the distribution of longitudinal strain in a transverse tensile test of a 2024-T3 friction stir weld. The map on top was determined experimentally using digital image correlation, the lower map is a FEM simulation of the test. Note the localization of strain in the low hardness HAZ regions.

advancing and retreating sides of the weld. While strain is localized in these regions, the local ductility may be quite high ($>20\%$). Due to strain localization in the HAZs, most FSW in precipitation hardened alloys will fail in the HAZ when loaded in transverse tension as long as no weld defects are present [2].

The least ambiguous transverse tensile property of a friction stir weld is the tensile strength. This is usually reported both as a raw value and as a percentage of the base metal tensile strength (the joint efficiency). The achievable joint efficiency depends strongly on the alloy. Alloys which rapidly and strongly natural age or which depend on a solid solution for strengthening tend to exhibit better properties than those which do not. For example, in a nonheat-treatable material like 5083-O, joint efficiencies approaching 100% are readily attainable. In 6061-T6, a joint efficiency of 60% is typical. Alloy 2024-T3, which exhibits good natural aging due to the combination of copper and magnesium, may exhibit joint efficiency of greater than 85% in relatively thick plate (8 mm). Alloy 2219, on the other hand, which does not contain magnesium to enhance natural aging response, typically exhibits joint efficiencies between 65% and 70%. Most of the 7XXX alloys (which naturally age) also exhibit high joint efficiency, sometimes exceeding 90% [2]. In general, the tensile strength of friction stir welds in age-hardening aluminum alloys improves with increasing welding speed.

The joint efficiency of friction stir welds is typically better than that of fusion welds in readily weldable alloys and far better than fusion welds when the more difficult to weld 2XXX and 7XXX alloys are considered. For example, Colligan et al. [21] presented data for 18 mm thick 2195-T8: FSW joint efficiency was 66% while that for variable polarity plasma arc (VPPA) welds was only 47%. For 25 mm thick 5083-H1131, an FSW joint efficiency of 89% was obtained while gas metal arc welds exhibited joint efficiency of 75%.

5.1.2 Fracture Toughness of Friction Stir Welds

Reports on the fracture toughness of friction stir welds are almost nonexistent in the open literature. Those reports that do exist are mainly confined to relatively thin plate and R-curve type experiments. Valid K_{IC} measurements in FSW have not been reported. The R-curve behavior of friction stir welds has generally been found to be superior to that of the parent metal, particularly for alloys that are welded in a peak-aged or near peak-aged condition. For example, Dawes et al. [35] tested friction stir welded plates of 2014-T651, 7075-RRA (retrogression reaged), and 5083-O. In the cases of the 2014 and 7075, the welds were substantially undermatched and exhibited significantly better fracture properties than the base metal (higher initiation toughness, $J_{0.2}$, and more steeply rising R-curves). For the 5083, the level of undermatching was small and the differences between the fracture behavior of the welds and the parent metal were small as well. On the other hand, Sutton et al. [36] have examined the fracture resistance of FSW in 2024-T351 plate, a high toughness temper in a high toughness alloy, and found that the fracture resistance is degraded relative to the base metal: a similar effect in 5 mm thick 2024-T3 plate was noted by von Strombeck et al. [37].

From a practical standpoint, it may be more useful to compare the fracture resistance of a friction stir weld to that of a fusion weld in the same material. Kroninger and Reynolds [3] have performed this comparison for a friction stir weld

and a VPPA weld in 8.1 mm thick plates of 2195-T8. R-curves and initiation toughness were measured for cracks parallel to the welding direction. Cracks were grown in the nugget and both HAZs of the FSW (advancing and retreating sides), through the center of VPPA weld, and in the parent metal. The toughness of the VPPA weld was lower than that of the parent metal both for crack initiation and growth. On the other hand, the toughness of the FSW nugget and HAZs were higher than the parent metal toughness. This is not solely an effect of reduced hardness because the VPPA weld was significantly softer than the parent metal as well. The primary reason that the FSW performs better than the VPPA weld is likely due to the absence of brittle solidification products in the FSW.

The overall effect of assembling structures by friction stir welding on fracture resistance, accounting for any reductions in strength and changes in toughness, will only be determined by testing of large-scale test articles.

5.1.3 Fatigue

Along with fracture properties, the fatigue behavior of friction stir welds will be of critical importance if FSW is to be used for structures that are designed on the basis of damage tolerance (e.g. airframes). Most of the available FSW fatigue data are in the form of S–N curves; however, some da/dN–ΔK data have become available recently.

Several salient features of the S–N performance of friction stir welds were illustrated in a study by Bussu and Irving [38]. In this study, 2024-T3 butt welds were tested in transverse and longitudinal orientations in tension–tension fatigue ($R = 0.1$). In the as-welded conditions, both longitudinal and transverse specimens exhibited substantially reduced fatigue lives relative to the parent metal specimens: a 50% reduction in stress range at a life of 10^7. However, when 0.5 mm was skimmed from the root and crown surfaces of the welds, the performance was nearly identical to the parent metal. In the welds that did not have surface irregularities machined off, the small lip of material that is normally produced on the advancing side of the weld crown acted as the fatigue crack initiation site in every case for the transversely loaded specimens (see Fig. 2). In the unskimmed, longitudinally loaded specimens, the failure initiation site was always at one of the semicircular tool marks on the crown side of the weld. When the lip and the tool marks were removed, fatigue failure in the transversely loaded specimens initiated in the minimum hardness regions of the HAZ and failure initiation in the longitudinal specimens was distributed randomly, as in the parent metal specimens.

At the same conference, Biallis et al. [39] demonstrated a very consistent dependence of the fatigue life on the welding speed for 2024-T3. It was shown that higher welding speeds improved the fatigue life for both 1.6 and 4 mm thick butt welds. The higher welding speeds also correlated with higher hardness in both the minimum hardness regions of the HAZ and in the nugget. It was surmised that the improved fatigue performance was due to the higher hardness in these regions.

At the third ISFSW, Kumagai et al. [40] compared parent metal, FSW, and MIG-welded 6N01-T5 (a heat-treatable Al–Mg–Si alloy) fatigue properties. The performance of the FSW was not as good as the parent metal, but substantially better than the MIG welds. At the second ISFSW, Magnusson and Kallman [41] showed results for 2024-T3, 7475-T76, and 6013-T6. Each of these alloys exhibited

the same effect of surface skimming that was demonstrated previously for 2024-T3 by Bussu and Irving [38]. In addition, the 6013 data indicate a rapid reduction in life relative to the parent material as the maximum stress level rises above 250 MPa. This was attributed to the reduced ultimate static strength of the weld relative to the parent material (291 MPa vs. 388 MPa). The welds in the 2024-T3 and 7475-T76 presumably did not exhibit this behavior because they had substantially greater joint efficiency than the 6013 (>90% vs. 75%).

In general, the stress-life fatigue behavior of friction stir welded joints appears to be governed either by geometric stress concentrations (all stress levels) or from the presence of regions of low hardness in the heat affected zones of the friction stir welds (high stress levels). Data indicate that effects of the geometric stress concentrations arising from FSW may be eliminated by post-weld machining. The hardness minima in the heat affected zones of friction stir welds may be unavoidable; however, their severity may be reduced by modifying the process parameters to enable higher welding speeds and hence, shorter times at temperatures producing overaged structures. The factors affecting stress-life fatigue performance are summarized in the schematic stress-life curves shown in Fig. 9. An effect that has not been carefully investigated to date is that of residual stress. It is certain that the residual stress state existing in a welded structure is not the same as that which exists in a relatively small fatigue coupon. The fatigue life of a *structure* may be strongly influenced by residual stress.
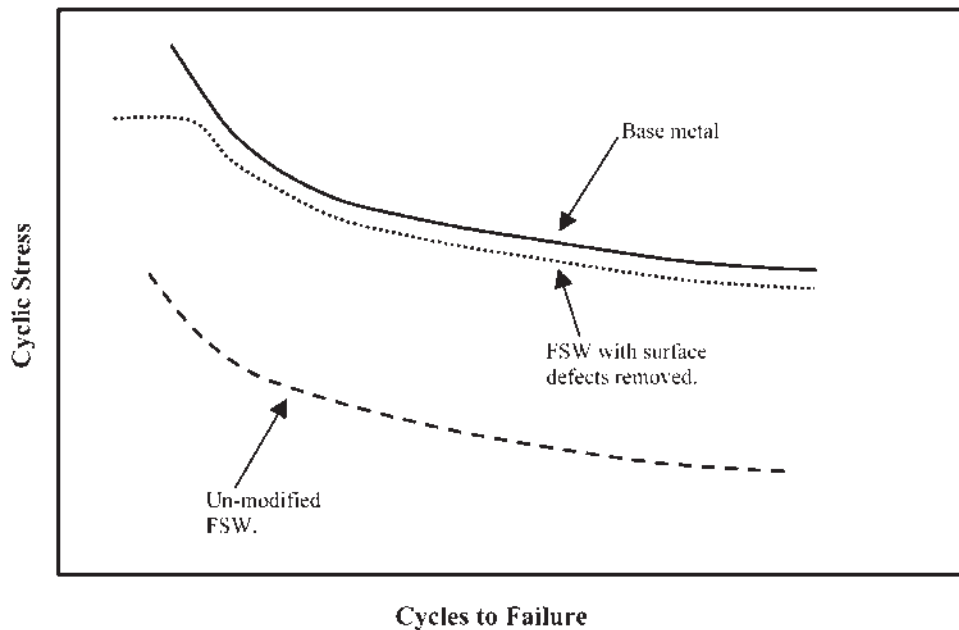

Figure 9   Schematic representation of the effects of FSW on high cycle fatigue life. Removal of surface defects (see Fig. 2) restores fatigue life to near base metal values. However, as the maximum stress approaches the strength of the minimum hardness regions of the weld, fatigue life may be degraded relative to the base metal.

The primary and unanimous conclusion of groups that have been performing research on fatigue crack growth in friction stir welds is that residual stress can exert a dominant effect on growth rates through the welds [30,42,43]. Dalle Donne et al. [43] observed that da/dN for cracks growing along the centerline of friction stir welds in C(T) specimens was substantially reduced relative to the base metal (2024-T3) when testing was performed at low load ratio, $R = 0.1$. This effect was especially pronounced at low ΔK. When crack growth testing was performed at $R = 0.7$, the performance of the base metal and the friction stir weld were indistinguishable. The effect of residual stress on the effective ΔK was measured as a function of crack length using the so-called "cut-compliance technique". The results of the cut-compliance testing were used to rationalize the observed fatigue crack growth behavior. The Dalle Donne group also performed some testing using middle crack tension (M(T)) specimens. The M(T) specimens revealed quite a different behavior from the C(T) specimens: at $R = 0.1$, the crack growth rates measured using the M(T) specimen configuration were greater than the base metal values. This finding is in accord with the results of Ohta et al. [44] who determined that for arc welds, residual stresses at the tips of M(T) specimen cracks were tensile, while C(T) specimens had compressive stresses.

5.2 Residual Stress

It has been stated that residual stress and distortion in friction stir welds are substantially less than that observed in equivalent fusion welds; however, very few one-to-one comparisons have been made to back up this claim. As described in the preceding section, it is believed that residual stresses have a large influence on the fatigue crack growth rate observed as cracks grow within and through friction stir welds. In addition to the cut-compliance measurements described in the section of this chapter on fatigue crack growth, several studies of residual stress in FSW have been made using x-ray and neutron diffraction.

One of the first x-ray diffraction studies of residual stress in FSW was performed by James and Mahoney [45] and presented at the first International Symposium on FSW at Thousand Oaks, California in 1999. The stress levels observed by this group in FSW of alloy 2219, 7050, and C458 were typically of the order of 50 MPa (7.2 ksi). However, these results were obtained by standard x-ray techniques applied to material near the surfaces of the weld. Later, studies have indicated higher residual stresses. For example, Sutton et al. [46] used neutron diffraction to map the residual stress state in a 2024-T3 FSW (7 mm thick). In the weld nugget, tensile residual stresses in the longitudinal direction were observed at levels of between 75 and 105 MPa. Transverse stresses were smaller and generally compressive in the weld zone. Dalle Donne et al. [47] observed longitudinal residual stresses up to 140 MPa (47% of parent metal yield) in 7108-T79. Although the residual stresses associated with FSW may be less than those resulting from fusion welding, the fatigue crack growth studies cited in the preceding section indicate that the residual stresses are by no means insignificant. Additional studies are warranted to quantify this issue and examine the effects of varying process parameters and tooling design on residual stress. It is interesting to note that longitudinal, tensile, residual stress levels equal to base metal yield stress have been observed in the nuggets of austenitic stainless steel friction stir welds [48].

5.3   Corrosion

As for fatigue performance, at present, only a limited amount of data regarding the corrosion behavior of friction stir welds has been generated and published; however, a few trends may be discerned. It has generally been observed that, in the highly corrosion resistant 5XXX series alloys, the corrosion resistance of friction stir welds is equivalent to the base metal. In some very limited studies of more complex precipitation hardening alloys, less desirable corrosion properties have been observed in some cases. For example, Biallas et al. [39] observed degraded performance in the EXCO test for the nugget region of a 2024-T3 FSW. The nugget also showed a propensity for IG attack. However, no increased susceptibility to stress corrosion cracking (relative to the base metal) was observed after 40 days of alternate immersion testing. In another study by Li et al. [49], the stress corrosion cracking resistance of 2219-T87 and 2195-T8 friction stir welds was found to be superior to both fusion welds in the same materials and the respective base metals.

While the limited data available make it impossible to draw any broad conclusions regarding the SCC, general corrosion, and pitting corrosion performance of friction stir welds, common sense indicates that the following is likely. In the weld nugget region, the performance will be similar to that of a solution treated and annealed (nonheat-treatable alloys) or naturally aged (heat-treatable alloys) material. In the heat affected zone, one should expect a behavior similar to that observed in fusion weld HAZs or in other highly overaged structures. It will be especially critical to examine the SCC behavior of FSW structure in high strength alloys that rely on complex tempers and favorable grain structures to prevent SCC (e.g. 7050-T7XX) as both the temper and the grain structure are destroyed by FSW and easy paths for SCC propagation may be produced by the welding operation.

6   EMERGING PROCESS ADVANCES

Friction stir welding process modifications have been developed over the past few years in an effort to expand the number of applications for friction stir welding. The bulk of the modifications have been designed to either increase welding speed, to enable welding of thicker sections, to enable friction stir welding of structures having complex curvature and/or variable thickness, or to reduce the necessity for support tooling to react the loads required by the process.

Modifications to increase welding speed and enable welding of thicker sections have both gone along similar paths. The key element in both the cases is development of pin designs that facilitate the transfer of material from the front of the tool to the back of the tool in the most efficient manner while still promoting full consolidation of the weld material. The specifics of tool design are often of a highly proprietary nature but, in general, increased speed and weld depth necessitate the addition of reentrant features (large pitch and depth thread forms and flutes) to the pins. In addition, rather than using right circular cylindrical pins, the pins may be tapered from top to bottom thereby reducing the bending moment on the pin and enabling higher welding speeds. Pin failure is normally observed at the interface between the pin and the shoulder which is the site of maximum pin bending moment in tools having cylindrical pins. As weld depth increases, the torque capability of the

pins must also be considered as the required torque will be a function of pin length and diameter [50].

While variations in pin design provide incremental improvement to the FSW process, a more fundamental change has been brought about by the development of the so-called "scroll shoulder" tools which enable welding with the tool axis perpendicular to the plates to be joined (normal to the surface) [50]. The ability to make welds with the tool normal to the surface rather than requiring a lead angle greatly eases the task of making other than straight welds. If a constant nonzero lead angle must be maintained during the welding process, then either the work must be made to approach the tool from a constant direction (e.g. as in sewing a curved seam on a sewing machine) or the spindle carrying the tool must be capable of motion in multiple axes requiring complex programming of the tool motion. In addition, multi-axis heads are generally less robust than those which move only in the z and x directions, making the reaction of large welding forces problematic. The shoulder design required for normal to the surface welding is relatively simple being a spiral groove machined into a planar surface rather than the concave dish typically used for conventional FSW tools. Property data for welds made with normal to the surface tools are not widely available but early trials indicate that weld quality should not be compromised by this process modification.

Production of a good quality, full penetration weld requires that the bottom of the pin remains in close proximity to the bottom of the plates being welded and that the shoulder maintain contact with the top surface of the plates. These requirements dictate the pin length for a given plate thickness. If a variable depth weld is required, as for transitions between structural sections of different thickness, then a tool with a variable pin length is required. Such a requirement led to the development of the retractable pin tool. The retractable pin tool is capable, as the name suggests, of changing pin length in midweld and with proper programming can be used to make full penetration welds on variable thickness sections [51].

An offshoot of the retractable pin tool is the self-reacting tool [52]. Self-reacting tool technology should open up a great number of new applications for FSW. The principle behind the self-reacting tool is very simple in concept: the pin rotates with, but is not connected to the shoulder, just as in the retractable pin tool. However, instead of terminating the pin near the bottom of the plate, the pin extends completely through the plate and is attached to a second shoulder which contacts the "root" side of the weld. It is this second shoulder that provides the self-reacting element. During a self-reacting weld, the forces exerted on the crown and root (or second crown) sides of the weld are balanced by applying a tensile load through the pin to the second shoulder thereby reacting the compressive load applied by the top shoulder. Hence, the need for substantial and expensive support tooling may be greatly reduced relative to single-sided FSW. Obviously, self-reacting tool welds place additional requirements on the material of the pin: during a self-reacting weld, torsional, bending, and tensile loads will be applied to the pin. In addition to reducing support tooling requirements, the self-reacting tool may enable higher welding speeds and, in some cases, improved weld properties. Also, the heating provided by the second shoulder creates a more symmetric weld than does single-sided welding; this can have significant effects on the fracture path in transverse tension and, may modify the residual stress state as well. A cross-section of a self-reacting weld is shown in Fig. 10.
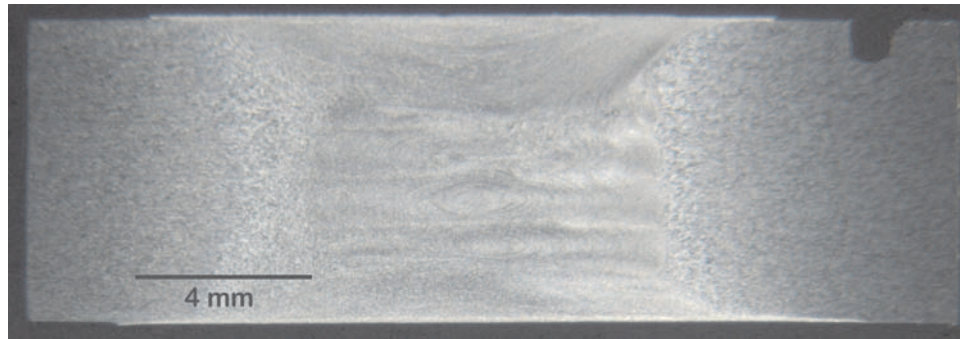
Figure 10 Transverse metallographic cross-section of a 2219 FSW made with a self-reacting tool. Note the symmetry between root and crown. The plate thickness is 8 mm. (Photo courtesy of MTS Systems Corporation, Advanced Engineering Solutions Division.)

7 SUMMARY

Since its invention in 1991, the friction stir welding process has experienced continuous development and rapid insertion into production applications. The solid-state joining process offers substantial benefits relative to fusion welding, particularly for high strength, aerospace, aluminum alloys. Friction stir welding is readily substituted for most fusion welding processes and with continued development of modifications such as the self-reacting tool, it may be capable of economically replacing mechanical joining in many structural applications. Although many FSW properties are demonstrably superior to those obtained using fusion welding techniques, the process does produce residual stresses and modified microstructures relative to the parent metal; hence, corrosion and stress corrosion cracking issues must be investigated. It seems likely that as understanding of the process increases, the friction stir welding process will continue to find additional applications in all areas of aluminum joining.

REFERENCES

1. W. M. Thomas, E. D. Nicholas, J. C. Needham, M. G. Murch, P. Templesmith, and C. J. Dawes, "Friction Stir Butt Welding," G. B. Patent Application No. 9125978.8, December 1991; U.S. Patent No. 5460317, October 1995.
2. A. P. Reynolds, W. D. Lockwood, and T. U. Seidel, "Processing Property Correlation in Friction Stir Welds," in: *Aluminum Alloys: Their Physical and Mechanical Properties*, Materials Science Forum, 2000, Vol. 331-3, pp. 1719–1724.
3. H. R. Kroninger and A. P. Reynolds, "*R*-Curve Behavior of Friction Stir Welds in Aluminum–Lithium Alloy 2195," Fatigue and Fracture of Engineering Materials and Structures 2002, 25(3), pp. 283–290.
4. S. Baxter and A. P. Reynolds, "Characterization of the Reinforcing Particle Size Distribution in a Friction Stir Welded Al–SiC Extrusion," in: Proceedings of the Symposium on Lightweight Alloys for Aerospace Applications, TMS 2001 Annual Meeting, February 11–15, 2001, New Orleans, LA (available on-line).
5. M. R. Johnsen, "Friction Stir Welding Takes Off at Boeing," The Welding Journal, 1999, 78(February; 2), pp. 35–39.

6. O. T. Midling, J. S. Kvale, and O. Dahl, "Industrialization of the Friction Stir Welding Technology in Panels Production for the Maritime Sector," in: Proceedings of the First International Symposium on Friction Stir Welding, Thousand Oaks, CA, June 1999, distributed on CD (available from TWI).

7. C. Jones and G. Adams, "Assembly of a Full-Scale External Tank Barrel Section Using Friction Stir Welding," in: Proceedings of the First International Symposium on Friction Stir Welding, Thousand Oaks, CA, June 1999, distributed on CD (available from TWI).

8. Anonymous, "New Jet to be Built with Friction Stir Welding," The Welding Journal, 2000, 79(November; 11), p. 7.

9. A. P. Reynolds and Wei Tang, "Alloy, Tool Geometry, and Process Parameter Effects on Friction Stir Weld Energies and Resultant Joint Properties," in: *Friction Stir Welding and Processing* (K. V. Jata, M. W. Mahoney, R. S. Mishra, S. L. Semiatin, and D. P. Field, eds.), TMS, 2001, pp. 15–23.

10. Private Communication with Dr. T. J. Lienert, University of South Carolina.

11. K. Colligan, "Material Flow Behavior During Friction Stir Welding of Aluminum," The Welding Journal, 1999, 78(7), pp. 229s–237s.

12. T. U. Seidel and A. P. Reynolds, "Visualization of Material Flow in AA2195 Friction Stir Welds Using a Marker Insert Technique," Metallurgical and Materials Transactions A, 2001, 32A(November), pp. 2879–2884.

13. Y. Li, L. E. Murr, and J. C. McClure, "Solid State Flow Visualization in the Friction Stir Welding of 2024 Al to 6061 Al," Scripta Materialia, 1999, 40(9), pp. 1041–1046.

14. O. T. Midling, "Material Flow Behaviour and Microstructural Integrity of Friction Stir Butt Weldments," in: Proceedings of the Fourth International Conference on Aluminum Alloys (T. H. Sanders and E. A. Starke, Jr., eds.), The Georgia Institute of Technology, Atlanta, GA, U.S.A., 1994, Vol. I, pp. 451–458.

15. A. Askari, S. Silling, B. London, and M. W. Mahoney, "Modeling and Analysis of Friction Stir Welding Processes," in: *Friction Stir Welding and Processing* (K. V. Jata, M. W. Mahoney, R. S. Mishra, S. L. Semiatin, and D. P. Field, eds.), TMS, 2001, pp. 43–54.

16. S. Xu, X. Deng, A. P. Reynolds, T. U. Seidel, "Finite Element Simulation of Material Flow in Friction Stir Welding," Science and Technology of Welding and Joining, 2001, 6(3), pp. 191–193.

17. A. P. Reynolds, X. Deng, T. U. Seidel, and S. Xu, "Recent Advances in FSW Process Physics," in: Proceedings of the Conference on Joining of Advanced and Specialty Materials, St. Louis, Missouri, U.S.A., October 2000, pp. 172–177.

18. G. J. Bendzsak, T. H. North, and C. B. Smith, "An Experimentally Validated 3D Model for Friction Stir Welding," in: Proceedings of the Second International Symposium on FSW, Gothenburg, Sweden, June 2000 (distributed by TWI on CD).

19. P. Colegrove, M. Painter, D. Graham, and T. Miller, "3-Dimensional Flow and Thermal Modelling of the Friction Stir Welding Process," in: Proceedings of the Second International Symposium on FSW, Gothenburg, Sweden, June 2000 (distributed by TWI on CD).

20. A. C. Nunes, Jr., E. L. Bersnstein, and J. C. McClure, "A Rotating Plug Model for Friction Stir Welding," Presented at the 81st American Welding Society Annual Convention, Chicago, IL, U.S.A., April 26–28, 2000.

21. K. Colligan, I. Ucok, K. McTernan, P. Konkol, and J. R. Pickens, "Friction Stir Welding of Thick Section 5083-H131 and 2195-T8P4 Aluminum Plates," in: Proceedings of the Third International Symposium on Friction Stir Welding, Kobe Exhibition Center, Port Island, Kobe, Japan, September 2001 (distributed by TWI on CD).

22. J. E. Gould and Z. Feng, "Heat Flow Model of Friction Stir Welding of Aluminum Alloys," Journal of Materials Processing and Manufacturing Science, 1998, 7(October; 2), pp. 185–194.

23. J. C. McClure, Z. Feng, T. Tang, J. E. Gould, and L. E. Murr, "A Thermal Model of Friction Stir Welding," in: Proceedings of the Fifth International Conference on Trends in Welding Research (J. M. Vitek, S. A. David, J. A. Johnson, H. B. Smartt, and T. DebRoy, eds.), Pine Mountain, GA, U.S.A., June 1998, pp. 590–595.

24. K. V. Jata and S. L. Semiatin, "Continuous Dynamic Recrystallization During Friction Stir Welding of High Strength Aluminum Alloys," Scripta Materialia, 2000, 43, pp. 743–749.

25. M. W. Mahoney, C. G. Rhodes, R. A. Spurling, and W. H. Bingel, "Properties of Friction Stir Welded 7075-T651 Aluminum," Metallurgical and Materials Transactions A, 1998, 29A(July), pp. 1955–1964.

26. Y. S. Sato, H. Kokawa, M. Enomoto, and S. Jogan, "Microstructural Evolution of 6063 Aluminum during Friction Stir Welding," Metallurgical and Materials Transactions A, 1999, 30A(9), pp. 2429–2437.

27. O. V. Flores, C. Kennedy, L. E. Murr, D. Brown, S. Pappu, B. M. Nowak, and J. C. McClure, "Microstructural Issues in a Friction Stir Welded Aluminum Alloy," Scripta Materialia, 1998, 38(5), pp. 703–708.

28. H. Salem, A. P. Reynolds, and J. S. Lyons, "Microstructure and Retention of Super-Plasticity of Friction Stir Welded Super-plastic 2095 Sheet," Scripta Materialia, 45(5), pp. 337–342.

29. A. P. Reynolds and F. Duvall, "Digital Image Correlation for Determination of Weld and Base Metal Constitutive Behavior," The Welding Journal Research Supplement, 1999, 78(October; 10) pp. 355s–360s.

30. K. V. Jata, K. K. Sankaran, and J. J. Ruschau, "Friction-Stir Welding Effects on Microstructure and Fatigue of Aluminum Alloy 7050-T7451," Metallurgical and Materials Transactions A, 2000, 31A(September; 9), pp. 2181–2192.

31. Y. Sato, H. Kokawa, M. Enomoto, S. Jogan, and T. Hashimoto, "Precipitation Sequence in Friction Stir Weld of 6063 Aluminum during Aging," Metallurgical and Materials Transactions A, 1999, 30A(December), pp. 3125–3130.

32. G. Oertlet, S. S. Babu, S. A. David, and E. A. Kenik, "Effect of Thermal Cycling on Friction Stir Welds of 2195 Aluminum Alloy," The Welding Journal, 2001, 80(March; 3), pp. 71s–79s.

33. A. P. Reynolds, K. Lindner, and W. Tang, University of South Carolina, Unpublished Research.

34. W. D. Lockwood, Borislav Tomas, and A. P. Reynolds, "Mechanical Response of Friction Stir Welded AA 2024: Experiment and Modeling," Materials Science and Engineering A, 2002, 323(January; 1/2), pp. 349–354.

35. M. G. Dawes, S. A. Karger, T. L. Dickerson, and J. Pryzdatek, "Strength and Fracture Toughness of Friction Stir Welds in Aluminium Alloys," in: Proceedings of the Second International Symposium on FSW, Gothenburg, Sweden, June 2000 (distributed by TWI on CD).

36. M. A. Sutton, A. P. Reynolds, B. Yang, and R. Taylor, "Mode I Fracture and Microstructure for 2024-T3 Friction Stir Welds," Materials Science and Engineering, A (in press).

37. A. von Strombeck, J. F. Dos Santos, F. Torster, P. Laureano, and M. Kocak, "Fracture Toughness of FSW Joints on Aluminium Alloys," in: Proceedings of the First International Symposium on Friction Stir Welding, Thousand Oaks, CA, June 1999, distributed on CD (available from TWI).

38. G. Bussu and P. E. Irving, "Static and Fatigue Performance of Friction Stir Welded 2024-T351 Aluminium Joints," in: Proceedings of the First International Symposium on Friction Stir Welding, Thousand Oaks, CA, June 1999, distributed on CD (available from TWI).

39. G. Biallas, R. Braun, C. Dalle Donne, G. Staniek, and W. A. Kaysser, "Mechanical Properties and Corrosion Behavior of 2024-T3," in: Proceedings of the First International Symposium on Friction Stir Welding, Thousand Oaks, CA, June 1999, distributed on CD (available from TWI).

40. M. Kumagai, S. Tanaka, H. Hatta, H. Yoshida, and H. Sato, "Integral Wing Panel for Airplanes Produced by Friction Stir Welded Extrusion," in: Proceedings of the Third International Symposium on Friction Stir Welding, Kobe Exhibition Center, Port Island, Kobe, Japan, September 2001 (distributed by TWI on CD).

41. L. Magnusson and L. Kallman, "Mechanical Properties of Friction Stir Welds in Thin Sheets of 2024, 6013, and 7475," in: Proceedings of the Second International Symposium on FSW, Gothenburg, Sweden, June 2000 (distributed by TWI on CD).

42. R. John and K. V. Jata, "Residual Stress Effects on Near-Threshold Fatigue Crack Growth in Friction Stir Welds," in: *Friction Stir Welding and Processing* (K. V. Jata, M. W. Mahoney, R. S. Mishra, S. L. Semiatin, and D. P. Field, eds.), TMS, 2001, pp. 35–42.

43. C. Dalle Donne, G. Biallas, T. Ghidini, and G. Raimbeaux, "Effect of Weld Imperfections and Residual Stress on the Fatigue Crack Propagation in Friction Stir Welded Joints," in: Proceedings of the Second International Symposium on FSW, Gothenburg, Sweden, June 2000 (distributed by TWI on CD).

44. A. Ohta, E. Sasaki, M. Nihei, M. Kosuge, and M. Inagaki, International Journal of Fatigue, 1982, 4, pp. 233–237.

45. M. James and M. W. Mahoney, "Residual Stress Measurements in Friction Stir Welded Aluminum Alloys," in: Proceedings of the First International Symposium on Friction Stir Welding, Thousand Oaks, CA, June 1999, distributed on CD (available from TWI).

46. M. A. Sutton, A. P. Reynolds, D. Wang, and C. Hubbard, "Residual Stress Analysis in 2024-T351 Aluminum Friction Stir Butt Weld by Neutron Diffraction," ASME Journal of Engineering Materials and Technology, 2002, 124, pp. 215–221.

47. C. Dalle Donne, E. Lima, J. Wegener, A. Pyzalla, and T. Buslaps, "Investigations on Residual Stresses in Friction Stir Welds," in: Proceedings of the Third International Symposium on Friction Stir Welding, Kobe Exhibition Center, Port Island, Kobe, Japan, September 2001 (distributed by TWI on CD).

48. A. P. Reynolds, M. Posada, J. DeLoach, M. J. Skinner, J. Halpin, and T. J. Lienert, "FSW of Austenitic Stainless Steels," in: Proceedings of the Third International Symposium on Friction Stir Welding, Kobe Exhibition Center, Port Island, Kobe, Japan, September 2001 (distributed by TWI on CD).

49. Z. X. Li, W. J. Arbegast, P. J. Hartley, and E. I. Meletis, "Microstructure Characterization and Stress Corrosion Evaluation of Friction Stir Welded Al 2195 and Al 2219 Alloys," in: Proceedings of the Fifth International Conf. on Trends in Welding Research, Pine Mountain, GA, U.S.A., June 1998, pp. 568–573.

50. C. J. Dawes and W. M. Thomas, "Development of Improved Tool Designs for Friction Stir Welding of Aluminum," in: Proceedings of the First International Symposium on Friction Stir Welding, Thousand Oaks, CA, June 1999, distributed on CD (available from TWI).

51. R. J. Ding, et al., U.S. Patent No. 5,893,507, Auto-Adjustable Pin Tool for Friction Stir Welding, 1999.

52. C. L. Campbell, M. S. Fullen, and M. J. Skinner, U.S. Patent No. 6,199,745, "Welding Head," March 2001.

16

Aluminum Intermetallics

GEORG FROMMEYER and SVEN KNIPPSCHEER

Max-Planck-Institut für Eisenforschung, GmbH, Düsseldorf, Germany

1   INTRODUCTION

The aluminium-rich intermetallic compounds of the Al_3X-type with transition metals (X = Ti, Zr, Nb, V) of groups IVb and Vb are of interest in the development of novel high temperature and light weight structural materials. The Al_3X compounds of main interest are: Al_3Ti and Al_3Nb with ordered DO_{22} superlattice structure and some modified ones of the composition $(Al, X)_3 Ti$ with $L1_2$ type of crystal structure. The important aspects of these aluminium-rich intermetallics are:

– Aluminium atoms are the main constituents of 75 at.%
– The low density of less than 3.5 Mg m^{-3} (Al_3Ti)
– High elastic moduli and high melting points

The first report on the deformation behaviour of this class of trialuminides appeared in 1974. The first paper on Al_3Nb was published by Shechtman and Jacobson in 1975 [1]. About 13 years later Yamaguchi et al. [2] reported on deformation behaviour of Al_3Ti. These preliminary studies showed that Al_3Ti and Al_3Nb were brittle at room and medium temperature. Almost at the same time detailed studies revealed that specific twinning occurs in the DO_{22} crystal structure, which does not disturb the ordering. However, this deformation mechanism plays an important role and inspired many researchers to improve the low temperature ductility of these compounds. Three main approaches have been attempted. These are:

– Microalloying for refining the microstructure.
– Macroalloying to change the crystal structure into higher symmetry ($L1_2$) and to increase the number of the independent slip systems.
– Rapid solidification and powder metallurgical techniques including thermo-mechanical treatment.

The microalloying approach in combination with rapid solidification has been producing some promising results. A series of ternary and quaternary aluminides based on $(Al, X, Y)_3Ti$ with cubic $L1_2$ structure revealed some ductility in bending

tests. However, the tetragonale DO_{22} and the modified cubic $L1_2$ Al_3Ti and Al_3Nb compounds are in general quite brittle for engineering applications. At higher temperatures above $700°C$ the modified trialuminides show sufficient ductility and strength properties for certain applications in internal combustion engines and in jet engines.

The present chapter decribes the important physical and mechanical properties of trialuminides with DO_{22} structure and their $L1_2$ variations. Particular emphasis is made on deformation modes and dislocation structures, the effect of alloying on the mechanical and oxidation properties, and on the application potential. In addition to Al_3Ti relevant investigations on Al_3Nb, Al_3V, Al_3Zr and on cubic $(Al, X)_3Ti$ variations are reviewed. For specific properties the reader is referred to review articles published by Yamaguchi et al. [3] regarding the deformation mechanism of Al_3Ti, by Yamaguchi and Umakoshi [4] for lattice defects, by Georges et al. [5] for deformation and the structure of $L1_2$ trialuminides, by Mikkola and Nic [6] for alloying and cubic phase stability, by Morris [7, 8] for antiphase boundary (APB) faults, and by Reip and Sauthoff et al. [9, 10] for structure and porperties of Al_3Nb.

2   STRUCTURE AND SELECTED PHYSICAL PROPERTIES

2.1   Crystal Structure

The ordered tetragonal DO_{22} structure is derived from the ordered face centered cubic $L1_2$ structure (Fig. 1a) by introducing an antiphase boundary (APB) with a displacement vector of $\frac{1}{2}\langle 110 \rangle$ on every (001) plane (Fig. 1b). Al_3Ti and Al_3Nb crystallize into this $L1_2$-derivative, long-period superlattice structure. The DO_{23} structure is formed by introducing the APB on every second (001) plane (Fig. 1c). Al_3Zr and Al_3Hf (at low temperatures), which are akin to Al_3Ti, possess this structure. The lattice parameters of Al_3Ti are a $=$ 0.3848 nm and c $=$ 0.8596 nm (c/a $=$ 2.23), see e.g. Villars and Calvert [11]. The density of Al_3Ti calculated from these lattice parameters is 3.36 $Mg\,m^{-3}$. For Al_3Nb the lattice parameters are quite similar to those of Al_3Ti. In c-axis of the unit cell the lattice constant is c $=$ 0.8615 nm and the value of a $=$ 0.3848 nm. Considering the higher atomic weight of Niobium atoms the calculated density of Al_3Nb is about 4,454 $Mg\,m^{-3}$. This value is comparable with the density of pure Titanium.

Table 1 represents some important physical properties of polycrystalline Al_3Ti and Al_3Nb compounds.

2.2   Hardness and Elastic Properties

Vickers' macro hardness values and elastic moduli of Al_3Ti and Al_3Nb are presented in Table 2a. The elastic stiffness constants C_{ij} of Al_3Ti and of the constitutents Ti and Al were determined by measurements of ultrasonic sound wave velocity in crystals of the compounds [12] and presented in Table 2b. Values of Young's modulus (E), shear modulus (G), bulk modulus (K) and Poisson's ratio (v) for the polycrystalline compound Al_3Ti were calculated from the stiffness constants. The elastic moduli and Poisson's ratio of Al_3Nb were determined by the resonance frequency method (Reip, [13]). For comparison the corresponding elastic constants of Al_3Ti were calculated by Fu [14] and Yoo and Fu [15]. The calculated elastic constants are in reasonably good agreement with experimental values.
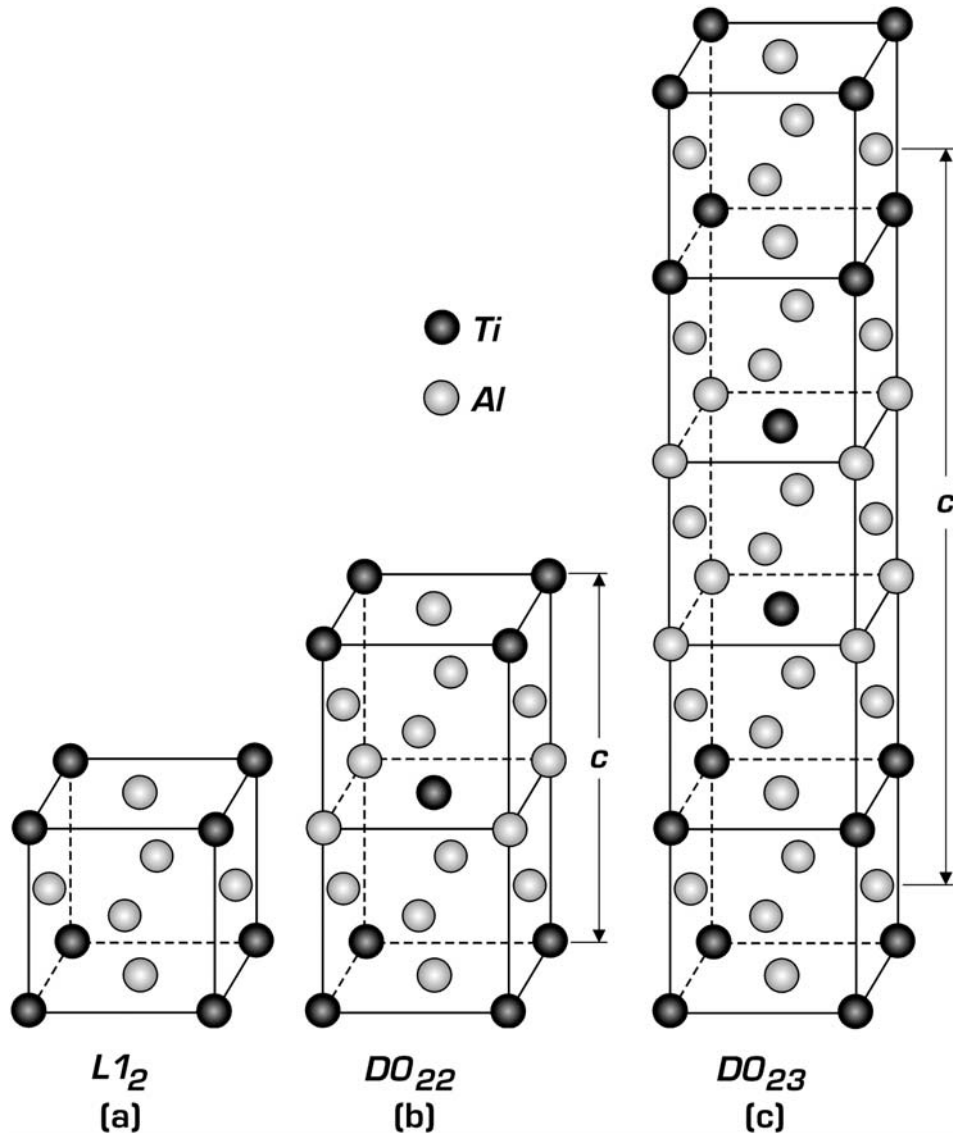
Figure 1 Elementary cells of the ordered L1$_2$, DO$_{22}$ and DO$_{23}$ structures.

Table 1 Physical Properties of DO$_{22}$-Trialuminides Al$_3$X

	Melting temperature (K)	Density ρ [Mg m^{-3}]	Linear thermal expansion coef. α 10^{-6}, [K^{-1}]	Spec. electrical resistance ρ 10^{-8}, [mΩ]
Al$_3$Ti	1620	3.36	9.75–9.9	27
Al$_3$Nb	1880	4.54	9.6–9.8	49

Table 2a   Hardness and Elastic Properties

	Vickers' Hardness (HV$_{10}$)	Young's Modulus E (GPa)	Shear Modulus G (GPa)	Poisson's ratio ν
Al$_3$Ti	390–420	166	90	0.16–0.19
Al$_3$Nb	565	246	105	0.17–0.19

Both, experimentally and theoretically determined elastic moduli E and G of Al$_3$Ti are approximately three and a half times those for pure aluminium and about two and a half times those of pure titanium, even though its primary constituent is aluminium and hence the density of Al$_3$Ti is only 24% higher than that of pure aluminium. The specific Young's and shear moduli of Al$_3$Ti are much higher than those of the constituent elements. The elastic stiffness constants of Al$_3$Zr, which is another example of a compound with an L1$_2$-derivative, long-period structure, have also been measured by Nakamura and Kimura [12], and the values of E and G have also been found to be much higher than the corresponding values for aluminium. Thus, this seems to be characteristic of this class of trialuminides. In contrast, the bulk modulus of Al$_3$Ti is only 44% higher than that of pure aluminium. This is also the case for Al$_3$Nb. Thus, these trialuminides, in general, have smaller K/G ratios, which correlate with the tendency toward brittleness for the trialuminides since a high value of G implies a high critical stress for dislocation generation at the crack tip.

3   DISLOCATIONS, BURGERS VECTOR

The atomic arrangement and planar faults on closed packed (111) planes in an Al$_3$ X compound with DO$_{22}$ structure is shown in Fig. 2. The periodic introduction of antiphase boundaries APB into the L1$_2$ structure transforms the triangular arrangement of B atoms on (111) in the L1$_2$ structure into a rectangular arrangement. From geometrical considerations, five possible planar faults have to be considered. These are the antiphase boundaries: APB-I, APB-II, the complex stacking faults: CSF-I, CSF-II and the superlattice intrinsic stacking fault: SISF. The stacking fault energies have been calculated by Yamaguchi and Umakoshi [4].

The SISF, which involves no first- and second-nearest neighbor wrong bonds, is expected to have generally a much lower energy than the other possible faults on {111} such as APB and CSF. The displacement vector $\mathbf{b}_{Al}$ forms an APB also on (001) without creating first nearest neighbor wrong bonds. Its energy is given by:

$$E_{\text{APB-I}}^{(001)} = \frac{2V^{(2)}}{a^2} \tag{1}$$

For the DO$_{22}$ structure to be stable with respect to both the DO$_{23}$ and L1$_2$ structures $E_{\text{APB-I}}^{(001)} > 0$ where $V^{(2)}$ is the second neighbour ordering energy. The energies of the five possible faults on {111} are thus generally expected to be in the following decreasing order:

$$E_{\text{CSF-I}} > E_{\text{CSF-II}} > E_{\text{APB-I}} > E_{\text{SISF}} \tag{2}$$

APB-I on (001) has a lower energy than CSF-I, CSF-II, APB-I and APB-II on {111}.

Table 2b: Stiffness and Elastic Constants of Al$_3$Ti and its Constituent Elements (10^2 GPa)

	C_{11}	C_{33}	C_{44}	C_{66}	C_{12}	C_{13}	Bulk moduli K	Young's moduli E	Shear moduli G	Poisson's ratio ν	specific moduli E/ρ	G/ρ
Al$_3$Ti$_{(exp.)}$	2.177	2.175	0.920	1.165	0.577	0.455	1.056	2.157	0.930	0.160	0.63	0.27
Al$_3$Ti$_{(theo.)}$	2.02	2.43	1.45	1.00	0.88	0.60	1.22	2.30	0.97	0.19	0.68	0.29
Al	1.082	C_{11}	0.285	C_{44}	0.613	C_{12}	0.752	0.706	0.262	0.345	0.26	0.10
Ti	1.624	1.807	0.467	C_{44}	0.920	0.690	1.084	1.202	0.456	0.361	0.27	0.10
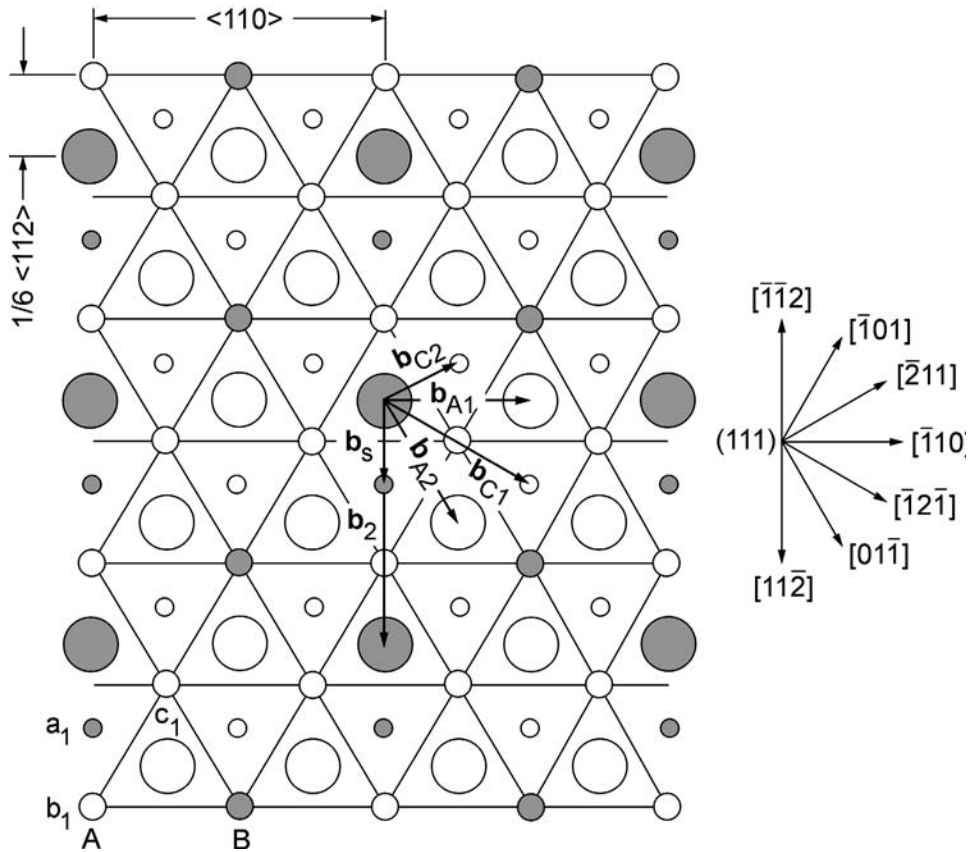
Figure 2 Atomic arragement on the (111) plane in A_3B compound with DO_{22} lattice structure. The a_1, b_1 and c_1 layers represent the part of the stacking sequence $a_1 b_1 c_1 a_2 b_2 c_2$...... along the [111] direction.

The APB-I on (001) is always stable for obvious symmetry reasons, while the stability of the faults on {111} is not guaranteed on symmetry grounds Yamaguchi et al. [16,17]. They can be unstable.

The DO_{22} structure exhibits relatively short lattice translation vectors along $\langle 110 \rangle$ and $\langle 112 \rangle$ on {111}, and along $\langle 110 \rangle$ and $\langle 100 \rangle$ on (001). These vectors are $\langle 110 \rangle$ (e.g. $2\mathbf{b}_{A1}$ in Fig. 2), $\frac{1}{2}\langle 112 \rangle$ (e.g. $\mathbf{b}_2$ in Fig. 2) and $\langle 100 \rangle$, respectively. Dislocations with Burger vectors corresponding to such relatively short lattice translation vectors are perfect ones, which are expected to be observed in compounds with this structure. Such perfect dislocations may be dissociated into partial dislocations, and the character of dislocation dissociation will depend on the relative stabilities and energies of the different faults involved.

The twinning systems of the {111}⟨112⟩ type are the variants in which the DO_{22} superlattice is mechanically twinned without disturbing the DO_{22} symmetry of the lattice. This is called "ordered twinning" since the ordered arrangement of atoms in the lattice is still retained. On the (111) plane in the DO_{22} structure, SISF corresponds to the stacking fault in the underlying f.c.c. structure. The ordered twinning on the

$(111)[11\bar{2}]$ system may result from the propagation of Shockley partials with $\mathbf{b} = \mathbf{b}_s$ on the adjacent (111) planes trailing the SISF.

4 PLASTIC DEFORMATION OF Al$_3$Ti AND Al$_3$Nb

4.1 Deformation Mode

The major deformation mode of these compounds is ordered twinning of the $\{111\}\langle 112]$ system as found earlier for the deformation of Ni$_3$V by Vanderschaeve and Sarrazin in 1977 [18]. The obtained results have been confirmed by Vanderschaeve and Escaig in 1983 [19]. TEM observation of deformation structures in Al$_3$Ti have been reported by Yamaguchi et al. in 1988 [3], Vasudevan et al. in 1989 [20], and by Morris and Lerf in 1991 [21]. Reip and Sauthoff [10] showed similar deformation mode for Al$_3$Nb.

Figure 3a represents deformation twins in Al$_3$Ti foils of single crystals deformed in compression at $800°$C. The (111) twin boundaries are viewed end-on. The corresponding selected area diffraction pattern (SAED) is shown in Fig. 3b, in which the fundamental and DO$_{22}$ superlattice reflections occur not only from the matrices, but also from the twins.

Figure 4 shows a high-resolution electron microscopy (HREM) image. The twin boundary, which is located horizontally at the center of the image, is seen to be atomistically flat and parallel to (111). The atomic arrangement characteristic of Al$_3$Ti projected onto $(1\bar{1}0)$ is evident in both the matrix and the twin. Image calculations indicate that the bright spots correspond to the atomic rows composed of alternately arranged Al and Ti atoms while weakly bright spots correspond to Al atoms. Twinning does not disturb the DO$_{22}$ order. This is consistent with the SAED pattern in Fig. 3b.

At high temperature, the four $\langle 112]$-type ordered twinning systems are augmented by slip of the types $\langle 110]$, $100]$ and $\langle 112]$, respectively. In particular, significant deformation occurs by slip on $(001)\langle 110]$ at temperatures higher than $600°$C. Dislocations with $\mathbf{b} = \langle 110]$ on (001) are widely dissociated into two superlattice partial dislocations with $\mathbf{b} = \langle 110]$ separated by a ribbon of APB-I on (002) as illustrated in Fig. 5, a bright-field image of superlattice dislocation with $\mathbf{b} = [110]$ in an Al$_3$Ti single crystal deformed at $800°$C. The figure is viewed nearly along $[001]$ with $\mathbf{g} = 0\bar{2}0$. The corresponding dark-field image with (201) superlattice reflection shows the ribbons of APB-I with the displacement vector $\frac{1}{2}[110]$.

These types of dislocations are splitting on $\{111\}$ planes; The spacings between the partials on (001) and $\{111\}$ have been measured, giving the values of APB-I energies of 25–32 mJm^{-2} on (001) and 190–200 mJm^{-2} on $\{111\}$ [22]. The energy of the $\frac{1}{2}\langle 110\rangle$-type APB in the L1$_2$ structure is also lowest on $\{001\}$, and the motion of pairs of superlattice partial dislocations with $\mathbf{b} = \frac{1}{2}\langle 110\rangle$ joined by the APB on (001) is the major high-temperature deformation mode in many L1$_2$ compounds. Slip of the type $(001)\langle 110]$ may be a common high-temperature deformation mode in L1$_2$ and L1$_2$-derivative, long-period ordered structures.

Perfect dislocations with $\mathbf{b} = \frac{1}{2}\langle 112]$ (e.g. $\mathbf{b}_2$ in Fig. 2) were not detected in the earlier studies [2]. Dislocations of the $\frac{1}{2}\langle 112]$-type have recently been found to exist in Al$_3$Ti. These are dissociated as $\frac{1}{2}\langle 112\rangle \rightarrow \frac{1}{6}\langle 112] + \frac{1}{3}\langle 112]$, and the separation has been measured [21]. From this the SISF energy of 91–1097 mJm^{-2} have been determined. This value is not so low as it would be expected from the abundance of ordered twins
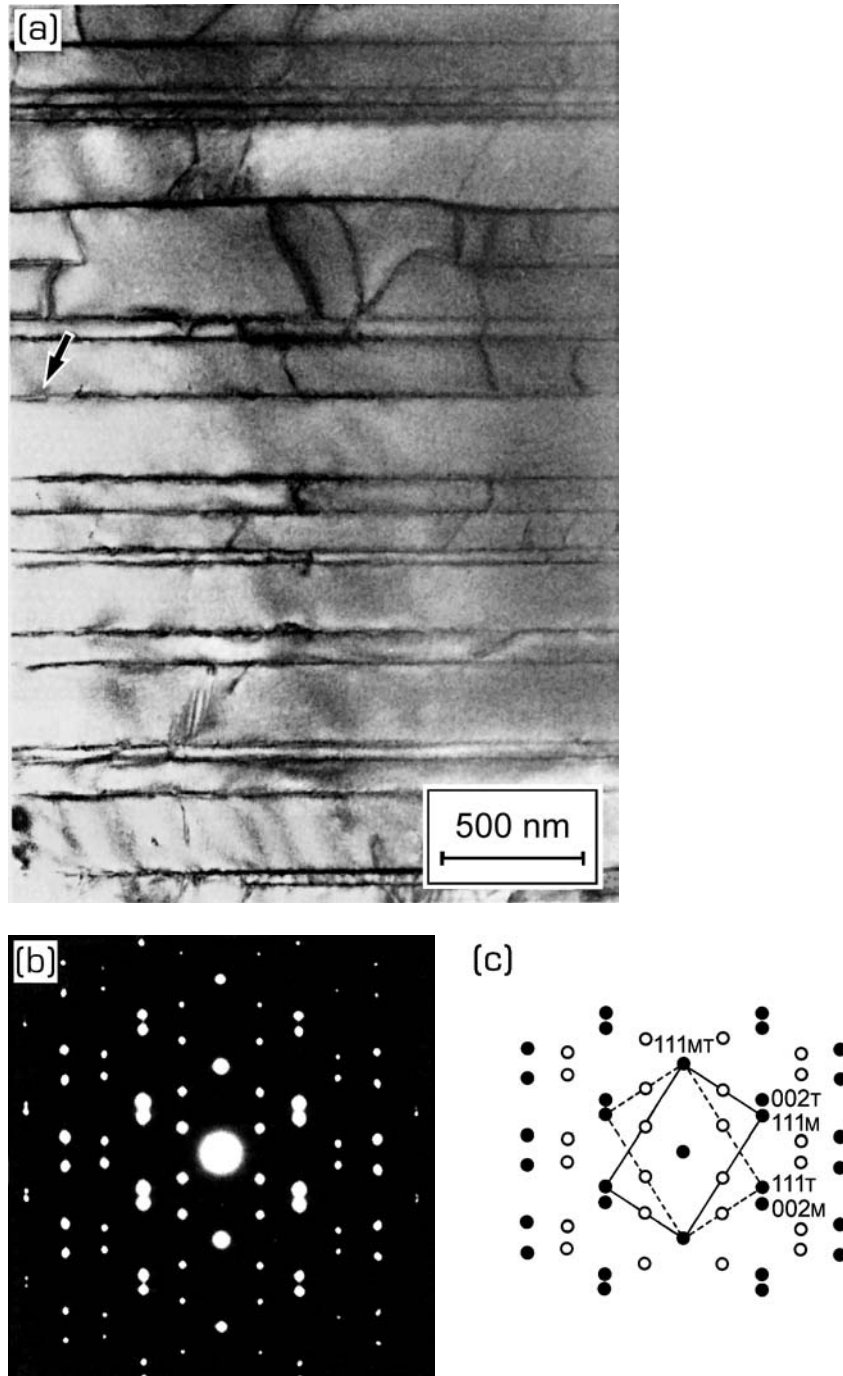
Figure 3 (a) TEM bright-field image of deformation twins in a single crystal of Al_3Ti deformed in compression at $800°C$. (b), (c) Corresponding SAEDX-pattern; indexed closed and open circles represent fundamental and superlattice reflections (Yamaguchi et al.[1988], Morris and Lerf [1991], Vanderschaeve et al.[1989]).
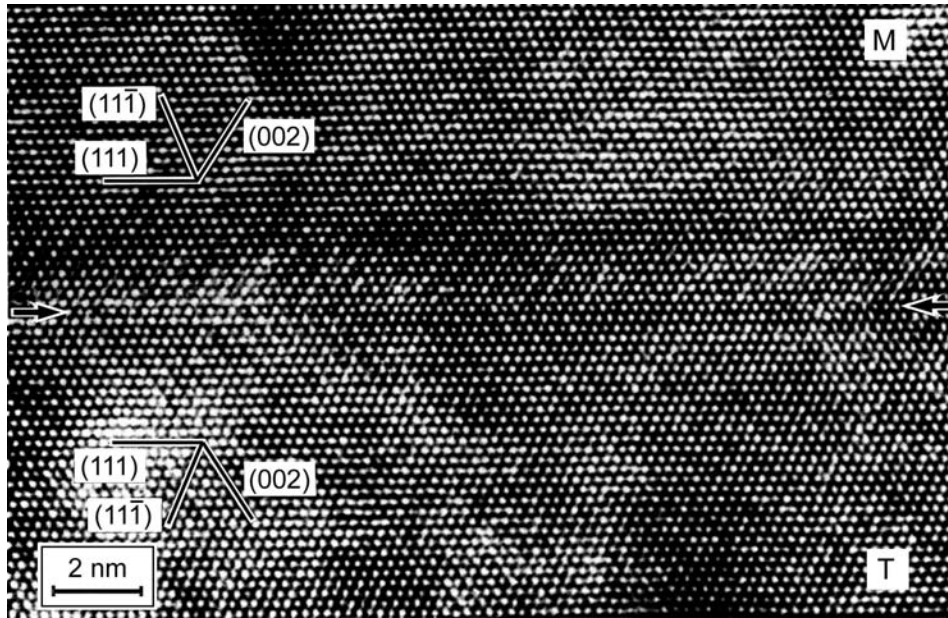
Figure 4   HREM image of deformation twin in the single crystal of Al₃Ti.
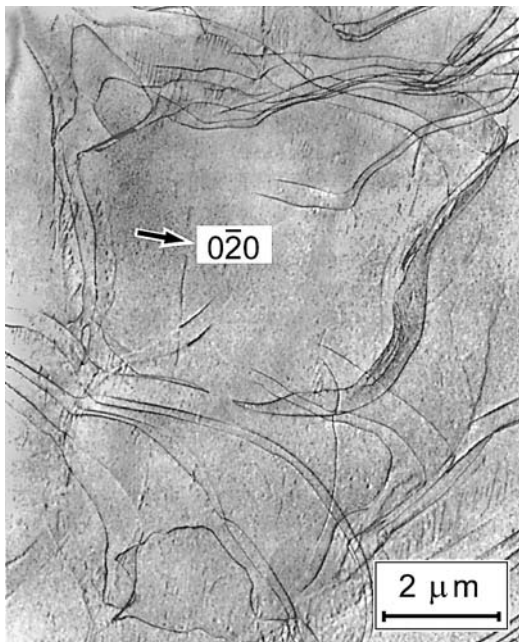
Figure 5   Bright-field image of dislocation structures in a Al₃Ti single crystal deformed in compression at 800°C.

in deformed specimens of Al_3Ti. The observation of extensive twinning in this compound is believed to be a reflection of the difficulty of dislocation propagation rather than a reflection of easy twinning. The only glissile screw dislocation with a planar core has been found to be the Shockly partial $\frac{1}{6}\langle 112]$ bounding an SISF. Other screw dislocations, such as the partial dislocation $\frac{1}{3}\langle 112]$,
and the superlattice partial dislocation $\frac{1}{2}\langle 112]$, bounding an APB-I on (001), are all sessile, and hence thermal activation is needed for their movement.

Bands of $\frac{1}{2}\langle 112]$ dislocations trailing the APB-I on (001) and loosely coupled in pairs have been often found where twins intersect. The $(001)\langle 110\rangle$ slip corresponds therefore to the conjugate deformation mode to the complementary ordered twinning system [23].

4.2 Temperature Dependence of Strength – Single and Polycrystals

Typical stress-strain curves of Al_3Ti single crystals at $800°C$ in two different compression axis orientations are shown in Fig. 6. In region I of the $[100]$–$[010]$–$[\bar{1}10]$ stereographic triangle all four ordered twinning systems can be activated since twins on the four ordered twinning systems produce compression strain, and hence deformation by ordered twinning can occur rather easily when compared with other orientations. As seen from the stress-strain curve for orientation (A) serrations occur for orientations in this region. This feature is due to the discontinuous generations of twins.

In region II, two of the four ordered twinning systems can be operative and none of them in region III. For orientations in region II with large values of the
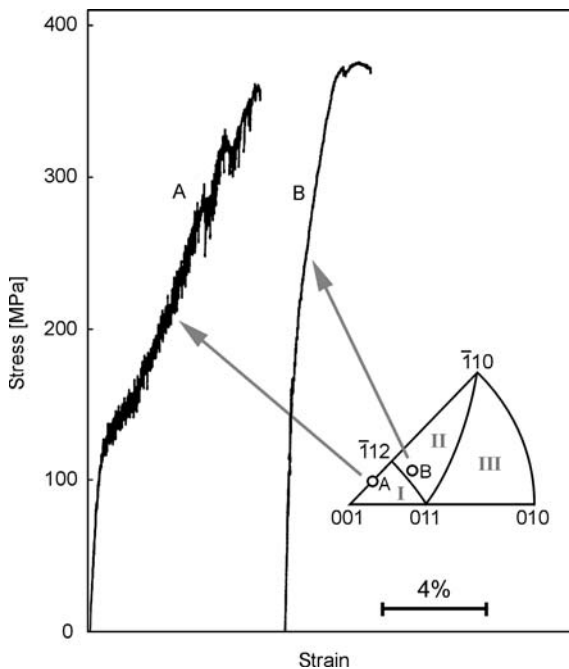
Figure 6   Stress-strain-curves of Al_3Ti single crystals with different orientations.

Schmid factor for slip on (001) [1$\bar{1}$0], a smooth stress strain curve exhibits which is like that for orientation (B). The twinning stress is plotted as a function of temperature in Fig. 7. Although the scatter in the results is large, the resolved twinning stress is almost independent of temperature. In contrast, the yield stress exhibited by specimens which are deformed by slip in the (001)⟨110] glide system shows a strong temperature dependence. The critical resolved shear stress (c.r.s.s.) for slip, calculated by resolving the yield stress on the corresponding slip system, is shown as a function of temperature in Fig. 7 (dashed line). This strong temperature dependence of the c.r.s.s. for slip of the (001)⟨110] type in the DO_{22} structure has been interpreted in terms of the low mobility of superlattice partial dislocations of the type $\mathbf{b} = \frac{1}{2}$⟨110] due to their sessile core structure by Khantha, Vitek, and Pope [24].

Polycristalline Al_3Ti material prepared by conventional vacuum metallurgy technology contains small volume fractions of finely dispersed secondary phases, mainly aluminium. This is due to the peritectic reaction and the formation of the in congruently solidifying intermetallic compound.

The dispersion of aluminium surrounding the compound grains will enhance the plastic deformation of Al_3Ti. The two-phase Al_3Ti/Al mixture containing more than
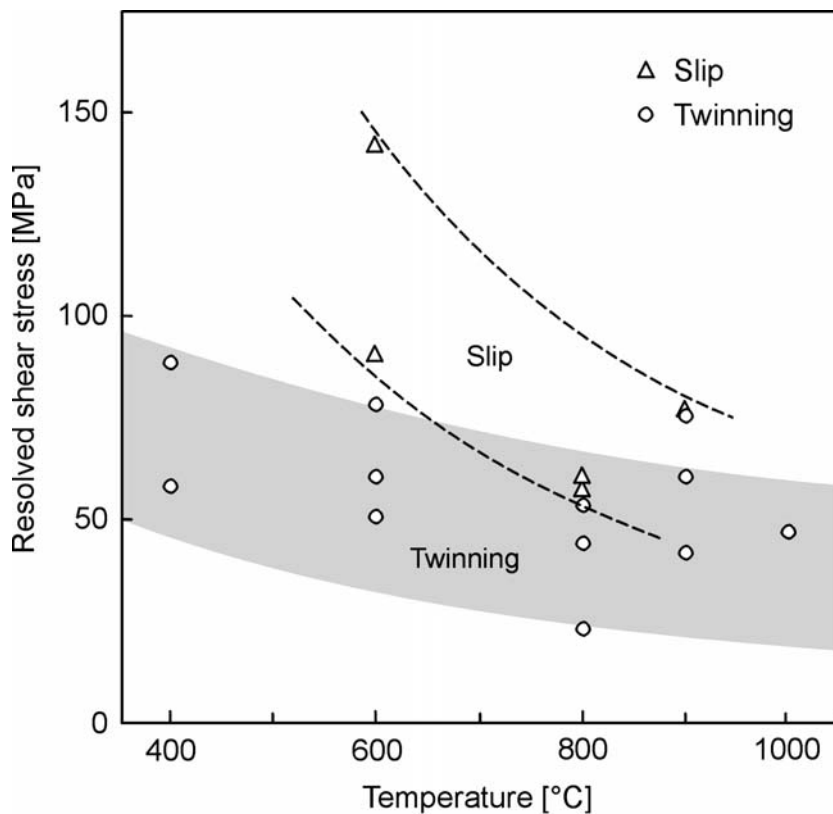
Figure 7 Resolved shear stress for slip and twinning in Al_3Ti single crystals in dependence on temperature.

85 vol% Al_3Ti can be extruded at 600°C with a ratio of reduction in area of 10:1. The overall deformation must be much greater than the deformation of the hard Al_3Ti component of the composite, which suffers significant plastic deformation. In particular, fine-grained, two-phase composites exhibit good compressive ductility even at room temperature. The following interpretation has been proposed: the presence of aluminium surrounding the Al_3Ti grains provides a sink for the relief of stress concentrations during the deformation of Al_3Ti, leading to a pseudo-augmentation of the low-temperature deformation of Al_3Ti [25].

In contrast to Al_3Ti and Al_3V Al_3Nb melts congruently. This makes the production of Al_3Nb in larger quantities easier. The remarkable high elastic stiffness of Al_3Nb at elevated temperature, e.g. Young's modulus of 225 GPa at 600°C and its high proof strength and creep resistance at temperatures beyond 1000°C are of considerable importance for high temperature applications. Reip and Sauthoff studied the short-term behaviour [10] and the creep properties of pure Al_3Nb and of some alloy modifications, such as $Al_3(Nb_{0.75}Ti_{0.25})$ and some others [10,13].

The temperature dependent yield stress of polycrystalline Al_3Nb and $Al_3(Nb_{0.75}Ti_{0.25})$ compounds and for comparison the Al_3Ti compound are shown in Fig. 8. At medium temperatures between 600°C and 775°C Al_3Nb exhibits pest phenomena because of preferential grain boundary oxidation [26,27]. However, this problem can be overcome by alloying titanium to the Al_3Nb compound. From Fig. 8 it can be seen that the proof stresses of the Al_3Nb and $Al_3(Nb_{0.75}Ti_{0.25})$ alloys are considerably higher than those of Al_3Ti.

The active slip systems are identical for all these compounds, and the major deformation mode is twinning revealing in individual Al_3Nb grains, shown in Fig. 9. The fracture toughness of Al_3Nb based compounds are rather poor. The critical stress intensity fractor K_{IC} at room temperature is about 2.5 to 3 MPa $\sqrt{m}$ and increases with increasing test temperature up to 5 MPa $\sqrt{m}$ at 800°C. However, above 400°C macroscopic plasticity, indicated by a plastic zone at the crack tip, gives evidence for thermally activated dislocation emission and dislocations glide, which is also strongly related to the applied strain rate.

The creep strength of the stoichiometric unalloyed Al_3Nb material is relatively low, e.g. a stress of 10 MPa leads to an amount of 1% creep strain in 500 h, and creep rupture occurs in 2300 h. This unique behaviour is in contrast to the high yield stresses achieved, which compare favourably with those of super alloys.

This fact illustrates the different high temperature strengthening behaviour of intermetallics with complex DO_{22} structure and multi component superalloys with large volume fractions of $L1_2$ or B2 compounds with higher symmetry finely dispersed through out an f.c.c. or b.c.c. nickel, cobalt or iron matrix.

Creep in Al_3Nb is due to dislocation climb, accompanied by subgrain formation. The secondary creep rate is described by the Dorn equation for dislocation creep:

$$\dot{\varepsilon} = A \frac{D_{\text{eff}} G |\boldsymbol{b}|}{k \cdot T} \left(\frac{\sigma}{G}\right)^n \tag{3}$$

where $\dot{\varepsilon}$ is the strain-rate, A is a structure term, $D_{\text{eff}} = D_0 \exp(-Q/RT)$ is the effective diffusion coefficient, G represents the temperature dependent shear modulus, $\boldsymbol{b}$ is the Burgers vector, T is the test temperature, k is the Boltzmann constant, σ is the applied stress, and n is the stress exponent. Fig. 10 illustrates secondary creep curves of a variety of $Al_3(Nb, Ti)$ and $Al_3(Nb, Ni)$ alloys [25]. The stress exponent n is between 3 and
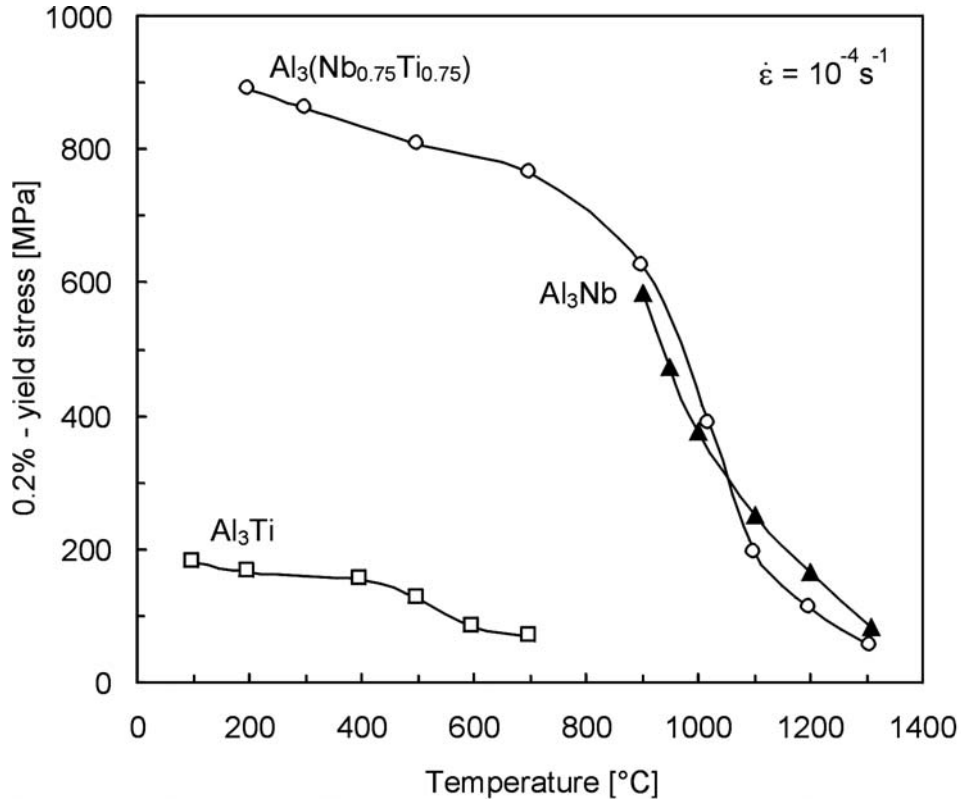
Figure 8   Yield stress vs. temperature curves for Al_3Ti, Al_3Nb and $Al_3(Nb_{0.75}Ti_{0.75})$ (Reip and Sauthoff [1993]).

5 and indicates clearly power low creep. The n value was determined to be 3.6 and the overall activation energy was 360 kJ/mol. The temperature dependent activation energy of the shear modulus is about 35 kJ/mol and almost the remaining 325 kJ/mol refers to thermally activated diffusion processes involving dislocation climb for subgrain formation and reinstating the ordering.

The isostructural compound Al_3V is regarded as promising for applications in nuclear reactor technology. The melting temperature is only slightly higher than that of Al_3Ti and a similar deformation behaviour is expected. However, the yield stress is higher than that of Al_3Ti by a factor of about 2 and there is no ductility in compression below about 400°C.

Microalloying with titanium improves the ductility in compression–at the expense of the yield stress – to such an extent that $Al_3(V_{0.95}Ti_{0.05})$ shows a compressive strain of about 7%. Increasing the Ti content leads to a deterioration in the properties, i.e. $Al_3(V_{0.75}Ti_{0.25})$ is significantly stronger than Al_3V, but can only be deformed plastically at 700°C and above.

It is noted that according to first-principles calculations, the Al_3Ru compound also has the DO_{22} structure and is closely related to Al_3V. The deformation behaviour of Al_3Ru is expected to be similar to that of Al_3V but with enhanced ductility. The main strategy for improving the ductility of Al_3X types of aluminides is to decrease the APB
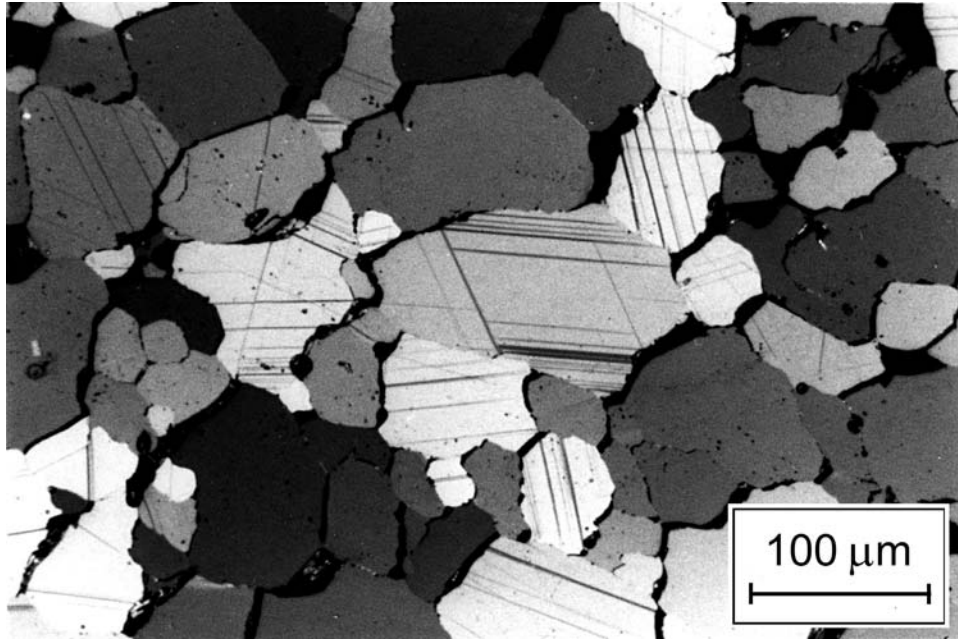
Figure 9   Twinning formation in Al$_3$Nb polycrystals (Reip and Sauthoff [1993]).

energy on the (001)-plane. Introducing a $\frac{1}{2}\langle 110\rangle$ APB on (001) in the DO$_{22}$ structure is equivalent to creating a local stacking of (001) layers of the DO$_{23}$ or L1$_2$ type. This would also reduce the magnitude of the Burgers vector of the dislocations carrying slip on (001)$\langle 110]$. Another important aspect is the refinement of slip and/or twins. The presence of massive macro twins often leads to brittle fracture [29]. As already pointed out, alloying of Al$_3$V by Ti leads to an enhancement of twinning and a refinement of twins und consequently the compressive fracture strain of the Al$_3$(V,Ti) aluminides will be improved.

5   (Al,X)$_3$Ti ALUMINIDES WITH L1$_2$ STRUCTURE

The ordered f.c.c. L1$_2$ structure of Al$_3$Ti will be stabilized if an appropriate number of d-electrons is removed from the Ti sites. This can be accomplished by decreasing the charge transfer from Al to Ti by replacing Al atoms with other less electro-positive elements which act as electron acceptors. The DO$_{22}$ type of structure will be changed to the L1$_2$ structure by alloying Al$_3$Ti with transition metals X, such as Cr, Mn, Fe, Co, Ni, Cu, Rh, Ir, Pd, Pt, Ag and Au, respectively.

X in (Al,X)$_3$Ti stands for one or more substitutional elements in ternary, quaternary and even in quinary compounds. After Kumar et al. the Cr-, Mn- und Fe-modified L1$_2$ trialuminides exhibit continuous solubility in each other [30–32]. However, the maximum concentration of substitutional elements in the L1$_2$ phase is in the range of 4 to 12 at%, regardless of the number of substitutional elements. Several parameters, such as electron-to-atom ratio, lattice constant of the L1$_2$ phase, Pettifor's Mendeleev number [33] and atomic radius, have been plotted versus the amount of
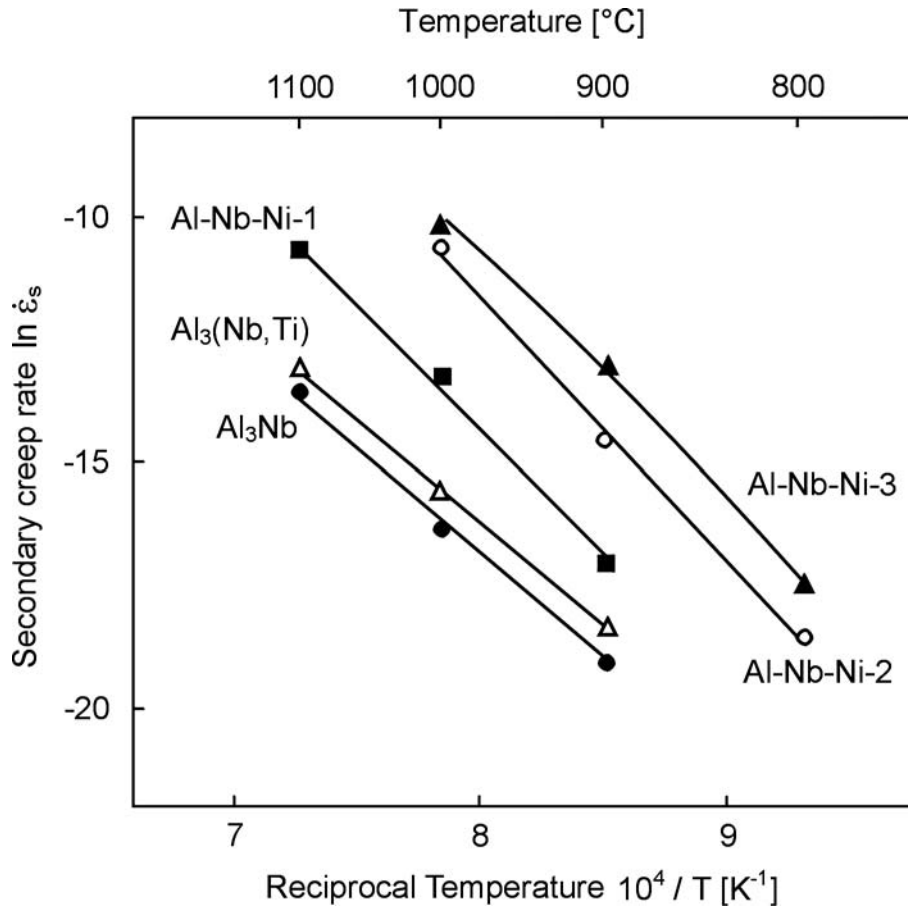
Figure 10 Secondary creep rate vs. reciprocal temperature of Al_3Nb and divers $Al_3(Nb, Ti, Ni)$ alloys (Srinivasan et al. [1992]).

element X in the $L1_2$ phase [34]. The best correlation has been found to be between the amount of element X in the $L1_2$ phase and the shortest distance between atoms of the element X in its pure structure. This seems to suggest that the simple charge transfer model may not be enough to describe quantitatively the observed $L1_2$ stabilization effects of ternary alloying additions. Recently, Hong and Freeman [35] have made first-principles calculations on the effects of substituting Cu for either Ti or Al in Al_3Ti and reported that Cu as a ternary addition strongly favors the Al site and the substitution of Cu for Al causes a similar change in bonding character in the new material as tretagonal distortion did in pure Al_3Ti [36]. Similar results have been achieved also by Eberhart et al. [37]. Such efforts to specify the role of ternary alloying elements in structural stability are needed for a more complete understanding of the stabilization of the $L1_2$ structure in the trialuminides.

Also for the DO_{23} compound Al_3Zr a partial substitution of ternary elements, such as Cr, Fe, Ni and Cu for Al has been found to result in ternary $(Al,X)_3Zr$-based $L1_2$ phases [38]. The elements V, Co [34] and Mn [39,40] have been added to the list of alloying elements known to form ternary $(Al, X)_3Zr$-compounds.

Table 3 Summary of results of experiments on different $(Al,X)_3Ti$ alloys

Alloy	Preparation method[a]	Measurement[b]	Temperature	Remarks	References
$Al_{67}Cr_8Ti_{25}$ $Al_{66}Cr_9Ti_{25}$	CIP→AM→HIP	Hardness Compressive strength Bend ductility	RT RT RT	0.5 % bend ductility	Zhang et al., (1990)
$Al_{67}Cr_8Ti_{25}$	CIP→ST→HP	Hardness Cempressive strength	RT RT – 100°C	Flow stress peak at 600C	Mabuchi et al., (1990)
$Al_{67}Cr_8Ti_{25}$	AM→HE	Compressive strength Bend ductility Elastic parameters	RT RT RT	0.16% bend ductility	Schneibel et al., (1992)
$AL_{67}Cr_8Ti_{25}$	SD→HIP	TEM	RT	APB-type dissociation (4–5nm)	Morris, (1992)
$Al_{67}Cr_8Ti_{25}$	IM→HT→HF	Tensile strength and ductility	RT – 800°C	0.2% tensile ductility at 350°C	Kumar and Brown, (1992)
$Al_{67}Mn_8Ti_{25}$	CIP→AM→HIP	Hardness Compressive strength	RT RT		Zhang, Nic and Mikkola, (1990)
$Al_{66}Mn_9Ti_{25}$	IM→HT→HF	Tensile strength and ductility TEM	RT – 800°C 350°C 650°C	0.2% tensile ductility at RT Undissociated APB-type dissociation	Kumar and Brown, (1992)
$Al_{66.7}Fe_{9.1}Ti_{24.2}$	CIP→ST→IM→H T→CIP→ST	Compressive strength	−196°C −750°C	Well-defined flow stress peak at 450°C	Kumar and Pickens, (1988)

Composition	Preparation	Measurement	Temperature	Comments	Reference
$Al_{65.2}Fe_{8.2}Ti_{26.6}$	single crystals grown by floating zone				
$Al_{66.8}Fe_{5.8}Ti_{27.4}$	Single crystals grown by Bridgman method	Compressive strength	$-196^{\circ}C$ $-1000^{\circ}C$	No flow stress peak	Wu et al., (1991)
$Al_{67}Co_8Ti_{25}$ $Al_{67}Ni_8Ti_{25}$	AM→HIP AT→HIP	Hardness TEM	RT $300^{\circ}C$	Undissociated	Nic et al., (1990) Vasudevan et al., (1989)
$Al_{67}Ni_8Ti_{25}$	AM→HT→HIP	Hardness Compressive strength Elastic parameters Bend ductility Fracture toughness	RT RT RT RT RT	No appreciable bend ductility $3\ MPa\ m^{-1/2}$	Turner et al., (1989)
$Al_{67}Pd_8Ti_{25}$	AM→HT	Hardness Bend ductility TEM	RT RT RT	No appreciable bend ductility Undissociated	Powers and Wert, (1990)
$Al_{67}Ag_8Ti_{25}$	CIP→ST→HP	Hardness Compressive strength	RT–$1000^{\circ}C$	Double flow stress peaks at 200 and $600^{\circ}C$	Mabuchi et al., (1990b)

[a] Glossary of preparation methods: CIP, cold isostatic pressing; AM, arc melting; IM, induction melting; ST, sintering; HP, hot pressing; HIP, hot isostatic Pressing; HF, hot forging; SD, spray deposition; LM, levitation melting; HT, heat treatment; HE, hot extrusion; RQ, rapid quenching; AT, atomizing.
[b] Glossary of measurements: TEM, transmission electron microscopy.

For the DO_{22} compound of Al_3Nb no ternary $L1_2$ phases have been found in the vicinity of the trialuminide composition [41]. This is because the differences in the electronic total energy for the Group IV transition metal trialuminides between the stable DO_{22} or DO_{23} and the hypothetical $L1_2$ phases are much smaller than those for Group V transition metal trialuminides [42]. Single-phase $L1_2$-structured Al_3Ti, Al_3Zr and Al_3Hf might be achieved through mechanical alloying of a mixture of pure elemental powders at the corresponding composition, although they transform to the stable structure in the temperature range from 500 to 700°C [43].

5.1   Processing and Microstructure

Table 3 lists a representative number of $Al_3(Ti, X)$ alloys – $L1_2$ variations – investigated together with the preparation methods and measured properties. The powder processing, rapid solidification, arc melting, induction melting, levitation melting, hot isostatic pressing (HIP), hot forging and hot extrusion will cause different mechanical properties, even though the compounds have the same composition. This indicates that the microstructure and mechanical properties of differently processed and investigated $Al_3(Ti, X)$ alloys with $L1_2$ structure may not be directly comparable, as evidenced by the fact that their mechanical properties often vary within a certain scatter band. Results of various studies [43] showed that the $L1_2$ phase field contracts with decreasing temperature, leading to precipitation of a second and sometimes a third phase at low temperatures. It is difficult to obtain single-phase $L1_2$ trialuminide compounds prepared by conventional ingot metallurgy. These normally contain precipitates of second and third phases. Even when compounds appear to be single phase on the scale of optical microscopy, finely dispersed secondary phase precipitates are often observed in TEM (Wu et al. [44]; DiPietro et al. [45]; Nick et al.[46]). The distribution of such precipitates tends to vary depending on heat-treatment parameters, such as temperature and cooling rate. The mechanical properties, specifically yield stress, strength, ductility and fracture toughness may vary with the distribution of precipitates and other defects, such as pores and micro blow holes. Precipitates may be eliminated by a homogenization heat treatment at temperatures generally above 1000°C. The same homogenization treatment, however, has been known to increase the volume fraction of pores through the Kirkendall mechanism.

Recently, hot forging, (Kumar and Brown, [31]; Chen et al., [47]) and hot extrusion (Schneibel et al. [48]) have been shown to be effective in minimizing the volume fraction of pores and secondary phase precipitates, and hence increasing the ductility of $L1_2$ trialuminide compounds. Another important fact is that hot-worked compounds generally show fine-grained microstructures with almost no pores. If the chemical composition is chosen properly, they contain only a small amount of secondary phase precipitates compared to as-cast ingots. No appreciable plastic strain has been observed in bending or tension test of $L1_2$ trialuminides produced without hot working. However, Schneibel et al. [48] have observed that small but definite plastic strain of 0.16% in bending of a hot-extruded Cr-modified compound at room temperature occured. This clearly indicates the importance of microstructural control to achieve optimum mechanical porperties in the $L1_2$ trialuminide compounds. Hardness measurements have been extensively used to evaluate the brittleness by measuring the total crack length formed around a Vickerś indent or the maximum crack-free load (Zhang et al. [49]; Nic et al. [46]). However, in evaluating the strength and brittleness

by hardness measurements, it has to be considered that the hardness of $L1_2$ trialuminides is load dependent. It decreases with load until a plateau value is reached. The maximum crack-free load is usually very close to the load value at which the hardness becomes load dependent. Hardness measurements showed that the strength and brittleness of ternary $L1_2$ trialuminides vary greatly with ternary alloying elements, but their hardness is always lower than that of binary Al_3Ti regardless of ternary alloying elements (Kumar and Pickens, [50]).

Interestingly, it was found that there is a systematic increase in hardness and a decrease in crack resistance with increasing atomic number of the ternary alloying element of the fourth period from Cr to Cu. There is a systematic decrease in lattice parameter and an increase in modulus with increasing atomic number of the ternary element. Mikkola et al. [51] suggested that the higher the atomic number of the ternary element, the stronger is the d-d bonding in the $L1_2$ trialuminide, leading to increased hardness and brittleness. A small plastic strain in bending has been observed only for Cr- and Mn-modified compounds. Winnicka and Varin [52], on the other hand, have shown that in trialuminides stabilized with Cr, Mn, Fe and Cu, the hardness is rather independent of the atomic number of the ternary element. The hardness increases almost linearly with increasing Ti content. A similar dependence of strength on Ti content has also been observed in compression for Mn- and Fe-modified compounds.

Nic et al. [46] investigated the variation in hardness with Ti and Cr content in Al-Cr-Ti compounds in which the phases present were identified by powder X-ray diffraction and TEM. As shown in Fig. 11, compounds with compositions within the $L1_2$ single-phase field have lower hardness values than compounds composed of two and three phases. In the composition range from 8 to 13 at% Cr and 23 to 25 at% Ti, where the lowest hardness occured, the hardness only slightly increases with Cr and Ti concentrations. However, in compounds with a Ti content higher than 25 at%, the hardness depends strongly on Cr and Ti content. In compositions containing more than 25 at% Ti, a large amount of finely dispersed precipitates have been observed in TEM. It is important to understand the mechanical properties of trialuminides in a microstructure/property relationship on the nanometer scale. X-ray and other measurements of the type and concentration of structural defects in trialuminides are needed to clarify the effects of off-stoichiometry on the mechanical properties.

5.2 Temperature Dependent Strength Properties

Several investigations on the temperature dependence of the compressive yield stress of various $(Al, X)_3Ti$ compounds have been performed since they are brittle under the action of tensile stress at low temperatures. Figure 12 shows some representative yield stress-temperature curves published for various ternary $L1_2$ trialuminides. The compositions are given in at%. The stress level and the shapes of the curves vary greatly with composition, even for the same nominal composition. This may be largely due to the difference in preparation method and hence the difference in microstructure of the compounds investigated. However, some general features of the yield stress-temperature profile occured.

The yield stress increases rather rapidly with decreasing temperature below room temperature. Since at low temperatures the yield stresses of Al_3Ti-based $L1_2$ compounds depend more strongly on temperature than those of $L1_2$ compounds, such as Ni_3Al – in which $\langle 110 \rangle$ dislocations dissociate into two $\frac{1}{2}\langle 110 \rangle$ superlattice partial
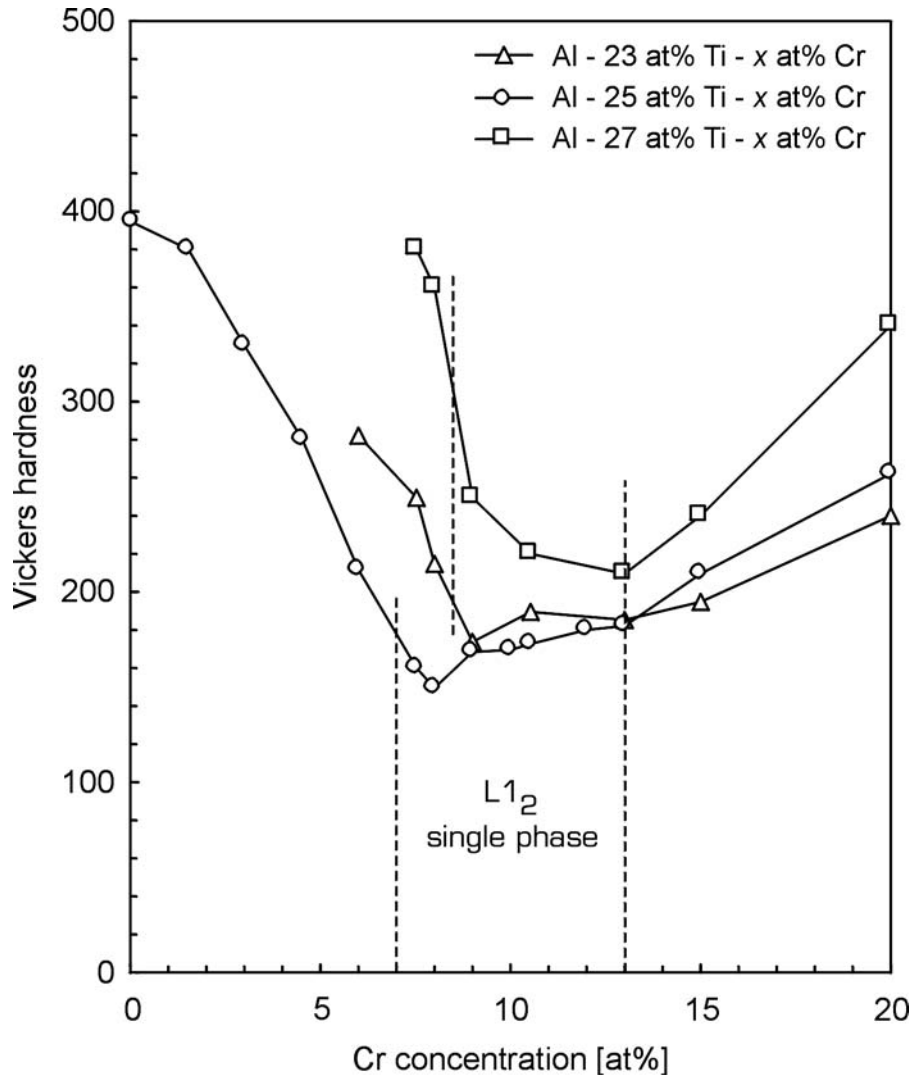
Figure 11   Variation in hardness of Al-Ti-Cr compounds (Nic et al. [1992]).

dislocations separated by an APB; a different dissociation scheme is believed to occur in the Al$_3$Ti-based trialuminides, if $\langle 110 \rangle$ dislocations dissociate into two $\frac{1}{2}\langle 112 \rangle$ superpartials separated by an SISF. The observed temperature dependence is well explained in terms of the nonplanar core structure of the superpartials (Yamaguchi et al. [53]; Paidar et al. [54]; Tichy et al. [55]). However, some investigatiors have reported either no dissociation or a dissociation of the APB type. Two exceptions are the results of Hu et al. [56] and Inui et al. [57]. The existence of a large number of antisite defects may contribute to strengthening at low temperatures.

Generally, the yield stress curves possess a plateau or a slight increase with temperature at intermediate temperatures. For iron and chromium modified compounds, most investigators have observed only a plateau or slightly anomalous behavior. The
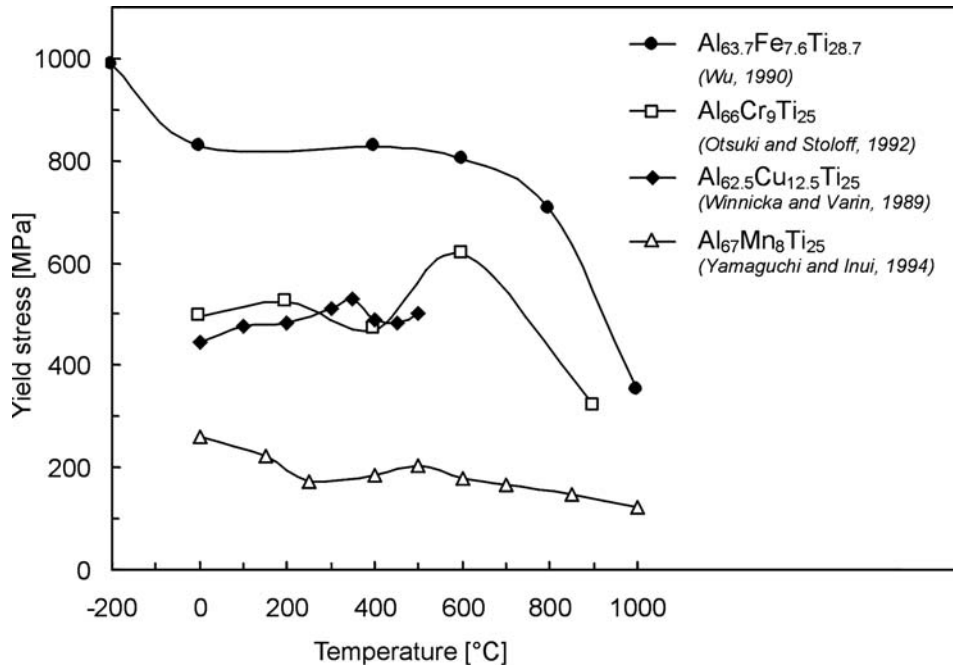
Figure 12   Yield stress vs. temperature curves of a variety of $L1_2$-Al-Ti-X alloys.

magnitude of the anomalous strengthening observed in $L1_2$ trialuminides is considerably smaller than that observed in $L1_2$ compounds such as Ni_3Al. Unlike Ni_3Al, the c.r.s.s. for $\{111\}\langle110\rangle$ slip in $L1_2$ trialuminides such as the Fe-modified ones is almost independent of orientation in the plateau or of the anomalous strengthening temperature range [44,58]. Thus, even if anomalous strengthening occurs in $L1_2$ trialuminides, the mechanism behind the anomaly in the $L1_2$ trialuminides may be different from the mechanism responsible for Ni_3Al. In the plateau or anomalous strengthening temperature range serrations have often been observed in stress-strain curves for most trialuminides, indicating that dynamic strain aging contributes to deformation [59–61].

The specific strength of high-strength aluminium alloys, such as 7075-T6 is higher than those of the trialuminides at temperatures lower than 150°C. However, the yield stresses of the $L1_2$ trialuminides are showing plateaus or a slight increase up to 600–800°C depending on composition, and thereby they are superior to high-strength titanium alloys in their specific strength at temperatures higher than 600°C. Even at room temperature, the strength can be markedly increased by utilizing precipitation hardening.

The yield stress rapidly decreases with increasing temperature up to 600°C or higher. The extent and the onset temperature of the rapid decrease of yield stress depend on the compound and the composition. For example, $Al_{65.2}Fe_{8.2}Ti_{26.6}$ exhibits a marked decrease in yield stress above 600°C, but to a much lesser extent than $Al_{63.7}Fe_{7.6}Ti_{28.7}$.

The yield stress markedly increases with increasing Ti content in compounds with higher Ti contents. The strong compositional dependence of the yield stress was

first noticed by Zhang et al. [49] in a series of Mn-modified compounds ($Al_{61}Mn_{10}Ti_{29}$, $Al_{67}Mn_8Ti_{25}$ and $Al_{68}Mn_9Ti_{23}$). The compound with the highest Ti content exhibits a yield stress that is almost twice as high as the others. The large increase in yield stress with Ti content in the Fe-modified compounds has been attributed to the existence of finely dispersed secondary phase precipitates, which were imaged in TEM. However, the effects of solid solution hardening due to structural defects, such as antisite atoms and vacancies, on the mechanical properties are not yet well known.

5.3 Plasticity and Fracture

$L1_2$ trialuminides apparently exhibit at room temperature ductility in compression, usually with a macroscopic fracture strain of more than 10%. However, this apparent ductility is accompanied by microcracking. Extensive internal and external microcracking are usually observed even at a plastic strain as low as 1%. In contrast, all $L1_2$ trialuminides appear to be very brittle under the action of a tensile stress, although a very small but definite plastic strain has been observed in Cr- and Mn-modified compounds when tested in bending. Consistent with this, the fracture toughness of most $L1_2$ trialuminides is in the range of 2–3 MPa m$^{1/2}$ (Turner et al. [62]; George et al. [5]). Fracture occurs by transgranular cleavage, largely along {110} and {111} planes. The addition of boron at levels of 25 to 1600 ppm has no beneficial effects on the improvement in ductility of $L1_2$ trialuminides. This may be partly because the fracture mode is transgranular cleavage. Wu et al. [44,48] have reported that single crystals of a Fe-modified compound show no higher ductility than polycrystals of the compound, even in compression. In single crystals tested in compression, fracture occurs by cleavage on a variety of planes including {110}, {111}, {001} and {013}, which is consistent with the result obtained for polycrystals tested in bending.

Uniaxial tensile tests have been carried out by Kumar and Brown [63] on Mn- and Cr-modified compounds. Plastic strains of 0.2% in the Mn-modified compound but not in the Cr-modified one were recorded at room temperature. These results are contrary to what has been observed in bending tests on Cr-modified compounds by Schneibel et al. [48]. It is reported that the ductility of Cr-modified $L1_2$ trialuminides is extremely sensitive to the surface-finish conditions of the specimens. Figure 13 shows the variation of the tensile yield stress and elongation of Mn- and Cr-modified compounds with temperature. The compressive yield stress data of the same compounds are also shown in the figure. Interestingly, the tensile yield stress exhibits a rather sharp drop above 350°C, while a relatively smooth decrease is observed in the compressive yield strength over a broad range of temperature. The fracture mode changed from transgranular cleavage at room temperature to a mixed cleavage and intergranular mode at 350°C, and to a completely intergranular mode at temperatures above 650°C. It should be noted that the temperature corresponding to the tensile ductility minimum is not in agreement with the ductility minimum at the temperature for the onset of intergranular fracture.

5.4 Fracture Mechanisms

Mechanisms for the poor room temperature ductility of the Al_3X (X = Ti, Nb, V) aluminides and the $L1_2$ variations have been discussed on the basis of the Pugh [64] criterion and the Rice-Thomson [65] criterion. The Pugh criterion states that a high value of the (K/G) ratio of the bulk modulus K to the shear modulus G is associated with
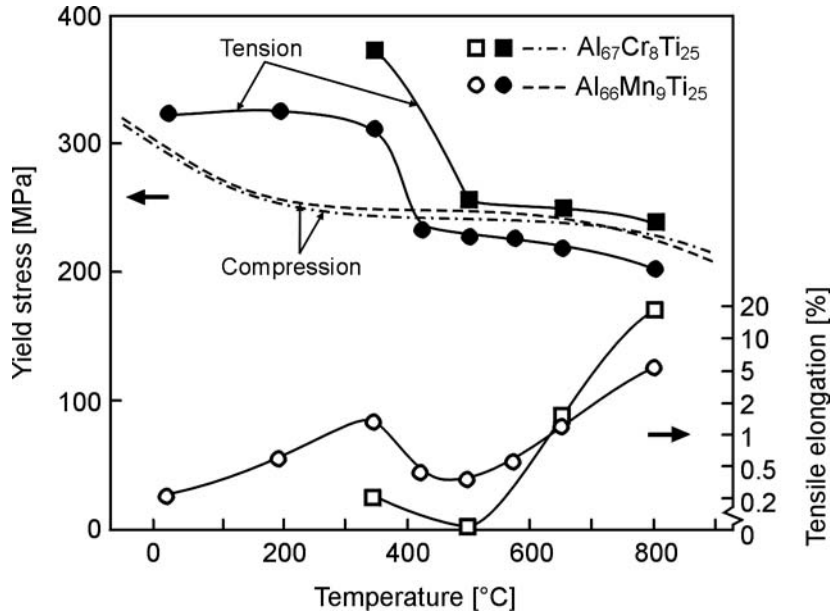
Figure 13 Variation in tensile and compression yield stress and tensile ductility as a function of the test temperature (Kumar and Brown [1992]).

ductility and a low value with brittleness. Modified $L1_2$ trialuminides such as $Al_{67}Ti_{25}Ni_8$ and $Al_{67}Ti_{25}Fe_8$ have in fact lower values of K/G than the ductile $L1_2$ compound Ni_3Al. However, in the review by Mikkola et al. [51] it was stated that the Cr- and Mn-modified trialuminides, which are somewhat more ductile than the other trialuminides, have the lowest (K/G) values of the Al_3Ti-based trialuminides. This result contradicts the predictions of the Pugh criterion.

The Rice-Thomson criterion is stating that brittle behavior occurs in those materials with values of $G/\boldsymbol{b}/\gamma > 7.5$–10, where /$\boldsymbol{b}$/ is the magnitude of the Burgers vector and γ is the surface energy of a crack. The $L1_2$ trialuminides have values of G which are comparable to that of the Ni_3Al compound. Although, there is no general agreement on whether $\langle 110 \rangle$ dislocations in trialuminides are dissociated into superlattice partial dislocations at low temperatures or not.

There is little knowledge on the γ-values of the surface energy in Al_3Ti-based trialuminides. However, recently first-principles calculations by Fu [14], have shown that the surface energy of Al_3Sc is significantly lower than that for Ni_3Al. Consistent with this, Fu and Yoo [15] have obtained a low value of the ideal cleavage strength in Al_3Sc. These results suggest that the values of γ in the Al_3Ti-based trialuminides would also be rather low. Considering the fact stated above: the brittleness of the $L1_2$ trialuminides may agree with the Rice–Thomson model. However, the high bonding energy giving rise to high elastic moduli which causes high cleavage strengths in the trialuminides. This might be the most probable reason for the general brittleness of the trialuminides.

6   OXIDATION BEHAVIOUR OF MONOLITHIC THE Al$_3$Ti AND Al$_3$Nb COMPUNDS AND OF (Al, X)$_3$Ti BASED ALLOYS

Al$_3$Ti is the only compound in the binary Ti-Al system which forms protective Al$_2$O$_3$ scales in air, thereby having much better oxidation resistance than Ti$_3$Al and TiAl. The rapid growth of TiO$_2$ in the scales unfavorably affects the oxidation behavior of trititanium and γ- titanium aluminides [66,67]. Smialek and Humphrey [68] have investigated the isothermal oxidation kinetics of Al$_3$Ti over a wide temperature range and related the formation rates to exclusive α-Al$_2$O$_3$ scale growth. As already pointed out Al$_3$Ti ingots prepared by conventional ingot metallurgy methods contain small amounts of aluminium since Al$_3$Ti does not melt congruently, but exhibits a peritectic reaction involving the aluminium phase near the melting temperature. Oxidation of the second phase, above 660°C, occurs rapidly and results in a high oxidation rate in shorter times and at lower temperatures. However, at temperatures above 1000°C, isothermal oxidation of Al$_3$Ti exhibits a parabolic oxidation behavior that is controlled by protective α-Al$_2$O$_3$ scale formation. The parabolic oxidation rate constant of about $k = 3.510^{-14} g^2/cm^4 s$ for Al$_3$Ti have been found to be comparable to α-Al$_2$O$_3$ rowth kinetics on NiAl. The intrinsic oxidation resistance of single-phase Al$_3$Ti is excellent. Therefore, attempts have been performed to use Al$_3$Ti as an oxidation resistant coating material on titanium base alloys α_2-Ti$_3$Al and γ-TiAl, respectively.

Coatings based on Al$_3$Ti have been produced with demonstrated improvements in the oxidation resistance of Ti and Ti-Al substrates using a pack aluminizing process. For protecting super-α_2 Ti-14Al-21Nb alloy in cyclic oxidation, an optimum coating thickness of 40–70 μm (this relates to a mass to area density of 8–14 mg cm^{-2}) was found by Smialek et al. [69]. They showed that thicker coatings are prone to penetrating cracks which widen, oxidize, and give rise to anomalously large weight gains, despite thin Al$_2$O$_3$ scales.

Cubic (AlX)$_3$Ti alloys with X = Cr and Mn additions do not show only reduced hardness and good compressive ductility but also some tensile ductility at room temperature. Wheter these modified cubic compounds have equivalent or better oxidation resistance at high temperatures than binary Al$_3$Ti is of first concern.

The oxidation resistance of these class of L1$_2$ variations of Al$_3$Ti has recently been evaluated by cyclic [70] and static [71,72] tests. The investigated materials are ternary compounds of compositions: Al$_{66-67}$M$_{8-9}$Ti$_{25}$ (M = Ag, Ni, Fe, Mn, Cr). The Ag- and Cr-modified ones show highest and lowest rates of oxidation, respectively [71]. The formation of TiO$_2$ has been detected in the Ag-modified compound. The Mn-containing compound shows better oxidation resistance than the Ag-containing one at lower temperatures, but the difference between the two compounds becomes smaller as the temperature is increased to 1200°C. The results of the oxidation resistance of the Cr- and Mn-modified compounds by cyclic oxidation testing at 1200°C in air are shown in Fig. 14. The Mn-modified compounds perform very poorly in this test, while the Cr-modified ones perform extremely well, and in fact display as good or better oxidation resistance than NiAl-based compounds which are considered to have excellent oxidation resistance. This is because the Cr-modified compound forms a protective scale of Al$_2$O$_3$, while the Mn-containing compounds form mixed Al$_2$O$_3$ + TiO$_2$ scales. The Al$_{67}$Cr$_8$Ti$_{25}$ compound with oxygen-active elements would likely reduce spallation to very low levels during the severe cyclic test. Such effects are similar to the

behaviour of NiAl, whose resistance to spalling is improved dramatically by the addition of small amounts of oxygen-active elements such as Zr and Y [73].

It is of interest that isothermal oxidation of binary Al_3Ti results exclusively in Al_2O_3 scale, while cyclic oxidation is degraded by the formation of TiO_2 along with Al_2O_3. The formation of TiO_2 during cyclic oxidation of binary Al_3Ti has been interpreted as follows: The spallation accelerates the rate of aluminium removal and may convert the surface of the Al_3Ti compound to a two-phase mixture of Al_3Ti + TiAl; then, nonprotective mixed TiO_2 + Al_2O_3 scales are formed on a TiAl layer and further oxidation proceeds at a higher rate. Such compositional resistance of a line compound may be avoided in ternary $L1_2$ compounds based on Al_3Ti, since they generally have a compositional range, e. g. an Al range of 62–67.5 at% for $(Al, Cr)_3Ti$ compounds. The Fe- and Ni-modified compounds have been reported to exhibit excellent static oxidation resistance in air at temperatures up to 1100 C, although their oxidation resistance has not yet been evaluated in severe cyclic oxidation testing.

Recently, Schneibel et al. [48] have examined compression creep properties, Young's modulus and thermal expansion of $Al_{67}Cr_8Ti_{25}$. The coefficient of thermal expansion of the Cr-modified $L1_2$ compound is about half that of aluminium and 30% higher than that of titanium alloys. The use of the Al_3Ti compound as protective coating on titanium alloys gives rise to tension strain in the coating layer during cooling to room temperature because of its larger thermal expansion coefficient than those of the Ti alloy substrates. However, cracking due to tension strain in the cubic Cr-modified compound is expected to be less severe than in binary Al_3Ti. This is due to its isotropic structure and the finite tensile ductility. In surveying the mechanical properties and oxidation resistance of the $L1_2$ variations of Al_3Ti, some of this class of materials, in particular the Cr-modified $L1_2$ compounds, exhibit improved oxidation resistance.
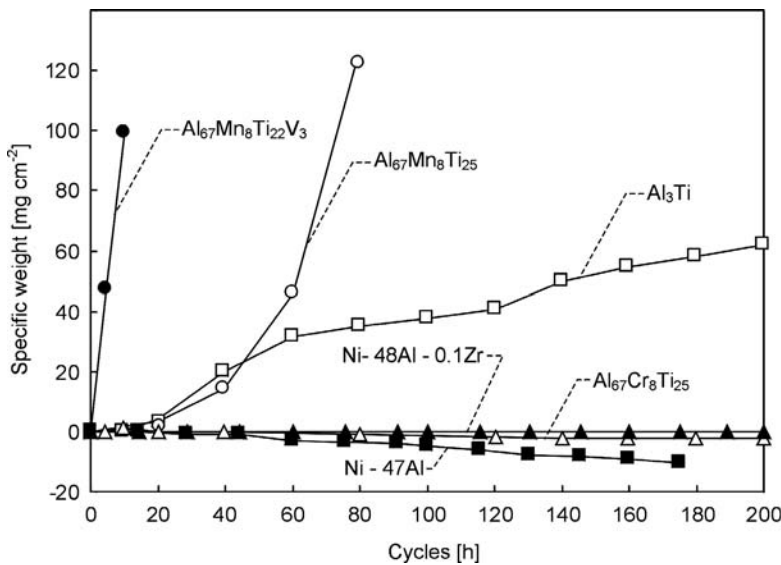

Figure 14 Specific weight of different Al_3Ti based compounds during cyclic oxidation at 1200°C (Parfitt et al. [1991]).

Pure Al_3Nb does not form a protective Al_2O_3 scale exclusively. However, by alloying to favor the selective oxidation of Al, a continuous Al_2O_3 scale can be grown on Al_3Nb. As a ternary addition to Al_3Nb, Cr is reported to be more effective than Ti and Si in growing a continuous Al_2O_3 scale. Oxygen-active elements have a benefical effect on the oxidation behavior of Al_3Nb. An Al_3Nb-based compound with a composition of Al-24Nb-7Cr-0.5Y-1.6Si (at%) has been reported to exhibit excellent cyclic oxidation resistance for 100 h at 1200°C, being nearly equivalent to NiAl + Zr [27,74]. The oxidation behavior and kinetics of Al_3Zr and its $L1_2$ modifications have not yet been investigated in greater detail.

7 POTENTIAL APPLICATIONS OF TRIALUMINIDE ALLOYS

In this chapter, a wide cross section of properties of trialuminides and its $L1_2$ variations have been reviewed. Early studies on Al_3Ti, Al_3V, Al_3Nb and Al_3Zr in the DO_{22} structure form stimulated much additional work on their $L1_2$ variations; however, interest in pure Al_3Ti quickly decayed since it is quite brittle at room temperature, and no significant improvement in ductility has been achieved in spite of various attempts to enhance slip and twinning by microalloying. In contrast, considerable interest on the $L1_2$ based Al_3Ti materials has been arisen because of its good compressional ductility and of some tensile ductility, especially in the Cr- and Mn-modified $L1_2$ compounds. In spite of alloying with transition metals such as Cr, Mn, Fe and Ni, the $L1_2$ variations of Al_3Ti still possess a desirable density of about 4 Mg m^{-3}. Their melting points are as high as ca. 1350°C and their creep properties are at least comparable with those of TiAl [48]. Oxidation resistance was a concern. However, the Cr-modified compound $Al_{67}Cr_8Ti_{25}$ has recently been found to exhibit excellent cyclic oxidation resistance at 1200°C. Its oxidation resistance is as good or better than that of NiAl alloys. From surveying the physical metallurgy and mechanical properties of the $L1_2$ variations of Al_3Ti, it is concluded that some of this class of materials in particular the Cr-modified compounds of superior oxidation resistance is considered as high-temperature coating and composite-matrix materials.

It should be noted that DiPietro, Kumar and Whittenberger et al. [45] have reported results on compressive studies of TiB_2 particulate composites with Al_3Ti-based $L1_2$ compounds as matrices in dependent on matrix composition, amount of TiB_2, processing technique, temperature and strain rate. Below 600°C, their strength-to-density ratio has been reported to be approximately four times those for superalloys. However, advantage is not maintained above 730 or 830°C. Creep responses (flow strength/density versus strain rate) of a monolithic Fe-modified $L1_2$ compound containing Nb have been found to be better than a nickel based superalloy. These results clearly imply that the Al_3Ti-based $L1_2$ trialuminides are worthy of more extensive studies as composite matrices showing good oxidation resistance.

REFERENCES

1. D. Shechtman and L. A. Jacobson, Metall. Trans, 1975, 6A, p. 1325.
2. M. Yamaguchi, Y. Umakoshi, and T. Yamane, Philos. Mag. A., 1987, 55, p. 301.
3. M. Yamaguchi, Y. Shirai, and Y. Umakoshi. In *Dispersion Strenghtened Aluminium Alloys*, Y.-W. Kim and W. M. Griffith, eds., 1988. TMS, Warrendale, p. 721.

4. M. Yamaguchi and Y. Umakoshi, Prog. Mater. Sci. 1990, 34, p. 1.
5. E. P. George, J. A. Horton, W. D. Porter, and J. H. Schneibel, I. of Mater. Res. 1990, 5, p. 1639.
6. D. E. Mikkola and J. P. Nic, Scripta Metall. Mater., 1990, 24, p. 1291.
7. D. G. Morris, J. Mater. Res., 1992, 7, p. 303.
8. D. G. Morris, Philos. Mag. A, 1992, 65, p. 389.
9. G. Sauthoff, Z. Metallkunde, 1990, 81, p. 855.
10. C.-P. Reip and G. Sauthoff, Intermetallics, 1993, 1, p. 159.
11. P. Villars and L. D. Calvert, *Pearson's Handbook of Crystallographic Data for Intermetallic Phases.* 1985, ASM Metals Park.
12. M. Nakamura and K. Kimura, J. Mater. Sci. 1991, 26, p. 2208.
13. C.-P. Reip, PhD Thesis, RWTH Aachen, 1991.
14. C. L. Fu, J. Mater. Res., 1990, 5, p. 971.
15. M. H. Yoo and C. L. Fu, ISIJ Int., 1991, 31, p. 1049.
16. M. Yamaguchi, D. P. Pope, V. Vitek, and Y. Umakoshi, Philos. Mag. A, 1981, 43, p. 1265.
17. M. Yamaguchi, V. Vitek, and D. P. Pope, Philos. Mag. A, 1981, 43, p. 1027.
18. G. Vanderschaeve and T. Sarrazin, Phys. Stat. Sol. (A), 1977, 43, p. 459.
19. G. Vanderschaeve, T. Sarrazin, and B. Escaig, Acta Metall., 1979, 27, p. 1251.
20. V. Vasudevan, R. Wheeler, and H. L. Fraser, in *High Temperature Ordered Intermetallic Alloys III*. C. T. Liu, A. I. Taub, N. S. Stoloff, and C. C. Koch, eds., 1989, MRS Proc., 133, p. 705.
21. D. G. Morris and R. Lerf, Philos. Mag.. A, 1991, 63, p. 1195.
22. C. Hug, J. Douin, and P. Veyssiere, in *High Temperature Ordered Intermetallic Alloys III*, C. T. Liu, A. I. Taub, N. S. Stoloff, and C. C. Koch, eds., 1989, MRS Proc., 133, p. 125.
23. M. H. Yoo, J. Mater. Res., 1989, 4, p. 50.
24. M. Khantha, V. Vitek, and D. P. Pope, Mater. Sci. Eng. A, 1992, 152, p. 89.
25. S. Srinivasan, S. R. Chen, and R. B. Schwarz, Mater. Sci. Eng. A, 1992, 153, p. 691.
26. M. Steinhorst, H. J. Grabke, Mater Sci A 1989, 120, pp. 55–59.
27. H. J. Grabke, et al. Mater Sci. Technol. 1992, 8, p. 339.
28. A. P. Paxton, D. G. Pettifor, Scr. Metall. et Mater. 1992, 26, pp. 529–533.
29. M. H. Yoo, C. L. Fu, and H. K. Lee, in *High Temperature Ordered Intermetallic Alloys IV* L. A. Johnson, D. P. Pope, and J. O. Stiegler, eds., 1991, MRS Proc., 213, p. 545.
30. K. S. Kumar, Intl Mater. Rev., 1990, 35, p. 293.
31. K. S. Kumar and S. A. Brown, Philos. Mag. A, 1992, 65, p. 91.
32. K. S. Kumar and S. A. Brown, Acta Metall. Mater., 1992, 26, p. 1923.
33. D. G. Pettifor, Mater. Sci. Tech., 1988, 4, p. 675.
34. C. J. Sparks, W. D. Porter, J. H. Schneibel, W. C. Oliver, and C. G. Golec, in *Alloy Phase Stability and Design*, G. M. Stocks, D. P. Pope, and A. F. Giamei, eds., 1991, MRS Proc., 186, p. 175.
35. T. Hong and A. J. Freeman, J. Mater. Res., 1991, 6, p. 330.
36. D. M. Nicholson, G. M. Stocks, W. M. Temmermann, P. Sterne, and D. G. Pettifor, in *High Temperature Ordered Intermetallic Alloys III*, C. T. Liu, A. I. Taub, N. S. Stoloff, and C. C. Koch, eds., 1989, MRS Proc., 133, p. 17.
37. M. E. Eberhart, K. S. Kumar, and J. M. MacLaren, Philos. Mag. B, 1990, 61, p. 943.
38. J. H. Schneibel and W. D. Porter, in *High Temperature Ordered Intermetallic Alloys III*, C. T. Liu, A. I. Taub, N. S. Stoloff, and C.C. Koch, eds., 1989, MRS Proc., 133, p. 335.
39. I. S. Virk and R. A. Varin, Metall. Trans. A, 1992, 23, p. 617.
40. I. S. Virk and R. A. Varin, Metall. Trans. A, 1992, 23, p. 1243.
41. F. R. Frasier, D. R. Baker and M. J. Kaufman, in *High Temperatur Niobium Alloys*, J. J. Stephens and I. Ahmad, eds., 1991, TMS, Warrendale, p. 121.
42. J.-H. Xu and A. J. Freeman, J. Mater. Res., 1991, 6, p. 1188.

43. S. Mazdiyasni, D. B. Mirack, D. M. Dimiduk, M. G. Mandratta and P. R. Subramanian, (Jahresz.?) Scripta Metall. 23, p. 327.
44. Z. L. Wu, D. P. Pope and V. Vitek, in *High Temperature Ordered Intermetallic Alloys IV*, L. A. Johnson, D. P. Pope, and J. O. Stiegler, eds., 1991, MRS Proc., 213, p. 487.
45. M. S. DiPietro, K. S. Kumar and J. D. Whittenberger, J. Mater. Res., 1991, 6, p. 530.
46. J. P. Nic, J. L. Klanski, and D. E. Mikkola, Mater. Sci. Eng. A, 1992, 52, p. 132.
47. X. F. Chen, X. H. Wu, S. P. Chen, and G. X. Hu, Scripta Metall. Mater. 1992, 26, p. 1775.
48. J. H. Schneibel, J. A. Horton, and W. D. Porter, Mater. Sci. Eng. A, 1992, 152, p. 126.
49. S. Zhang, J. P. Nic, W. W. Milligan and D. E. Mikkola, Scripta Metall. Mater., 1990, 24, p. 1441.
50. K. S. Kumar and J. R. Pickens, in *Dispersion Strengthened Aluminium Alloys*, Y.-W. Kim, and W. M. Griffith, eds., 1988, TMS, Warrendale, p. 763.
51. D. E. Mikkola, J. P. Nic, S. Zhang and W. W. Milligan, ISIJ Intl, 1991, 31, p. 1076.
52. M. B. Winnicka and R. A. Varin, Scripta Metall. Mater., 1991, 25, p. 1289.
53. M. Yamaguchi, V. Paidar, D. P. Pope, and V. Vitek, Philos. Mag. A. 1982, 45, p. 867.
54. V. Paidar, M. Yamaguchi, D. P. Pope, and V. Vitek, Philos. Mag. A, 1982, 45, p. 883.
55. G. Tichy, V. Vitek, and D. P. Pope, Philos. Mag, A, 1986, 53, p. 467.
56. G. X. Hu, S. P. Chen, X. H. Wu, and X. F. Chen, J. Mater. Res., 1991, 6, p. 957.
57. H. Inui, D. E. Luzzi, W. D. Porter, D. P. Pope, V. Vitek and M. Yamaguchi, Philos. Mag. A, 1992, 65, p. 245.
58. Z. L. Wu, D. P. Pope, and V. Vitek, Scripta Metall. Mater., 24, 1990, p. 2191.
59. L. Petez, G. Lapasset, and L. P. Cubin, Scripta Metall. Mater., 1992, 26, p. 841.
60. H. Mabuchi, K. Hirukawa, K. Katayama, H. Tsuda, and Y. Nakayama, Scripta Metall. Mater., 1990, 24, p. 1553.
61. R. Lerf and D. G. Morris, Acta Metall. Mater., 1991, 39, p. 2419.
62. C. D. Turner, W. O. Powers and J. A. Wert, Acta Metall., 1989, 37, p. 2635.
63. K. S. Kumar, S. A. Brown, and J. D. Whittenberger, in *High Temperature Ordered Intermetallic Alloys IV*, L. A. Johnson, D. P. Pope, and J. O. Stiegler, eds., 1991, MRS Proc., 213, p. 481.
64. S. F. Pugh, Philos. Mag., 1954, 45, p. 823.
65. J. R. Rice and R. Thomson, Philos. Mag., 1986, 29, p. 73.
66. J. Subrahmanyam and J. Annapurna, Oxid. Met., 1986, 26, p. 275.
67. Y. Umakoshi, M. Yamaguchi, T. Sakagami, and T. Yamane, J. Mater . Sci., 1989, 24, p. 1599.
68. J. L. Smialek and D. L. Humphrey, Scripta Metall. Mater., 1992, 26, p. 1763.
69. J. L. Smialek, M. A. Gedwill, and P. K. Brindley, Scripta Metall. Mater., 1990, 24, p. 1291.
70. L. J. Parfitt, J. L. Smialek, J. P. Nic, and D. E. Mikkola, Scripta Metall. Mater., 1990, 25, p. 727.
71. K. Hirukawa, H. Mabuchi, and Y. Nakayama, Scripta Metall. Mater., 1991, 25, p. 1211.
72. S. Chen, W. Zhang, Y. Zhang, G. Hu, T. Li, J. Shen and L. Zhou, Scripta Metall. Mater., 1991, 27, p. 455.
73. J. L. Smialek and R. Browning, in *High Temperature Materials Chemistry III*, Z. A. Munir and D. Cubicciotti, eds., 1986, Electrochemical Society, Pennington, NJ, p. 258.
74. M. G. Hebsur, J. R. Stephens, J. L. Smialek, C. A. Barrett and D. S. Fox, in *Oxidation of High Temperature Intermetallics*, T. Grobstein and J. Doychak, eds., 1989, TMS, Warrendale, PA, p. 171.
75. G. Sauthoff, in *Mat. Science and Technology*, (Chap. 11, Intermetallics) Vol. 8: Structure and Properties, H. Matucha, ed., R. W. Chan, P. Hacker and Dr. Kramer, Gen. Eds., 1996, VCH Verlagsgesellschaft mbH, Weinheim.

17

Aluminum-Based Metal Matrix Composites

GEORG FROMMEYER and SVEN KNIPPSCHEER

Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf, Germany

1 INTRODUCTION

1.1 Classification

Metallic matrix composites are combinations of two or more different metals, intermetallic compounds, or second phases having particular microstructural arrangements within a metallic matrix. They are produced by mixing the components in such a way that the morphologies of the constituents are controlled so as to achieve optimum properties. The manifold possibilities of combining metallic and nonmetallic materials allow the creation of a new class of metallic alloys with superior or unusual properties.

The essence of composite materials and their technology is the ability to modify materials with hard particles, strong stiff fibers or layers in the proper distribution, orientation, and volume fraction. Ceramics and polymers alone or in combination with metals cover the complete spectrum of properties for all composite materials. However, in the interest of confining the topic to the realm of aluminum metallurgy, specifically aluminum-based composites, these composites may be classified in many ways, depending on the concept of interest.

One useful classification system is based on microstructural aspects, differentiating between composite materials according to the *morphology* of the constituent phases. Using this system, composite materials fall into the following categories:

- Dispersion- and particle-strengthened composites.
- Fiber- reinforced composites.
- Layer- composite materials.

Dispersion and *particle strengthened composites** contain the second phase as fine particles randomly distributed in the matrix [Fig. 1(a)]. In fiber-composites, the second phase is mainly one-dimensional, and may be continuously or discontinuously aligned in the matrix [Fig. 1(b)]. In *layered* or *laminated composites*, the second components and phases are two-dimensional. They can also be continuously or discontinuously arranged [Fig. 1(c)]. These different types of microstructures can be achieved on nano size, micron size, and macro size levels.

Another classification system is based on the *size and distribution* of the second phases in the composite. Using this system the following categories are defined:

- Nano-composite materials (particle size or fiber diameter: $d \approx 10$ to 100 nm), e.g. dispersion-strengthened alloys and fiber composites with extremely thin fibers.
- Micro-composite materials (particle size, fiber or layer thickness: $d \approx 1$ to 10 μm), e.g. particle-, fiber- and laminate-reinforced composite, and directionally solidified eutectics.
- Macro-size materials (macroscopic dimensions of the components: $d \gg 100 \, \mu$m), e.g. steel wire-reinforced aluminum alloys or laminate.

1.2 Preparation Techniques

In the first section of this chapter the relation between structure and properties of the dispersion, particle or fiber reinforced matrix phases, and interfaces will be described and discussed. Macro-composite materials belong to the category of technological structures and structural engineering products.

The predominant production methods of composites will be presented shortly. For a more detailed discussion see Broutman and Krock [12].

The widely used approaches for preparing metallic composites are:

- Powder-metallurgical techniques.
- Directional solidification.
- Roll bonding/plating.
- Coextrusion combined with wire-drawing (in-situ formation of fibers).

Powder-metallurgical techniques and their variations use metals and/or hard second phases in powder form as starting materials. The diameter of the powder particles ranges from 0.05 μm to about 1 μm of the reinforcing dispersion or particles and to 50 or 100 μm for the matrix powders. For alloy preparation, the components are mixed or mechanically alloyed and compacted by pressing. This can be carried out at room or higher temperature by sintering. Mass transport by diffusion decreases the volume fraction of cavities, thus simultaneously increasing the density of the compacted composites. Hot isostatic pressing (HIP) and extrusion allow the density of the material to increase to its theoretical limit. Frommeyer and Wassermann [16] have shown that extensive plastic deformation by rolling or wire-drawing enables very thin layers or fibers of about 10–50 nm in size to be produced in-situ in the matrix. A promising technique especially for producing laminate

---

*The term composite often refers to fiber composites only in the literature. Here, we use it in the more general sense.
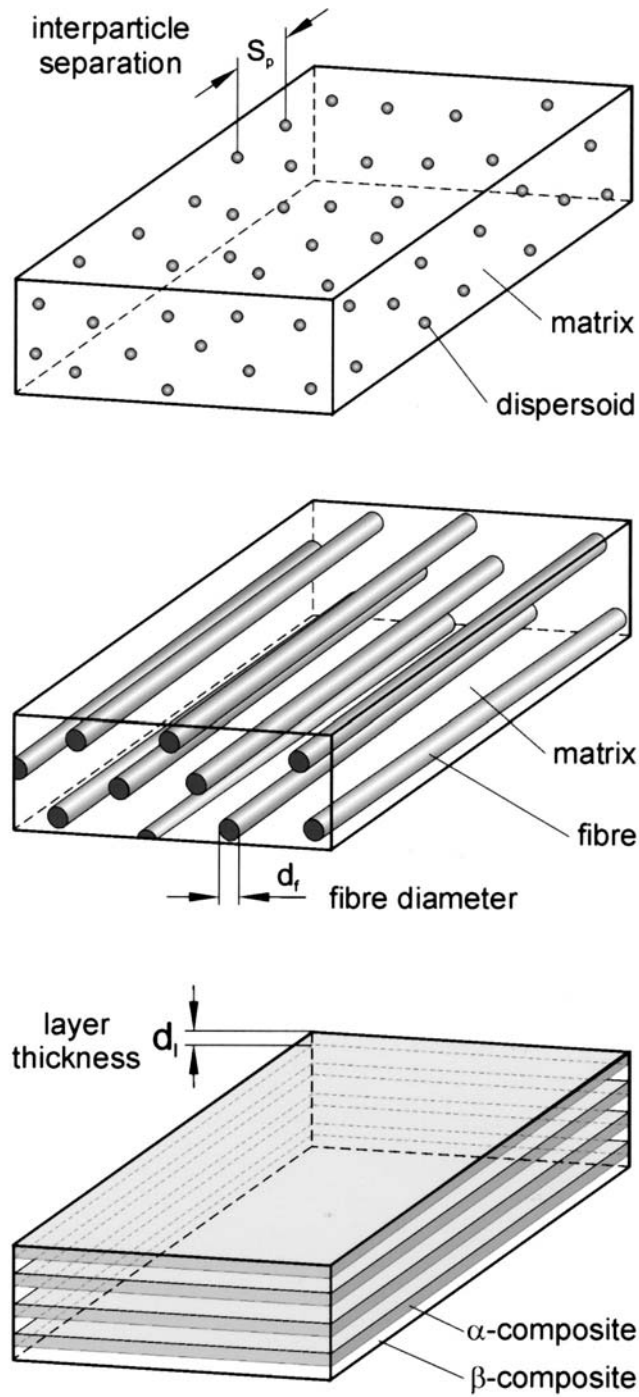
Figure 1 Schematic representation of different microstructures of composite materials.

composites with sandwich structure is superplastic bonding. Aluminum matrix composites, classified in dispersion- and particle-strengthened, as well as in fiber- and layer-reinforced ones can be prepared by powder-metallurgical techniques. In addition, such modified techniques as internal oxidation (in situ oxide-particle formation in a metallic matrix by oxidation), and mechanical alloying enable the preparation of elevated temperature creep and oxidation resistant superalloys. Readers who are interested in this topic are referred to the treatment by Hausner [58] and Lenel [49]. Other methods, such as chemical and physical vapor deposition, and explosive plating may be used, especially for preparing high melting point b.c.c. and h.c.p. metal composites with layer structure.

An elegant method for producing metallic composites with fiber and layer structures from the melt is *directional solidification* of binary or ternary eutectic alloys. This process can be used to take advantage of the simultaneous formation of the separate phases which form an eutectic upon cooling. The same principles are valid for directionally transformed eutectoids. For *directional solidification*, it is necessary to solidify the eutectic melt at a controlled rate, with the temperature gradient perpendicular to the liquid-solid interface. This can be accomplished by Bridgman or Czochralski technique applying a specific solidification rate and high temperature gradient. The solidification rate of the phases is then controlled by the velocity of the crucible.

Figure 2 represents as an example the fibrous microstructure of the unidirectionally solidified eutectic Al-Al$_3$Ni composite material in longitudinal section (a) and transverse section (b) shown by Hertzberg et al. [27]. The rod like fibers are of the order of 1 μm in diameter for most binary and ternary eutectic alloys. The morphology of the fibers is characteristic of many alloy systems, e.g. NiAl-Cr, Al-Pd, Al-Si, Al-Al$_3$Ni, etc. Many factors influence the fibrous shape and morphology. The most important parameters are the solidification rate, $v_{cr.}$, and the volume ratio of the second phase in the eutectic V_{eut}. If this ratio is very small, e.g. $V_{eut} < 0.1$, then the eutectic alloy will show anomalous directional-solidification characteristics with irregular shapes of the fibers (e.g. Al-Si eutectic).
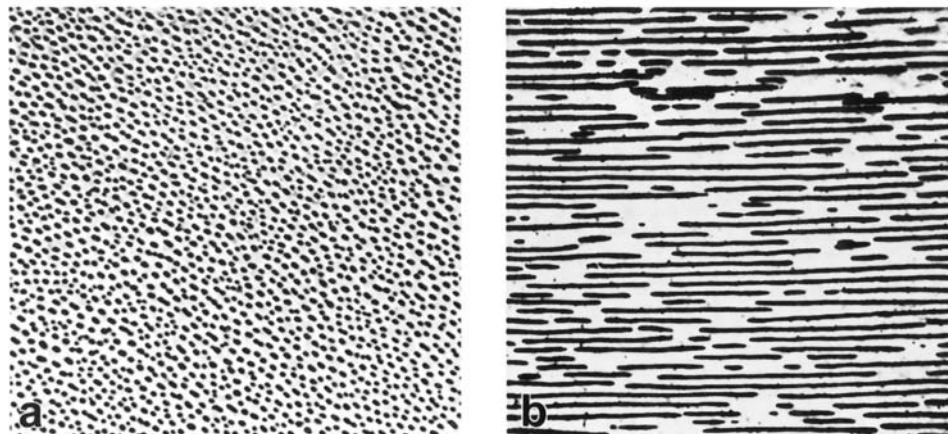
Figure 2   Fibrous microstructure of an unidirectionally solidified Al-Al$_3$Ni eutectic alloy; (a) transverse section; and (b) longitudinal section. 500×. (From Ref. 27.)

Figure 3 Lamellar microstructure of an unidirectionally solidified Al-CuAl$_2$ eutectic (transverse section). $1000\times$. (From Ref. 27.)

Another common morphology of directionally solidified composites is the lamellar structure [Fig. 3]. Here the two phases form as alternating platelets. This microstructure is characteristic for eutectics with approximately equal volume fractions, e.g. Al-Al$_2$Cu, Al-PdAl$_3$, Al-Ag(ζ), Al-Zn etc. It can also occur in systems with unequal volume fractions if the energy associated with the interface is particularly low for a particular crystallographic orientation. The controlled solidification of eutectic alloys is discussed more thoroughly, see e.g. Kurz and Sahm [46].

2 THEORETICAL APPROACH OF STRENGTHENING MECHANISMS

2.1 Dispersion-Hardening of Composite Materials

In dispersion-strengthened composite materials (particle size $d \ll 1$ μm) and also in particle-reinforced composites (particle diameter $\cong 1$ μm) the matrix is the major load-bearing constituent. The fine dispersion of a second phase is present primarily to impede the motion of dislocations. The metallic matrix will be strengthened or reinforced in proportion to the effectiveness of the dispersion as a barrier to the motion of dislocations.

The structure parameters determining the effectiveness of a dispersion are the mean free path between particles in the dispersion and the interparticle separation. The mean free path between particles, $\bar{L}_p$, is defined as the mean separation of particles along a randomly drawn straight line across a random section of a specimen. The interparticle separation of randomly distributed particles, S_p, may be evaluated as the radius of the smallest circle about one particle in a random plane of finding a second particle, for which the probability is unity, see e.g. Ashby [4].

These parameters are related to the particle diameter, d, and volume fraction of particles, V_p, by the following equations:

$$\bar{L}_p = \frac{2}{3}\,d\,\frac{1 - V_p}{V_p} \tag{1}$$

$$S_p = \frac{d}{\sqrt{v_p}}\,\sqrt{\frac{2}{3}} \tag{2}$$

For a matrix dislocation to pass through a dispersion of fine particles, the applied stress must be sufficient to bend the dislocation into a semicircular loop. The smallest radius R of curvature to which a dislocation can be bent under the influence of an internal stress field 'τ_i' is:

$$R = G_m|\vec{b}|/2\tau_i \tag{3}$$

where G_m is the shear modulus and $\vec{b}$ is the Burgers vector of the matrix. For a given interparticle spacing S_p, the stress required to bow dislocations around this dispersion is:

$$\tau_i = G_m|\vec{b}|/S_p \tag{4}$$

If the applied stress is large enough to create a dislocation loop of radius $R = S_p/2$, then the loop can be expanded without a further stress increment. The limits for S_p can be predicted by using the matrix yield strength, $\tau_{th} \approx G_m/1000$, as the minimum stress to expand a dislocation loop between the particles. The maximum stress is equal to the theoretical matrix shear stress, which for f.c.c. metals is $\tau_{th} \approx G_m/30$ and for b.c.c. metals is $\tau_m \approx G_m/10$. Substituting these values of the stress into Eq. (3), and using $|\vec{b}| \approx 3\,\text{Å}$ for the Burgers vector, we see that the interparticle separation for effective dispersion strengthening should be $0.01 \leq S_p \leq 0.3\,\mu\text{m}$. Many properties of the metallic matrix, such as ductility, electrical and thermal conductivity, and impact strength, will be preserved. Exceptions are strength improvement and creep resistance. The former restrictions require that the volume fraction of the dispersoid should be kept small.

Figure 4 shows clearly that the particle diameter must be less than 0.1 μm to keep the mean free matrix path 'L_p' in the desired range between $0.01 < \bar{L}_p < 0.3\,\mu\text{m}$, and to maintain an optimal volume fraction of $V_P \approx 15\%$ of the second phase. The described stress-induced dislocation bowing creates a loop of slipped area around the dispersed particles. This circular slipped area effectively reduces the interparticle spacing. The additional increase in stress is required to move dislocations through the dispersion and produces work-hardening.

The main advantage of dispersion-strengthened composites is not their ability to improve the room-temperature yield strength or work-hardening rate, but more the ability to maintain this yield-strength increase and attendant creep resistance increase over a wide temperature range, up to $T \approx 0.8T_m$ of the melting point of the metallic matrix, see e.g. [2]. The effectiveness of the dispersion and dispersoids is determined by their insensitivity to high temperatures. This distinguishes dispersion-hardened composites from precipitation-hardened alloys, which soften with increasing temperature. The dispersoids in aluminum-based composites may be
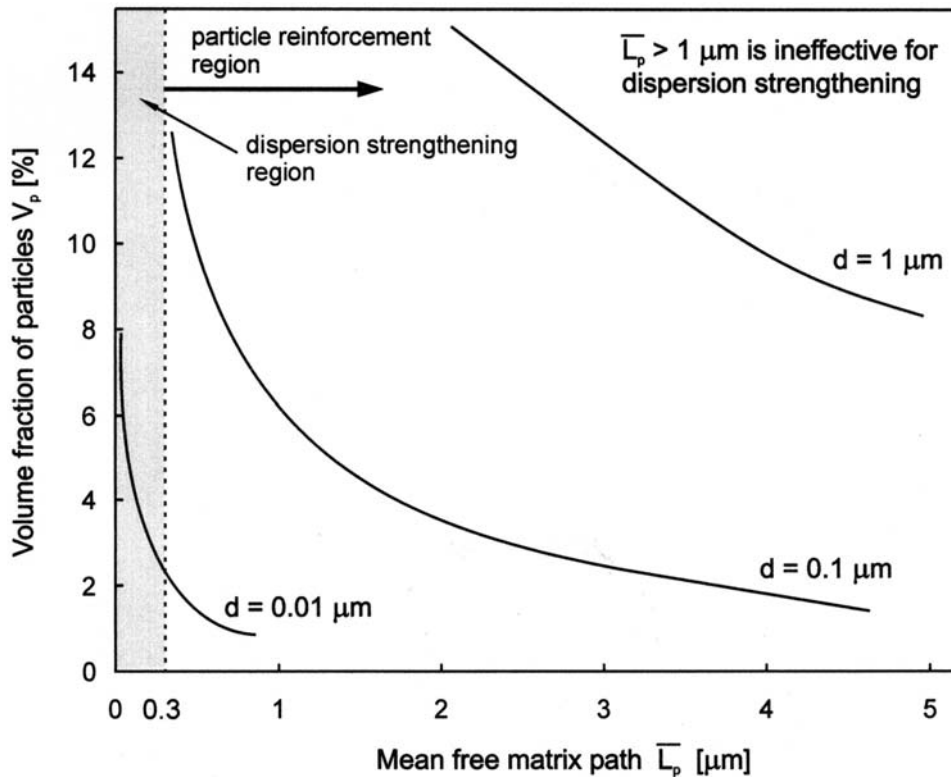
Figure 4 Relationship between volume fraction and mean free matrix path for dispersion-strengthened composites of different particle size. (From Ref. 12.)

oxides (Al_2O_3), carbides (SiC, Al_4C_3), silicides (VSi_2), borides (TiB_2) or refractory metal particles (Cr, W, etc.), which are insoluble or incoherent with the matrix.

2.2 Development of Commercial Dispersion Strengthened Aluminum Composites

The first developed and commercialized dispersion-strengthened aluminum alloy by its own refractory oxide Al_2O_3 was SAP—Sintered Aluminum Powder—introduced by Alusuisse (Switzerland). An important indication of the great interest in this material has been the extent of the technical literature describing these SAP alloys containing different aluminia contents of about 6–14 wt%. More than 60 articles have been published, including reviews by Goetzel [20], Bloch [9], and Hansen [26], since the first experimental work has been performed by Stern [79].

Some physical and mechanical properties of selected SAP materials are presented in Tables 1, 2 and 3.

The production of SAP alloys requires relatively complex and extensive milling operations which led to some problems encountered, e.g. non reproducibility of the microstructure and properties, and high cost; thus SAP are no longer produced commercially on large industrial scale. Jangg et al. [35] started working with a modified milling process of graphite with aluminum powder in a high-energy ball mill attritor.

Table 1   Physical Properties of Sintered Aluminum Alloys

Property	USA Swiss	XAP 002 SAP 930	XAP 003 SAP 895	XAP 004 SAP 865
Density (Mg/m^3)		2.74	2.75	2.77
Electrical conductivity at 25°C $(\times 10^{-4}\ \Omega^{-1}\ cm^{-1})$		44	39	34
Average thermal expansion coefficient, 20–400°C $(10^{-6}\ K^{-1})$		23.9	23.0	22.1

Source: Ref. 79.

Table 2   Mechanical Properties of Sintered Aluminum Alloys

Material	Al_2O_3 content (wt%)	Test temperature (°C)	Yield strength, 0.2% offset (MPa)	Tensile strength (MPa)	Elongation (%)
SAP 930	6–8	20	115–144	221–241	18–16
		200	115–126	144–159	17–25
		400	67–82	67–86	6–11
SAP 895	9–11	20	174–220	281–356	8–12
		200	144–174	193–217	4–8
		400	86–106	96–115	3–7
SAP 865	12–14	20	202–232	328–356	6–9
		200	173–193	212–250	4–7
		400	106–114	114–137	2.5–6

Source: Refs. 25, 51, 72.

Table 3   Rupture Properties of Sintered Aluminum Alloys

Material	Al_2O_3 content (wt%)	Test temperature (°C)	Stress to failure at given time (MPa) 100 hrs	Stress to failure at given time (MPa) 1000 hrs
AMP-M257 (SAP 930)	7	≈ 200	120	110
		400	55	54
		≈ 480	49	43
SAP 895	10	400	74	71
		≈ 480	56	54
SAP 865	13	200	165	152
		400	88	85
		≈ 480	59	55

Source: Refs 24, 26, 71.

The final product consisted of Al_4C_3, which was synthesized by reaction milling, and was finely dispersed throughout the aluminum matrix. This pioneer work has been carried forward to technical production, where the alloy powder is consolidated

by cold isostatic compaction and hot extrusion. The commercialized alloys, given the trade mark DISPAL, are established since 1984 by Eckhart-Werke, Germany.

The properties of dispersion strengthened Al-C materials depend primarily on the volume percent of the effective dispersoids. Focus on the reaction-milled products has been on the unalloyed aluminum matrix with properties for these DISPAL alloys reported in the literature. Most commonly, data for DISPAL 2, 3 and 4 have been presented; see Table 4. The tensile properties show the strong influence of carbon content, i.e. dispersoid content, as well as the excellent high-temperature strengths of these alloys. An interesting characteristic is the approximately linear decrease in strength with temperature. This linear decrease is also found for ductility, in contrast to the increase typically found for conventional aluminum alloys.

The structural features of the Al-Al_4C_3 composites are simply described as dispersoid in an aluminum matrix with the presence of some dislocation substructure. The dispersoids present are Al_4C_3 and in smaller amounts Al_2O_3. Kubasta [45] described the Al_4C_3 as having dimensions of 10–20 nm with lengths typically 100 nm. These are distributed in a matrix with a grain size of about 0.3–0.6 μm. Some dispersoids are concentrated at grain boundaries with only limited incidence at the grain interiors. The Al_2O_3 present exhibits characteristics similar to those found in the SAP alloys. Also similar to the SAP alloys is the presence of iron-rich, MoS_2, and silicate inclusions. These are reported by Jangg et al. [36] to result from milling and extrusion procedures.

The mechanical milling (MA) process that was developed and introduced by Benjamin [5] in the 1960s at INCO Research and Development Center in New York for the production of advanced dispersion-strengthened superalloys (ODS). Like the SAP and reaction milling process, mechanical alloying involves a milling step followed by powder conditioning, consolidation, and finally thermomechanical working. This high-energy dry-milling technique produces composite metal powders of aluminum besides other metals, such as Ni, Fe, etc. with nano-scale homogeneity [5,19].

The properties of MA aluminum materials will be discussed in two groups: unalloyed aluminum and aluminum alloys. In the case of MA aluminum, the dispersoids have been determined by Singer et al. [77] to be γ-Al_2O_3 and Al_4C_3.

Table 4 Tensile Properties of DISPAL Alloys

Alloy	Carbon level (wt%)	Test temperature (°C)	Yield strength, 0.2% offset (MPa)	Tensile strength (MPa)	Elongation (%)
DISPAL 2	2	25	340	310–385	9–11
		200	—	210–245	8.5
		400	—	95–125	4–6
DISPAL 3	3	25	350–410	380–420	2–6
		200	—	260–275	4
		400	110–140	130–160	1–3
DISPAL 4	4	25	—	465	—
		200	—	305	—
		400	—	175	—

Source: Refs 3, 34, 36, 61.

Table 5   Tensile Properties of MA Aluminum

Dispersoid content (vol%)[a]	Test temperature (°C)	Yield strength, 0.2% offset (MPa)	Tensile strength (MPa)	Elongation (%)	Reduction of Area (RA) (%)
10	25	424	454	3.5	19.5
	316	147	179	2.0	4.5
7.5	25	377	410	11.0	34.0
	316	154	156	3.5	12.5
4,5	25	325	397	11.5	46.5
	316	143	156	7.5	30.5
3.5	25	283	323	14.5	44.0

[a]Dispersoids composed of Al_2O_3 + Al_4C_3.
Source: Refs 5, 11.

Benjamin and Bomford [5] stated that the development of MA aluminum was aimed at producing materials with good electrical conductivity combined with strength. Table 5 shows room- and elevated-temperature tensile properties of MA aluminum. With increasing dispersoid contents the strength is also increasing. The desire for achieving good electrical conductivity was the interest of Benjamin and Bomford [5]. The results are presented by them in Table 6, where the higher strength is impaired with a given level of conductivity than other dispersion-strengthened aluminum or conventional aluminum alloys.

Many MA aluminum alloys have been developed mainly for higher temperature applications. The following three MA alloys commercially produced have been chosen as representative materials. The INCO MAP alloy Al 9052 is on the binary aluminum–magnesium system revealing high strength combined with excellent corrosion resistance. Another class is the ternary aluminum–copper–magnesium alloy Al 9021 that is heat treatable and is characterized by high strength and ductility combined with good intermediate-temperature stability [75]. The aluminum–magnesium–lithium based alloy Al 905XL exhibits lower density combined with high stiffness [74]. Table 7 summarizes some typical properties at room temperature.

Table 6   Electrical Conductivity of DS Aluminum Alloy Wire

Material	Yield strength 0.2% offset (MPa)	Conductivity (% IACS)[a]
MA Al	332–350	52–53
SAP 930	241	53.5
SAP 895	324	49.8
SAP 865	371	39.5
EC Al (H-19 temper)	186	62.5
5005 Al alloy	186	53.5

[a] % IACS = $(\rho_{Al}/\rho_{Cu}) \times 100$, where ρ_{Cu} = 1.7241 $\mu\Omega$ cm. = International Annealed Copper Standard and ρ_{Al} = resistivity of Al alloy.
Source: Refs 5, 11.

Table 7 Typical Properties of MA Aluminum alloys

Tensile properties at 20°C	AL-9052	Al-905XL[a]	AL-9021
Density (Mg/m^3)	2.66	2.58	2.80
Elastic modulus (GPa)	71	78	74
Yield strength			
(0.2% offset) (MPa)	380	448	469
Tensile strength (MPa)	450	517	538
Elongation (%)	13	9	13
Fracture toughness			
(MPa m$^{1/2}$)	44	29.7	39.6

[a] Solution annealed.
Alloy IN-9052, Strong corrosion-resistant Al-Mg alloy [1985], Military Standardization Handbook [1976], and Aluminum Standards and Data [1984].
Source: Refs 74, 75.

The microstructure of mechanically alloyed powder shows extensive deformation, which resulted in a layered structure. Electron microscopy revealed the powder to be composed of nano crystals about 20–50 nm in thickness [38]. Apparently, dynamic recrystallization occurred during the milling process.

The deformation inherent in the mechanical alloying process strongly influences subsequent grain structures in MA aluminum materials. Thus, the consolidated bar grain size of all three alloys are similar. Singer et al. [77] reported grain size of MA-Al as 0.2–0.5 nm. Similar grain sizes have been reported for AL-9052, AL-9021, and AL-905XL by Kang [38], Gilman [19], and Narayanan et al. [60]. Some elongation of the grain shape has been found in worked materials. For example, the same author reports grain aspect ratios ranging between 1.3 and 3 for extruded rod.

The alloying elements in the aluminum alloys can also introduce intermetallic precipitates, such as Al_3Li, into the microstructures. Contaminations, particularly iron, are also present. Fine iron-containing particles have been shown by Narayanan et al. [60] to be involved in some fracture initiation of these alloys.

3. PARTICLE-REINFORCED COMPOSITES

The important role of the matrix phase in particle-reinforced composite materials is to transmit load to the reinforcing particles. These types of composites continue where dispersion-strengthened composites leave off. In this case, the metallic matrix and the particles share the applied load. The volume fraction of the second phase exceeds $V_p = 0.25$, and the particle diameter as well as the mean free path in the matrix is of the order of 1 m. Therefore, particle-reinforced composites are intermediate between dispersion-strengthened and fiber-reinforced materials.

Hardening and strengthening of particle-reinforced composites occur initially when the dispersed particles restrict matrix deformation by mechanical restraint. The amount of the restraint effect is complex, and depends on structural factors. In general, it is a function of the ratio of the particle separation distance to the particle diameter, S_p/d, and the ratio of the elastic properties of the matrix and

the second phase. The elastic modulus of particle-reinforced composites can be described using the *rule of mixtures*

$$E_c = V_m E_m + V_p E_p \tag{5}$$

where V_m and V_p are the volume fractions of the matrix and the particles, and $E_{c/m/p}$ are the elastic moduli of the composite and the components, respectively. Krock [43] showed that if the constraint is severe enough, this bound may be exceeded, the isostrain conditions will not be met and deviations from eq. (5) occur. Some typical data are represented in Fig. 5 The *elastic properties* of the aluminum Mg_2Si particulate composite do not follow the rule of mixtures under the isostrain condition so that remarkable deviations from the predicted upper bound modulus given by Eq. (5) will occur.
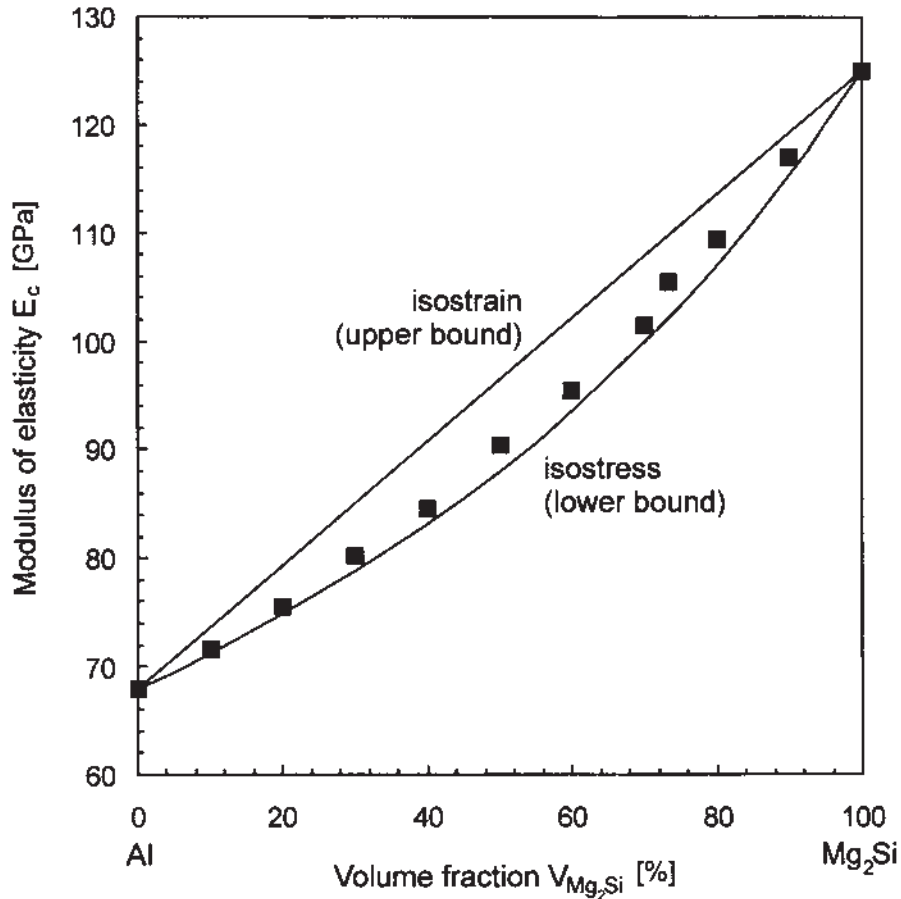

Figure 5 Elastic modulus of aluminum–magnesium silicide particulate composite versus volume fraction of dispersoid. The experimental values fall between of the rule-of-mixture upper (isostrain) and lower (isostress) bounds. (From Ref. 17.)

In this case, the elastic moduli of particle composites should follow more the isostress criterion:

$$\sigma_c = \sigma_m = \sigma_p \tag{6}$$

The elastic properties can then be describes as

$$\frac{1}{E_c} = \frac{V_m}{E_m} + \frac{V_p}{E_p} \tag{7}$$

All particle-reinforced metallic composites investigated exhibit also positive deviation from Eq. (7) and signify matrix constraint. This has been observed in several material systems such as: Al-BN, Al-Glass, Al-Cr, etc. The positive deviations of the elastic modulus of Al-Mg_2Si particle composites from the lower bound isostress curve described by Krock [42] is due to a certain amount of hydrostatic stress generated in particle-reinforced metallic composites. The amount of deflection by the degree of hydrostatic constraint depends on the mechanical and microstructural parameters.

The deformation behavior beyond the elastic range of particle-reinforced composites may be classified into categories according to whether the particles themselves undergo *plastic deformation* before fracture occurs. Under an applied stress, the softer matrix phase is restrained from deformation by the hard particles. A hydrostatic stress field, identical to that in the elastic range, is built up. With continued load increase, the magnitude of the hydrostatic stress increases. When the hydrostatic stress component reaches about three times the unconstrained matrix yield strength, fracture is likely. If a stress in the particles is about 3–3.5 times the unconstrained matrix flow stress, fracture through the matrix occurs, initiated by *particle cracking*. Such behavior is typical for carbide and cermet materials, where the hard dispersion of carbide and oxide particles cannot be deformed, see e.g. Bergmann and Wassermann [7]. Under load, the internal shear stress, τ_i, is equal to the product of the applied stress and the number of dislocation loops piled up against the particles:

$$\tau_i = n\sigma \tag{8}$$

The number of piled-up dislocations, n, is proportional the interparticle spacing, S_p, by

$$n = \sigma S_p / G_m |\vec{b}| \tag{9}$$

Combining Eqs. (8) and (9) we obtain the internal shear stress:

$$\tau_i = \sigma^2 S_p / G_m |\vec{b}| \tag{10}$$

When the shear stress acting on the particles is equal to their fracture stress σ_{fp}, the particles will fracture, nucleating a crack which will cause yielding of the composite:

$$\tau_i = \sigma_{fp} = G_p / C \qquad \text{at yield,} \tag{11}$$

where C defines the strength of the composite as:

$$\sigma_{cy} = (G_m G_p |\vec{b}| / C S_p)^{1/2} \tag{12}$$

This expression predicts that the composite yield strength is proportional to the inverse square root of the interparticle spacing, S_p, when the particles do not yield under load, see Fig. 6 [6]. When the interparticle spacing is $S_p > 0.5$ μm, the strength drops due to crack-nucleation at weak particle-particle boundaries.

In cases where the constrained matrix creates stresses high enough to *deform the dispersed particles*, the yield criterion is:

$$\tau_i = \tau_{py} = \sigma^2 S_p / G_m |\vec{b}| \tag{13}$$

Under the hydrostatic stress, the particle yield strength is reduced from its unconstrained yield strength by an amount proportional to the constraint:

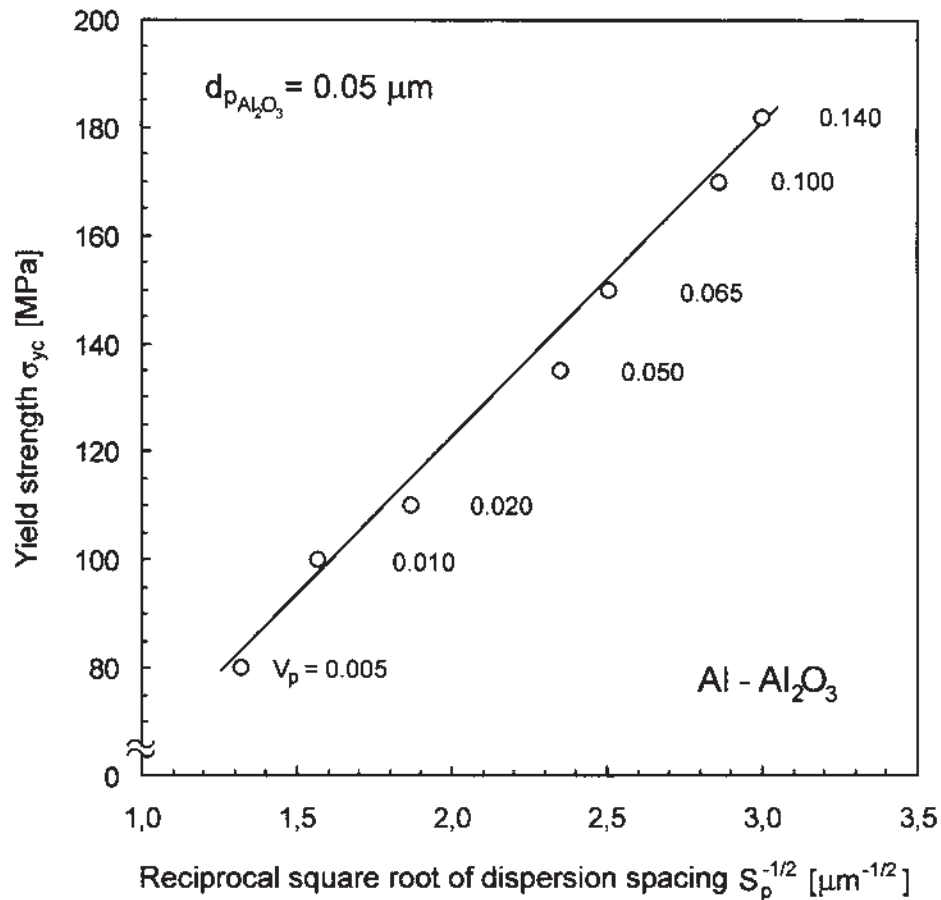$$\tau_{py} = \frac{G_p S_p}{C*} \tag{14}$$

Figure 6 Yield strength of an alumina (Al_2O_3) particle reinforced Al matrix composite as a function of the reciprocal square root of dispersion spacing, S_p. (From Ref. 6.)

Combining Eqs. (13) and (14) we obtain the yield strength relation for deformable particle composites as

$$\sigma_{cy} = (G_p G_m |\vec{b}|/C^*)^{1/2} \tag{15}$$

This expression describes the independence of the yield stress of the composite of the interparticle spacing, see e.g. [44]. This unusual behavior has contributed much information to an understanding of the deformation and strengthening behavior of particulate composites in general.

4 FIBER- AND LAMINATE-REINFORCED COMPOSITES

4.1 Continuous Fibers and Layers

In microcomposite materials with fiber or lamellar structure, the second components are definably dispersed throughout the metallic matrix. In laminated composites, the phases are sequentially arranged, as shown in Fig. 1(c). For this discussion, we assume that the second phase is uniform, continuous, and unidirectional, and is bound by the matrix so that no slippage can occur at the matrix-fiber or matrix-lamellae interface. Under the isostrain condition, $\varepsilon_c = \varepsilon_m = \varepsilon_f$, the total load, F_c, that is applied to the composite, is shared between the fiber load, F_f, and the load on the matrix, F_m. Therefore:

$$F_c = F_m + F_f$$

or, in terms of stresses,

$$\sigma_c = \sigma_m V_m + \sigma_f V_f \tag{16}$$

where V_m and V_f represent the volume fractions of matrix and fibers, respectively. When slippage does not occur at the interface, the isostrain criterion is fulfilled, and Eq. (16) may be rewritten as:

$$\sigma_c = E_m \varepsilon_c V_m + E_f \varepsilon_c (1 - V_m) \tag{17}$$

since $V_m + V_f = 1$. The ratio of the load carried by the fibers or layers to the load carried by the matrix is:

$$\frac{F_f}{F_m} = \frac{E_f}{E_m}\left(\frac{1 - V_m}{V_m}\right) \tag{18}$$

This ratio is plotted in Fig. 7 as a function of the ratio of the elastic moduli E_f/E_m and the volume fraction of the second phase. To obtain optimum reinforcement by the fibers or lamellae, the modulus of the second phase should be higher than the modulus of the metallic matrix. Under this condition, the reinforcement carries the main contribution of the total applied load and attains the maximum stress compared to the matrix. The theoretical maximum of the volume fraction of a fibrous second component is almost $V_f = 0.91$. However, above a volume fraction of the fibers of about $V_f \geq 0.8$, most artificially prepared composites exhibit binding problems at the interfaces and many superior properties of the composites begin
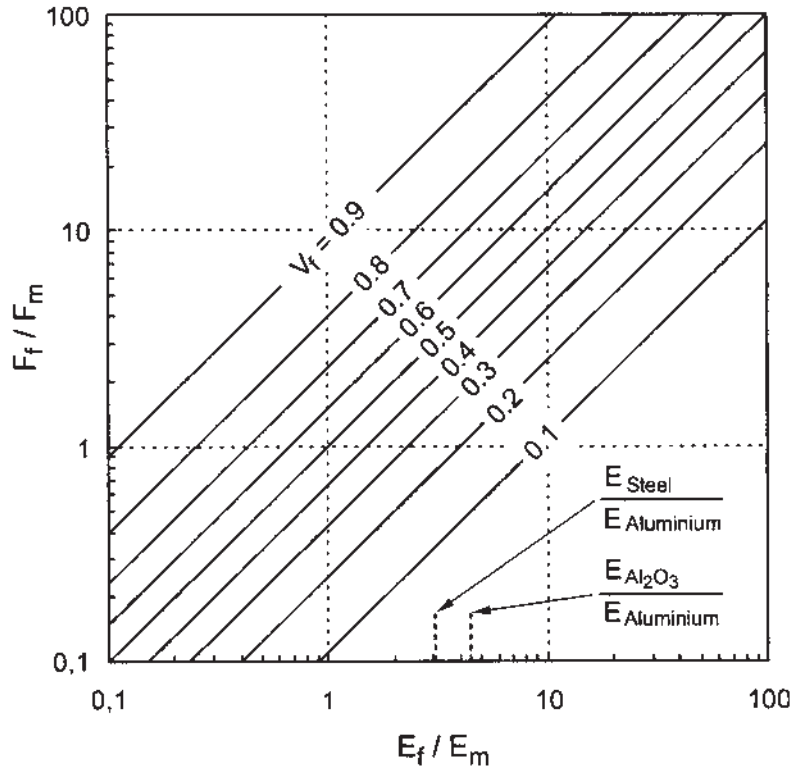
Figure 7 Ratio of the fiber-to-matrix load versus ratio of the elastic moduli for various volume fraction of the fibers. (From Ref. 12.)

to deteriorate. When a composite with uniaxially aligned and continuous fibers or laminates is stressed in an isostrain mode, the deformation behavior can be classified into *four different stages*:

- The fibers or lamellae *and* the matrix deform elastically;
- The matrix deforms plastically and the fibers or lamellae continue to deform elastically;
- The matrix *and* the fibers or lamellae deform plastically;
- Fiber fracture is followed by complete fracture.

These four stages are shown schematically in Fig. 8. In Stage I, the elastic modulus of the composite can be predicted accurately using the rule of mixtures,

$$E_c = V_p E_f + V_m E_m \tag{19}$$

The largest portion of the composite stress-strain curve belongs to Stage II, where the matrix stress-strain curve is not linear. The composite modulus may be determined at different stress-strain levels by

$$E_c = E_f V_f + \left(\frac{d\sigma_m}{d\varepsilon_m}\right)_{\varepsilon_f} \cdot V_m, \tag{20}$$
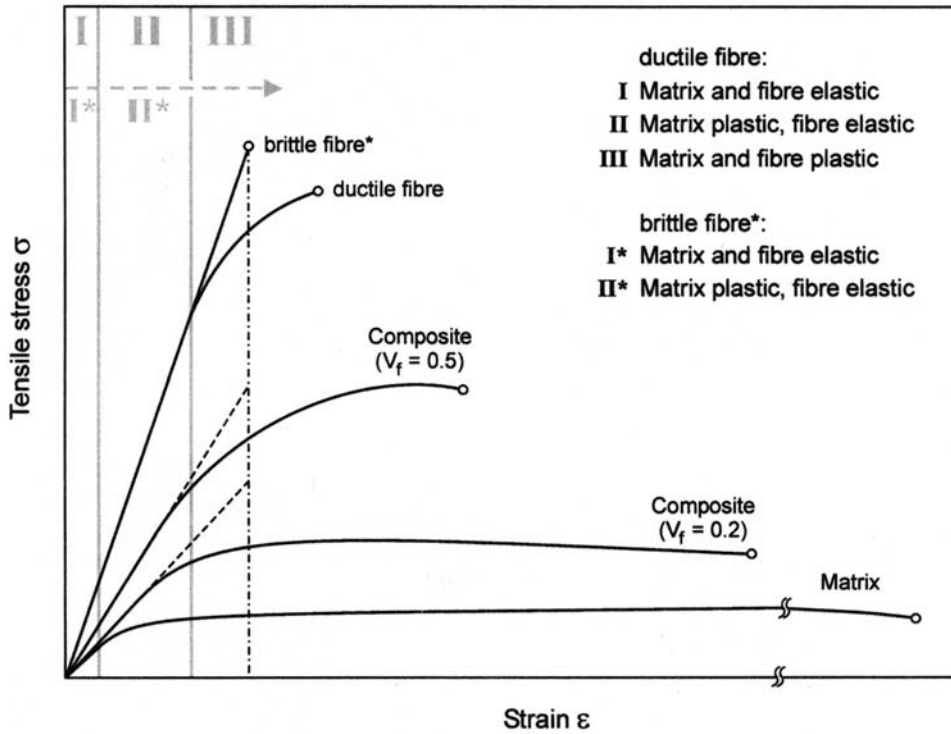
Figure 8 Stress strain curves of a ductile metal reinforced with ductile and / or brittle fibers of equal fracture stress. (From Ref. 39.)

where $(\mathrm{d}\sigma_m/\mathrm{d}\varepsilon_m)_{\varepsilon_f}$ is the slope of the matrix stress-strain curve at different strains ε_f of the fibers. In Stage III, the modulus of the composites with ductile fibers must be modified in a similar way to the reduction of the matrix modulus as

$$E_c = \left(\frac{\mathrm{d}\sigma_f}{\mathrm{d}\varepsilon_f}\right)_{\varepsilon_f} \cdot V_f + \left(\frac{\mathrm{d}\sigma_m}{\mathrm{d}\varepsilon_m}\right)_{\varepsilon_f} \cdot V_m. \tag{21}$$

In composites with ductile fibers which fracture by necking during a larger amount of plastic deformation, additional factors such as a hydrostatic stress state exerted by the ductile matrix to prevent necking of the fibers cause deviations from the rule of mixtures.

For metallic composites containing more than a minimum volume fraction, V_{min}, of continuous fibers or lamellae, the ultimate tensile strength will be ideally reached at a total strain of the fibers at their ultimate tensile strength. The following expression describes the ultimate tensile strength of a composite:

$$\sigma_c^u = \sigma_f^u V_f + (\sigma_m)_{\varepsilon_f}(1 - V_f), \tag{22}$$

where $V_f > V_{min}$ and σ_f^u is the ultimate tensile strength of the reinforcing fibers. $(\sigma_m)_{\varepsilon_f}$ is the stress of the metallic matrix when the fibers are strained to their ultimate tensile strain. The value of $(\sigma_m)_{\varepsilon_f}$ for metallic matrices is equal to the flow stresses at the ultimate tensile strain of the fibers. This value can be determined from the

stress-strain curve of the metallic matrix. The extensive *work-hardening rate* of the matrix can be explained by dislocation pile-up at the matrix-fiber interfaces. This was shown by Neumann and Haasen [62] on tungsten-fiber reinforced copper composites. Equation (22) predicts a linear increase of the yield strength of the composites with increasing volume fraction of the reinforcing components. This relation has been verified for many systems, such as Al-Al$_2$O$_3$ whiskers, Al-stainless steel, Al-SiC, and Al-Inconel 601 fibers, etc. Some of the data are represented in Fig. 9, and show fairly good agreement with the rule of mixtures curve. However, the exception is the Al (99.99)-Inconel 601 composite, where the experimental values fall slightly above the predicted linear relationship. In this case, the aluminum matrix provides a hydrostatic stress component on the ductile steel fibers and thus restrains necking. This effect leads to enhanced strength, and to a positive deviation from the predicted composite strength.

It should be noted that for very small volume fractions of the second phase, V_f, the mechanical properties of the composite may not follow the prediction by Eq. (22). This is because there is an insufficient number of fibers and lamellae to restrain
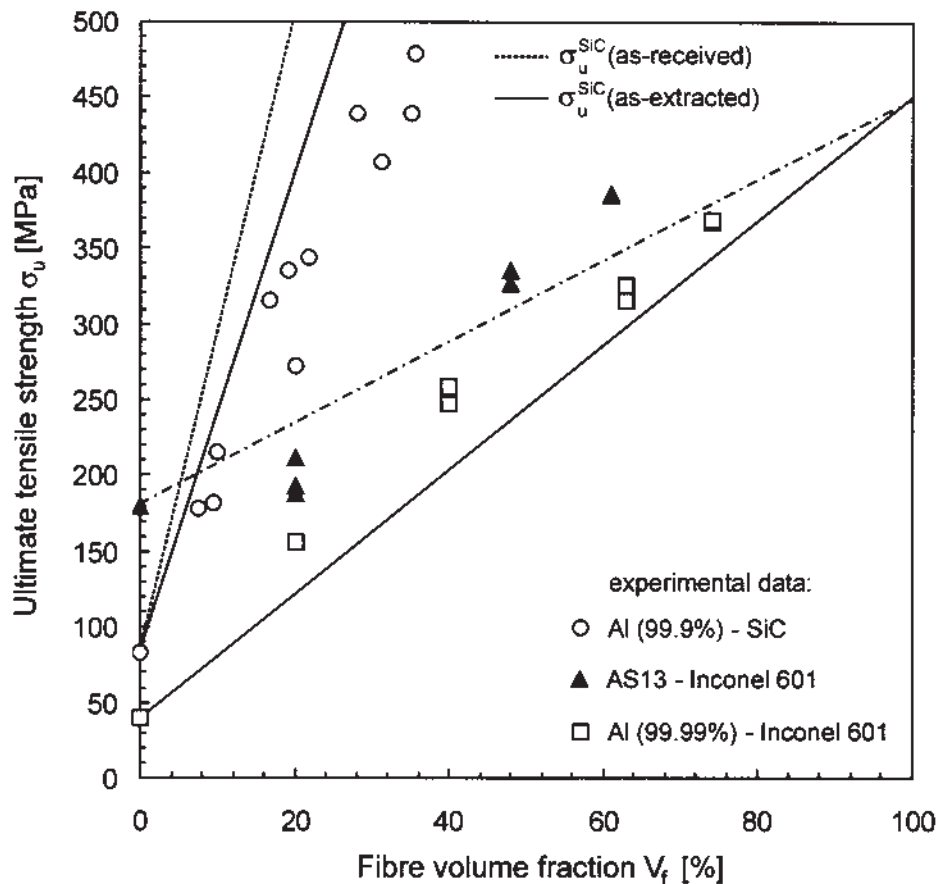
Figure 9 Strength of SiC and Inconel 601 fiber reinforced strength Al matrix composites versus fiber volume fraction. (From Ref. 10.)

the elongation of the matrix phase, so that the fibers are rapidly stressed to their fracture point.

For effective reinforcement of the matrix phase (e.g. in work-hardened or precipitation-hardened metallic matrices), the strength of the composite must exceed the *ultimate tensile strength of the matrix*, σ_m^u, so that $\sigma_c^u \geq \sigma_m^u$. This condition defines a critical volume fraction of the reinforcing second component which must be exceeded for effective strengthening, see Kelly and Davis [39]. This volume fraction can be determined by the expression

$$V_f > V_{crit} = \frac{\sigma_m^u - (\sigma_m)_{\varepsilon_f}}{\sigma_f^u - (\sigma_m)_{\varepsilon_f}}, \tag{23}$$

which is graphically represented in Fig. 10. The minimum volume fraction is given by the following expression:

$$V_{min} = \frac{\sigma_m^u - (\sigma_m)_{\varepsilon_f}}{\sigma_f^u + \sigma_m^u - (\sigma_m)_{\varepsilon_f}}. \tag{24}$$
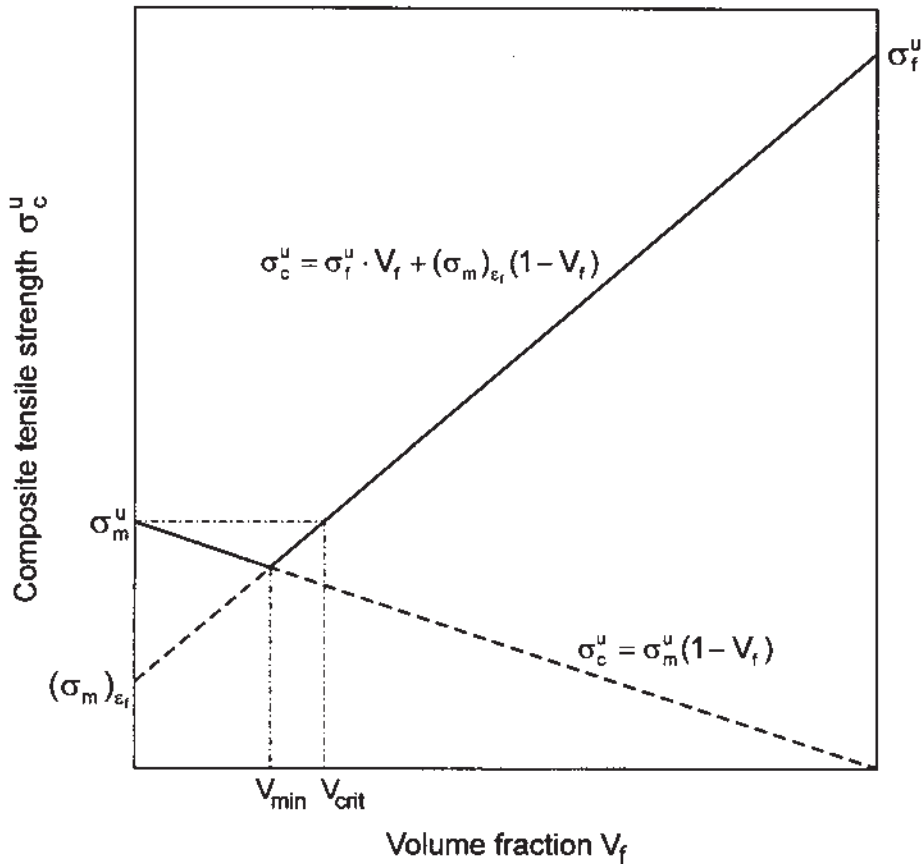
Figure 10 Schematic diagram shows the variation of composite tensile strength, σ_c^u, with volume fraction, V_f, for reinforcement with continuous brittle fibers. (From Ref. 39.)

Table 8 Values of V_{crit} for Fibers of Various Strengths, σ_{f}^{u}, in a Ductile Al Matrix

Matrix metal	$(\sigma_{\text{m}})_{\epsilon_{\text{f}}}$ (MPa)	σ_{m}^{u} (MPa)	Critical fiber volume fraction V_{crit}			
			σ_{f}^{u} 700 MPa	σ_{f}^{u} 1750 MPa	σ_{f}^{u} 3500 MPa	σ_{f}^{u} 7000 MPa
Al	28	84	0.083	0.033	0.016	0.008

In all cases V_{min} is less than V_{crit}. Table 8 presents some selected values of V_{crit} for composites consisting of strong fibers in a ductile Al metal matrix at room temperature. It is evident that the critical volume fraction increases with increasing work-hardening in the matrix $(\sigma_{\text{m}}^{u} - (\sigma_{\text{m}})_{\varepsilon_{\text{f}}}) \Rightarrow \max$, and also as the strength of the metallic matrix phase approaches that of the fibers $(\sigma_{\text{f}}^{u} - (\sigma_{\text{m}})_{\varepsilon_{\text{f}}}) \Rightarrow \min$. Thus, reinforcement of a strong matrix by fibers of comparable strength requires a large critical volume fraction of the second component.

An interesting example of a composite with a continuous layer structure is the directionally solidified Al-CuAl$_2$ eutectic, which consists of alternating ductile Al and reinforcing intermetallic CuAl$_2$ single crystal platelets. The volume fractions of the two phases are approximately equal, and the thickness of the platelets is about 1–2 μm (see Fig. 3). The strong anisotropy of the mechanical properties of this system was investigated by Hertzberg et al. [27]. Other composite systems like this also show strongly anisotropic properties. The elastic stiffness and the strength of the composite decrease as the Al and Al$_2$Cu lamellae are rotated away from the maximum applied normal stress direction $(\theta = 0)$. The influence of the microstructural anisotropy on the tensile strength of this laminated composite materials is illustrated in Fig. 11. As mentioned above, the tensile strength decreases continuously with increasing degree of misorientation of the lamellae over the whole range from the isostrain to the isostress limits. A particularly surprising feature of the plastic deformation of the platelets is the unusual deflection, which develops a maximum for lamellar orientations between $30° \leq \theta \leq 45°$ to the tensile axis, see Fig. 11. This extraordinary increase in ductility of the composite may be caused by slip systems activated in the Al$_2$Cu lamellae. This behavior of the laminated composites indicates both microstructural and crystallographic anisotropy.

4.2. Discontinuous Fiber Composites

The properties of composite materials reinforced with continuous fibers are relatively simple to describe using the rule of mixtures, since the stress profile along any single fiber is nearly uniform. The rule of mixtures must be modified for describing the properties of discontinuous fiber-reinforced composites, see e.g. [14]. The fiber ends will not support the same load as the middle part, hence there is a critical length, l_{c}, called the fiber transfer length, which is the smallest fiber length for which the maximum fiber stress, $\sigma_{\text{f}}^{\max}$, may be achieved (Fig. 12).

For fibers longer than the critical lengths l_{c}, a certain fraction will support the maximum fiber stress, while a fixed fiber length $l_{c/2}$ at each end has a stress profile which decreases from $\sigma_{\text{f}}^{\max}$ to zero. When these *fiber-end effects* are taken into account, an equation analogous to Eq. (22) may be developed. The quantity σ_{f} is defined as the ratio of the area under the stress curve to that part which supports
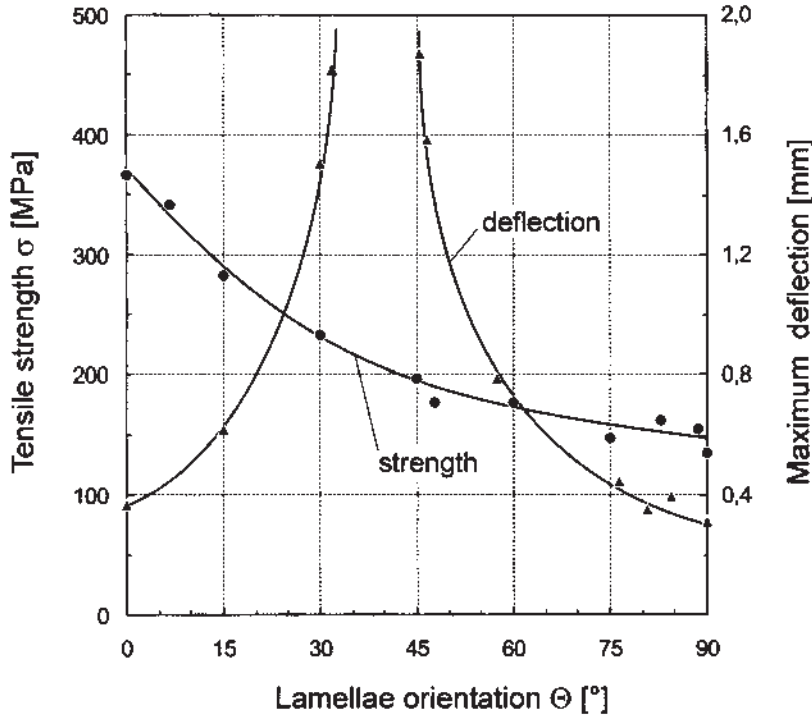
Figure 11 Tensile strength and lamellae deflection versus lamellae orientation relative to tensile axis, Θ, of a directionally solidified Al-CuAl$_2$ eutectic alloy. (From Ref. 27.)

σ_f. Under this assumption, the ineffective fiber ends contribute a reduced average stress $\alpha \sigma_f^{max}$ to the transmitted maximum stress σ_f^{max}. The average stress, $\bar{\sigma}_f$, in the discontinuous fibers is given by expression:

$$\bar{\sigma}_f = \frac{1}{l} \int_0^l \sigma_f dl = \sigma_f \left[1 - (1 - \alpha)\frac{l_c}{l} \right]. \tag{25}$$

In the case of an ideal plastic matrix with yield stress τ_m^0, α is equal to 0.5 and the average fiber stress can be described as

$$\bar{\sigma}_f = \sigma_f \left(1 - \frac{l_c}{2l} \right). \tag{26}$$

The critical ratio of the fiber length, l_c, to fiber diameter, d_f, is defined as

$$\frac{l_c}{d_f} = \frac{\sigma_f^{max}}{2\tau_m^u}. \tag{27}$$

For a work-hardened metallic matrix, τ_m^0 is identical with the ultimate shear strength of the matrix, and in the case of interfacial failure τ_m^0 must be considered as the interfacial shear strength between the fiber and the matrix. However, in an aligned

short fiber composite two important failure mechanisms are possible, depending on the relative values of τ^0, σ_f^u and σ_f^u. In one case, fiber fracture occurs before fracture in the matrix takes place. The build-up of stress in the fibers, as the load on the composite is increased, can be described by Eq. (26). If the average stress in the fiber reaches the stress for fiber fracture, then the ultimate tensile strength of the composite can be described by

$$\sigma_c^u = \left(1 - \frac{l_c}{2l}\right)\sigma_f^u V_f + (\sigma_m)_{\varepsilon_f}(1 - V_f). \tag{28}$$

The second failure condition, illustrated in Fig. 12, occurs when the build-up of stress in the fiber is insufficient to cause fiber fracture and the average fiber stress,

$$\bar{\sigma}_f = \tau\frac{l}{r}, \tag{29}$$

reaches the fracture stress of the matrix. This relationship is valid for $l < l_c$. The tensile strength of an aligned short fiber composite is given by

$$\sigma_f^u = \left(\frac{l}{2r}\tau\right) + \sigma_m^u(1 - V_f). \tag{30}$$

Equations (28) and (30) allow the prediction of the variation in strength as a function of the geometry of the discontinuous fibers if σ_f^u and τ are known. Equation (28) indicates that if the discontinuous fiber length, l, is equal to the critical fiber length, l_c, then 50% of the maximum fiber strength will be realized. When $l \geq l_c$, then 90% or more of the tensile strength of the fibers will be reached. Therefore, if the fiber strength or the aspect ratio, l/d_f, is sufficiently large, the composite with reinforced discontinuous fibers approaches the strength of the composites with continuous fibers. This case is illustrated in Fig. 13 for a boron-fiber reinforced aluminum matrix [59].

It is shown that for fibers having an aspect ratio $l/d_f < \infty$ a deviation from the linear upper bound line, predicted by the rule of mixtures, occurs. This deviation increases with decreasing values of the fiber aspect ratio. For $l/d_f = 60$ and fiber
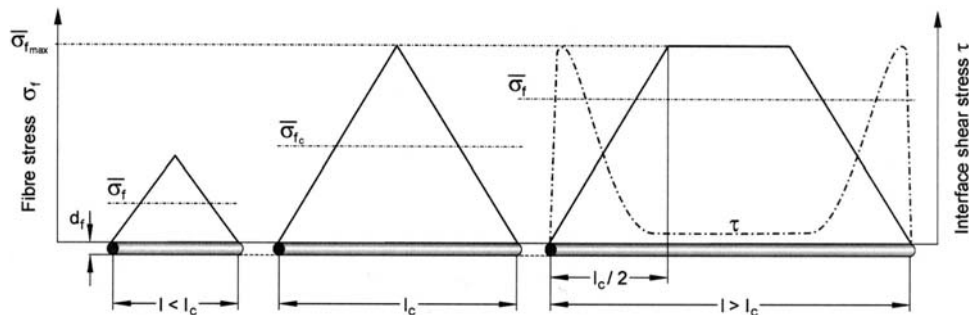
Figure 12   Schematic representation of fiber tensile stress, σ_c, and interfacial sheas stress, τ, when the matrix exhibits elastic deformation. (From Ref. 39.)
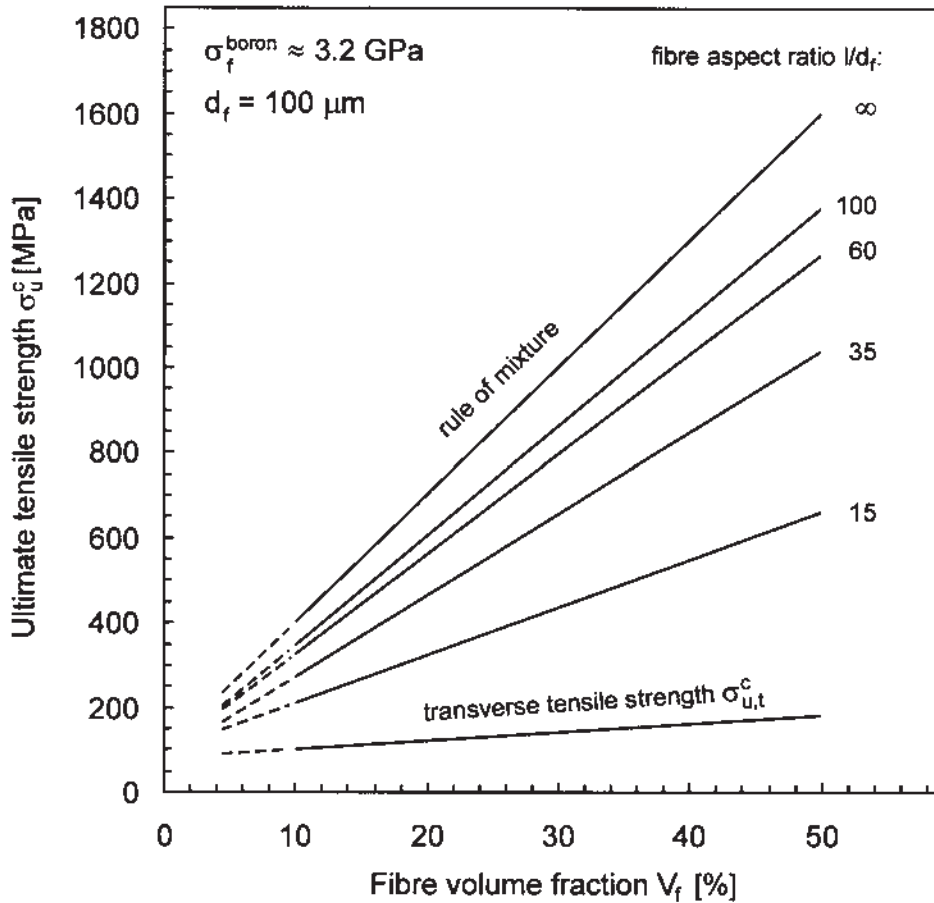
Figure 13 Effect of fiber aspect ratio (l/d_f) and fiber volume fraction V_f on tensile strength of aluminum–boron composite at room temperature. (some data were taken from Ref. 59.)

volume fraction $V_f = 0.5$, the tensile strength of the composite materials reaches 80% of that with continuous fibers. Similar data are obtained from other systems, such as aluminum-steel fibers and aluminium-Al_2O_3 whiskers, etc. For fibers with small fiber aspect ratio, misalignment is the strongest factor which contributes to a decrease in the elastic modulus and also the tensile strength of composite materials. Kelly and Davis [39] described the effect of the fiber orientation on the elastic properties and the tensile strength, that is qualitatively represented in Fig. 14.

A misalignment of about $\Phi_{crit} \leq 8°$ has no influence on the discussed mechanical properties of the composites. Beyond this critical angle, tensile strength and elastic modulus diminish rapidly with increasing orientation angle, and at $\Phi = 45°$ the curve passes its minimum. With further increase of the Φ values tensile strength and elastic modulus increase slightly until the curve reaches the intersection with the isostress-condition curve. When the fibers are oriented perpendicularly to the load direction, then the isostress condition is exactly fulfilled, and the properties
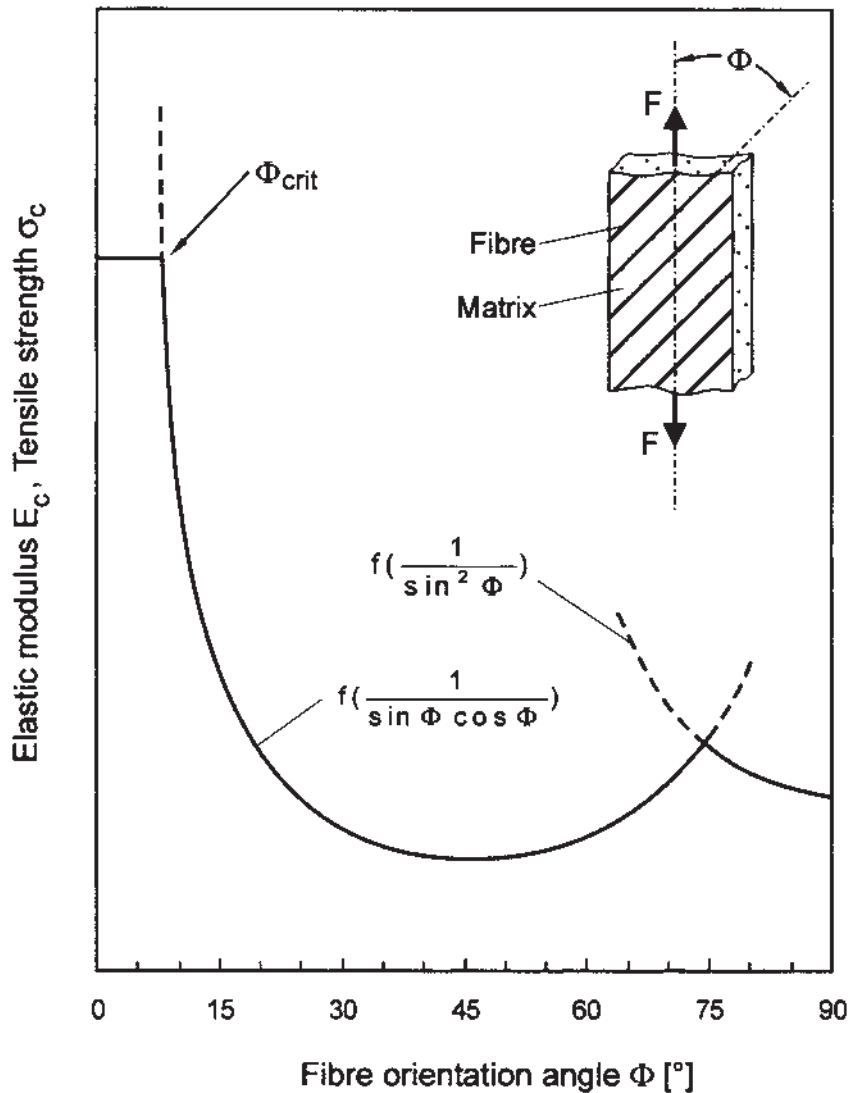
Figure 14 Variation in elasticity and tensile strength of a fiber-reinforced composite as a function of fiber misorientation. (From Ref. 39.)

of the composite material can be described by the lower-bound isostress rule of mixtures.

Koppenaal and Parikh [40] determined the tensile strength as a function of the volume fraction of the second phase for different orientations of steel fibers in a silver matrix. The obtained results are represented in Fig. 15. It is evident that composite materials with discontinuously aligned fibers of varying orientation relative to the stress axis have lower strengths than composites with continuously aligned fibers. Composites with randomly distributed fibers exhibit a minimum reinforcement effect. Similar behavior is observed in directionally solidified eutectics, where stiff
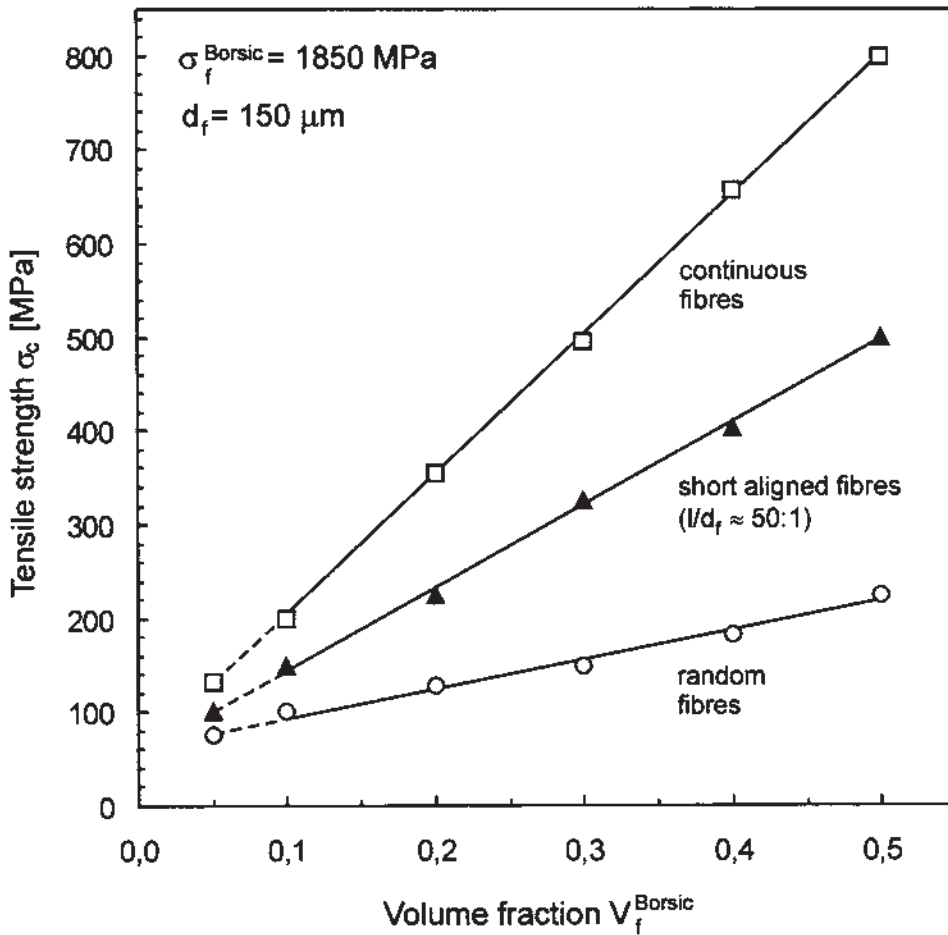
Figure 15 Effect of fiber volume fraction and fiber orientation on the strength of Al-Borsic fiber composite (From Ref. 12.)

whiskers of pure metallic and intermetallic compounds are embedded in a ductile metallic matrix.

The first example of a fiber-reinforced eutectic alloy prepared by controlled solidification was the Al-Al$_3$Ni system. This eutectic alloy contains 10% volume fraction of Al$_3$Ni whiskers, surrounded by the ductile Al matrix. The reinforcement effect of the unidirectionally aligned intermetallic Al$_3$Ni whiskers is compared with that of randomly distributed Al$_3$Ni phase in the as-cast condition in Fig. 16.

All composite materials containing misaligned fibers are weaker than those with an aligned fiber structure. This graph shows that the directionally solidified fiber structure causes a marked improvement in strength to about $\sigma_c^u = 350$ MPa. In contrast to this, the as-cast eutectic alloy exhibits a tensile strength of $\sigma^u = 90$ MPa and an elongation to failure of $\varepsilon_{tot} = 15\%$. Calculations of the ultimate tensile strength of the controlled-solidified Al-Al$_3$Ni composite, using Eq. (28), yield $\sigma_c^u = 335$ MPa, in good agreement with the experimentally determined values.
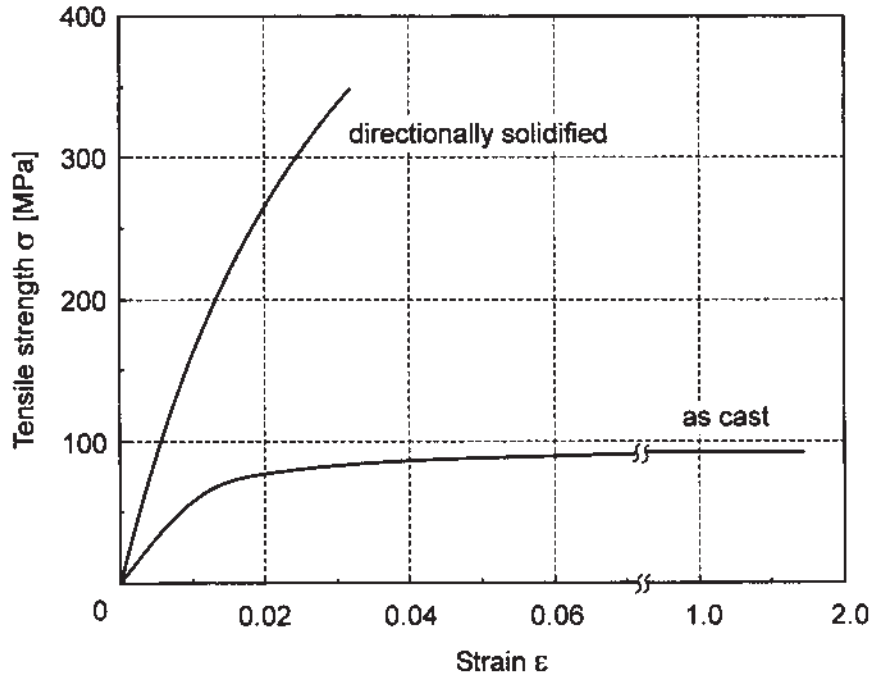
Figure 16 Tensile properties of as-cast and unidirectionally solidified Al-Al$_3$Ni eutectic alloy. (From Ref. 48.)

Improvements in stiffness and strength by controlled eutectic structures have been obtained in many other aluminum-based alloys or high aluminum containing intermetallics, such as Al-CuAl$_2$, Al-Si, Al-AlLi, Al-Al$_2$Ca, Al-Be, NiAl-X (X = Cr, Cr, V, Mo, Re, W) etc. see e.g. [46].

5 SUPERPLASTICITY IN REINFORCED ALUMINUM-BASED METAL-MATRIX COMPOSITES (MMCS)

5.1 Fine-structure superplastic properties

Particle or short fiber reinforced MMCs can behave superplastically [29, 52, 65]. This unusual material property is characterized by high uniform and total elongations of several hundred or thousand percent of strained tensile samples at temperatures above 0.6 T_m (T_m —absolute melting temperature) and at medium or high strain rates of the order of 10^{-4} to 10^0 sec^{-1}.

The important parameter which describes the superplastic behavior at a given temperature T and average grain size is the strain-rate sensitivity exponent m, defined as:

$$m = \left(\frac{\delta \ln \sigma}{\delta \ln \dot{\varepsilon}}\right)_{T,GS} \tag{31}$$

Comparative data of representative superplastic aluminum-based SiC and Si_3N_4 particulate or whisker reinforced composites are summarized in Table 9. These materials have been produced either by powder metallurgy (PM) or ingot metallurgy (IM), specifically stir-casting [41].

The composites listed in Table 9 were produced by conventional PM techniques, except the 15 vol% SiC_w-7475 Al, which was manufactured using SiC_w layers between specially prepared foils of superplastic 7475 Al alloy [54]. The IM composite, 10 vol% SiC_p-2024 Al [81], was produced by stir-casting. For discussion purposes, the composites are hereafter abbreviated as reinforcement-matrix alloy, e.g. 20 vol% SiC_w-2124 Al becomes SiC_w/2124. The reported strain rate sensitivity exponents 'm' vary significantly. An exceptionally large elongation of 1400% was recorded in one of the thermally cycled materials (SiC_w/6061). The matrix grain size for the SiC_w/7475, SiC_p/64, and SiC_p/2024 alloys were about 6 μm, whereas the grain sizes for the other composites were of the order of 1 μm.

The results in Table 9 can be classified into two groups according to the test method. The first group, in which test samples are thermally cycled between two temperatures, is termed *internal stress superplasticity*; this phenomenon has been described elsewhere [67]. The other group, which is the most commonly known, is based on isothermal tests.

In the case of SiC_w/2024 and SiC_w/6061, superplasticity was observed under nonisothermal test conditions [23, 82]. The concept of internal-stress superplasticity—an example of which is phase-transformation superplasticity [37, 70]—was used to demonstrate superplasticity under conditions of thermal cycling between 100°C and 450°C. Under the thermal-cycling conditions, internal stresses are generated from the thermal expansion coefficient mismatch between the reinforcement, i.e. SiC whisker, and the aluminum matrix. With a small externally applied stress, the composites behave like Newtonian viscous fluids ($m = 1$). This results in superplastic elongations of about 300% in the SiC_w/2024 and an exceptionally large elongation of 1400% in the SiC_w/6061. In contrast, the composites exhibit a low m value (~ 0.1) with only 10% elongation under isothermal creep deformation conditions. For both the thermal-cycling and isothermal tests, the average strain rate is about 10^{-4} to 10^{-5} sec^{-1}, i.e. a range of values similar to those used for conventional superplastic forming of aluminum alloys. A direct comparison of the results from SiC_w/2024 between the two different test conditions is shown in Fig. 17.

For superplasticity under isothermal conditions, the results can be further categorized into two subgroups according to the strain rate at which superplasticity occurs. As illustrated in Table 9, the strain rates at which superplasticity occurs in all the composites except the SiC_w/2124 composite are approximately 10^{-4} to 10^{-5} sec^{-1}, which results in tensile elongations of about 300–400 %. These strain rates are of the same orders of magnitude as for conventional superplasticity. The m values for these composites are about 0.4.

The SiC_w/7475 composite was manufactured using SiC_w layered between specially prepared foils of superplastic 7475 Al alloys. As a result, the superplastic behavior is similar to that of the monolithic 7475 Al alloy. The three PM SiC particulate-reinforced composites are noted to show superplasticity only under conditions in which a back pressure was applied in order to reduce cavitation and thus delay fracture. In the absence of back pressure, the composites behave like

Table 9 Superplastic Properties of SiC Reinforced Al Composites

Material composition[a]	Material code	T (°C)	Superplastic elongation (%)	m-value	Strain rate sec^{-1}	Stress (MPa)
20 vol% SiC$_w$-2124Al [66]	SiC$_w$/2124	475–550	~300	~0.33	3.3×10^{-1}	~10
20 vol % SiC$_w$-2024Al [83]	SiC$_w$/2024	100 450[b]	~300	1.0	5×10^{-4}	~15
20 vol % SiC$_w$-6061Al [22]	SiC$_w$/6061	100 450[b]	~1400	1.0	1×10^{-5}	~7
15 vol % SiC$_w$-7475Al [55]	SiC$_w$/7475	~520	350	0.5	2×10^{-4}	~7
10 vol % SiC$_p$-PM64Al[c] [55]	SiC$_p$/64	~516	~250	0.5	2×10^{-4}	~1.4
15 vol % SiC$_p$-2014Al[d] [70]	SiC$_p$/2014	480	395	0.4	4×10^{-4}	
15 vol % SiC$_p$-7475Al[d] [70]	SiC$_p$/7475	515	442	0.38	2×10^{-4}	
10 vol % SiC$_p$-2024Al [82]	SiC$_p$/2024	515	685	0.4	5×10^{-4}	~5

[a] w= whisker; p = particulate.
[b] Thermal cycling.
[c, d] Back pressures of 4.14 MPa and 5.25 MPa applied.
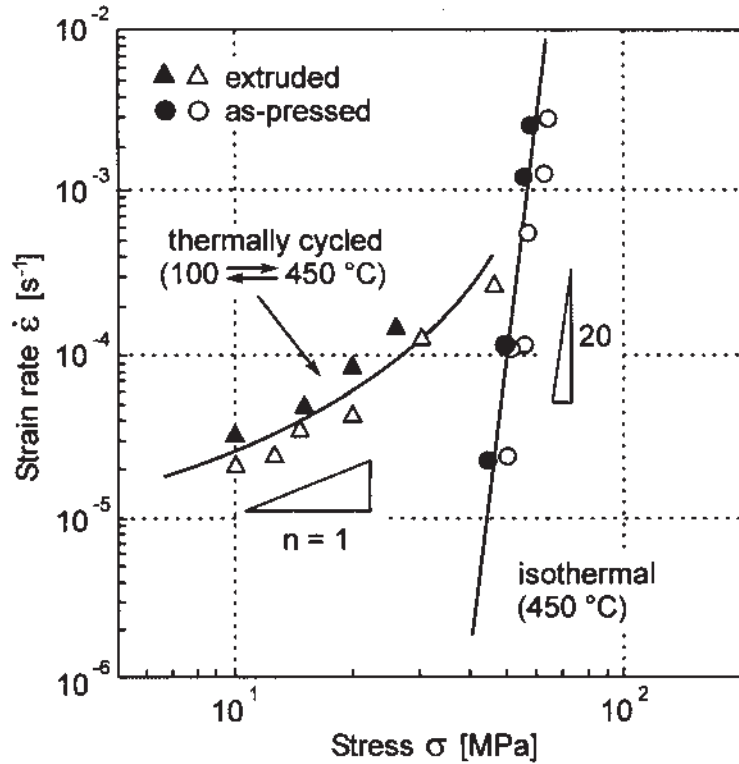
Figure 17 Comparison of the plastic properties of 20 vol% SiC_w-reinforced 2024Al deformed under isothermal and thermal-cycling conditions. (From Ref. 82.)

conventional alloys. The fact that the materials do not behave superplastically without back pressure indicates severe cavitation in these materials at strain rates of approximately 10^{-4} sec^{-1}. This is consistent with Fig. 17., i.e. under isothermal conditions, it is difficult to make Al-based composites superplastic at strain rates usually observed in conventional superplastic Al alloys. In contrast, the composite $SiC_p/2024Al$ (prepared by IM) is superplastic without the application of back pressure. This may be attributed to a relatively homogeneous microstructure in the stir-cast composite. To achieve superplasticity in PM composites without applying back pressure, it is necessary to carry out experiments at an extraordinarily high strain rate (higher than 10^{-2} sec^{-1}). This high-strain-rate phenomenon, denoted as high-strain-rate superplasticity (HSRS), has been also observed in Al composites reinforced with other ceramic phases, e.g. Si_3N_4, TiC, and in some other advanced Al alloys.

5.2 High-strain-rate Superplasticity

It is often thought that superplasticity is only found at relatively low strain rates, typically about 10^{-4} to 10^{-3} sec^{-1}. Several recent studies have indicated, however, that superplasticity can exist at strain rates considerably higher than 10^{-2} sec^{-1}. This high-strain-rate superplasticity (HSRS) phenomenon has now been observed

Table 10   HSRS Aluminum Alloys and Composites

Material/reference[a]	Test temperature (C)	Solidus (C)	Strain rate (sec^{-1})	Stress (MPa)	Elongation (%)
Composites					
β SiC$_w$/2124Al [65]	525	502	0.3	~10	~300
β Si$_3$N$_{4(w)}$/2124Al [33]	525	502	0.2	~10	~250
α Si$_3$N$_{4(w)}$/7064Al [32]	525	~525	0.2	~15	~250
β Si$_3$N$_{4(w)}$/6061Al [52]	545	582	0.5	~20	~450
β SiC$_w$/6061Al [30]	550	582	0.2	6.5	300
27 wt% β SiC$_w$/7075Al [55, 73]	~500	< 538	0.2	40	—
AlN/6061Al [31]	600	582	0.5	10	350
TiC/2014Al [31]	545	507	0.2	15	250
Mechanically alloyed Al					
IN9021 [28]	550	495	50	22	~1250
MA SiC$_{(p)}$/IN9021 [29]	550	495	10	~7	~500

[a] w = whisker; p = particulate.

in metal-matrix composites [33, 52], mechanically alloyed materials [8, 66], and even in the more conventionally produced aluminum-based alloys and other materials [18, 63]. We will discuss the phenomenon in greater detail in the following.

HSRS was initially observed in the Al-based metal-matrix composite: SiC$_w$/2124Al [65], and has continued to be studied mainly in Al-based alloys. Composite reinforcements include SiC and Si$_3$N$_4$ whiskers and SiC particles; matrix alloys include the 2000, 6000, and 7000 series Al-based alloys. A list of published HSRS results is presented in Table 10. Despite the differences in the type of reinforcement and matrix composition, all of these composites are noted to exhibit approximately similar deformation and microstructural characteristics. In the following, we use the powder-metallurgy 20% SiC whisker-reinforced 2124Al composite (SiC$_w$/2124Al) as an example to reveal the key experimental observations of HSRS.

5.2.1   Silicon Carbide Whisker-reinforced Aluminum Composites

Up to now, reports on HSRS are found in aluminum composites mainly produced by powder-metallurgy methods. High-temperature deformation investigations on the SiC$_w$/2124Al indicated that the material was not superplastic in as-extruded conditions, over the conventional strain-rate range of 1.7×10^{-3} to 3.3×10^{-1} sec^{-1}, elongation-to-failure values of 30–40% were recorded.

After proper thermomechanical processing the composite became superplastic; Fig. 18(a) shows the elongation-to-failure value at 525ºC as a function of the initial strain rate. The strain rate range examined covered about four orders of magnitude. At the slowest strain rate in this range, 9.3×10^{-5} sec^{-1}, a non-superplastic value of elongation-to-failure of 40% was measured. This strain rate is typical of that for optimum superplasticity in many aluminum alloys. As the strain rate increased, the tensile ductility increased for the case of the SiC$_w$/Al composites. At a strain rate of 1.3×10^{-2} sec^{-1}, values of about 100% were observed both in the longitudinal
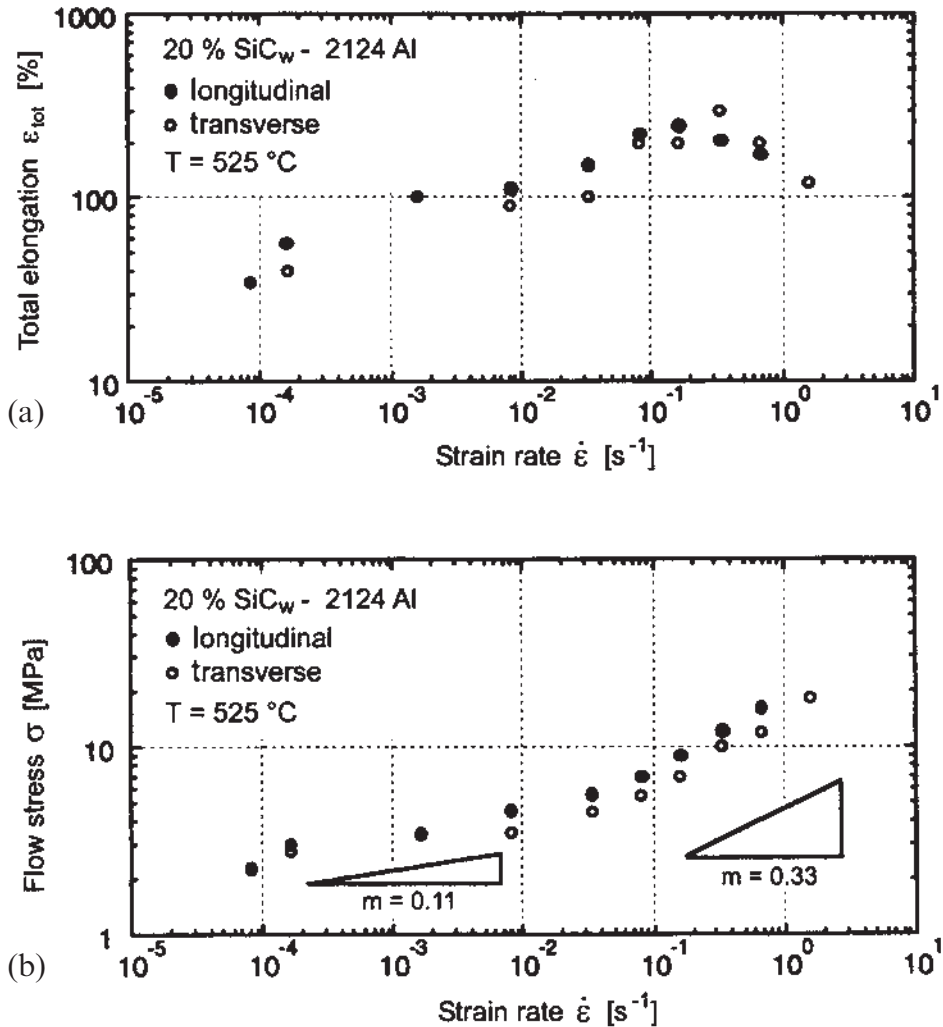
Figure 18 (a) Elongation and (b) flow stress as a function of strain rate for a Al 2124 composite containing 20 vol%SiC whisker at 525°C. (From Ref. 67, 80.)

and transverse directions. Even at higher strain rates, the elongation-to-failure values continued to increase until a maximum value of 300% was observed at a very high strain rate of 3.3×10^{-1} sec^{-1} (2000% min^{-1}). Beyond this, a decrease in elongation occurred. It is pointed out that at these high strain rates, no significant difference was observed between samples from the longitudinal and transverse directions of the processed material.

Figure 18(b) shows the corresponding flow stress (at a fixed strain of $\varepsilon = 0.3$) as a function of strain rate. As may be observed, the maximum in elongation corresponds to an increase in the strain-rate-sensitivity value from $m = 0.1$ for the low-strain-rate region to $m = 0.33$ for the high-strain-rate region. This correspond-

ing increase in *m* with higher elongations-to-failure is consistent with the observations in conventional superplastic metals.

The data in Fig. 18(a) are noted to have been measured at a temperature of 525°C, which is above the solidus temperature of the 2124Al matrix. (The solidus temperature for 2124 is 502°C [80]). Therefore, mechanical tests were carried out on the composite at temperatures of 475°C and 500°C (below the solidus temperature) and 550°C (yet further above the solidus temperature); these results are depicted in Fig. 19. In a manner similar to that found at 525°C, the data in Fig. 19 show that at high strain rates, *m* increased to about 0.3 at all temperatures, and corresponding high values of elongation-to-failure were also recorded in these strain-rate regimes. It should be emphasized particularly that 550°C is nearly 50°C above the solidus temperature of the 2124Al matrix. The data in Fig. 19, however, indicate that the flow characteristics do not exhibit a sharp discontinuity across the solidus temperature, implying that the presence of a partial liquid phase may not be the main contributing factor to the observed superplasticity. Rather low strain-rate sensitivities ($m \sim 0.1$) and therefore, low elongations to failure ($< 100\%$), were generally observed at low strain-rate regions ($< 10^{-3} \, \text{sec}^{-1}$) at each
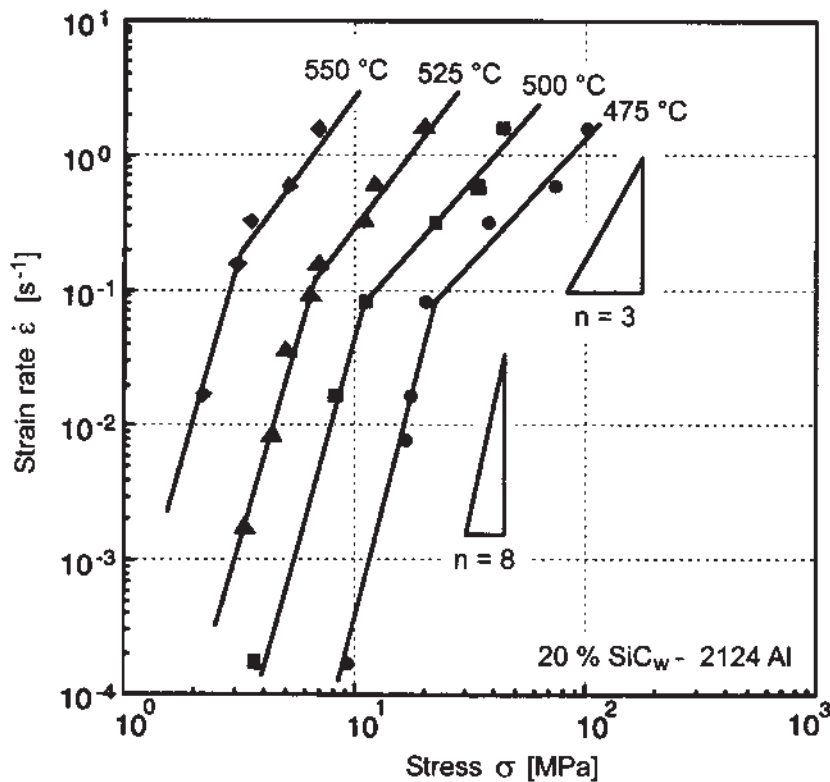

Figure 19 Strain rate-stress data for an 2124Al composite containing 20 vol%SiC whiskers at four different temperatures that bracket the solidus temperature for the matrix. (From Ref. 57.)

temperature. Low strain-rate sensitivity in the low strain-rate-regions has been suggested to be associated with the existence of a threshold stress [13, 22].

The above discussed transition in superplastic behavior from low to high strain rates is also manifested by a change in the apparent activation energy for superplastic flow. The apparent activation energies of $SiC_w/2124Al$ at the two regions are computed to be 920 and 218 kJ/mol for the low- and high-strain-rate-regions, respectively; this is shown in Fig. 20. Pandey et al. [68] argued that, analogous to oxide dispersion strengthened alloys [50], the high apparent activation energy at the low strain-rate region was a result of the existence of a threshold stress in metal-matrix-composites. Although uncertainty remains regarding the origin of threshold stresses, a relatively high activation energy at low strain rates is noted to occur in all the HSRS materials. In fact, in the present case, even the activation energy of 218 kJ/mol measured at the high-strain-rate region, where superplasticity is observed, is notably higher than the value of 140 kJ/mol, the activation energy for self-diffusion of aluminum [76].

A general question regarding the data in Fig. 19, specifically the data at 475°C and 500°C, is that of adiabatic heating. Although these test temperatures are below the solidus temperature for the 2124Al matrix, the question arises as to whether
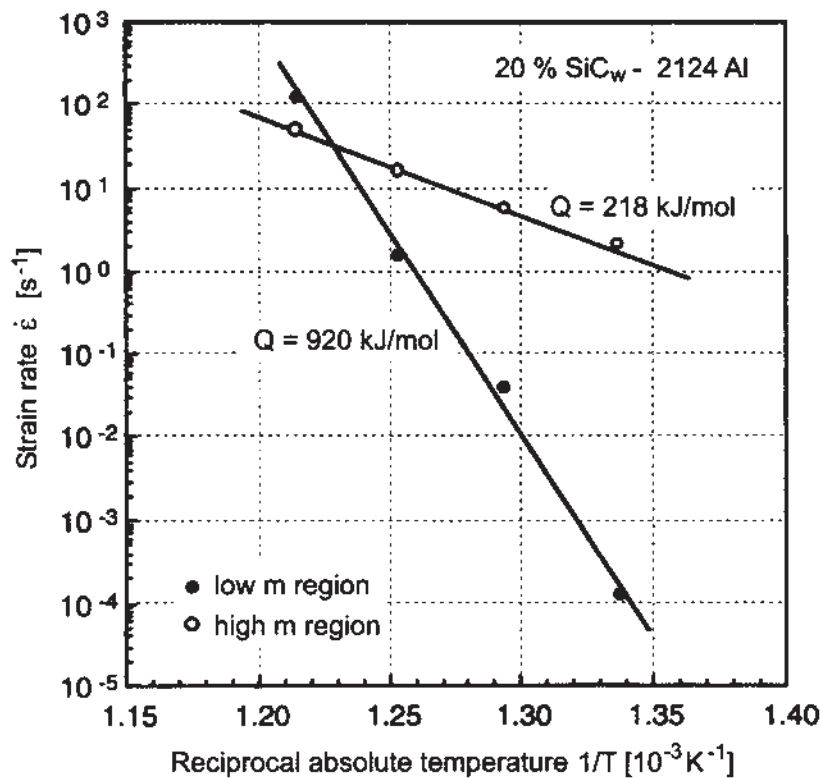

Figure 20   Apparent activation energies for the 2124Al composite containing 20 vol%SiC whisker at four different temperatures that bracket the solidus temperature for the matrix. (From Ref. 67.)

adiabatic heating during high-strain-rate deformation may have caused partial melting of the alloy matrix. By assuming that all the strain energy (i.e. the area under the stress-strain curve) was converted into heat, which was then considered to result in a uniform temperature increase in the test sample, the temperature rise in the test sample was estimated to be approximately 15°C. Macroscopic localized heating can result in a higher local temperature rise.

Only limited microstructural characterization has been performed on the superplastic $SiC_w/2124Al$ composite. This is primarily because the microstructures of metal-matrix composites are generally quite complex. Transmission electron microscopy from the gage lengths of the $SiC_w/2124Al$ sample, deformed at 525°C at a strain rate of $0.6\ sec^{-1}$ to an elongation of about 100%, revealed that some clustering of the whiskers is apparent; also, a range of diameters from less than 1 μm to about 2 μm is found for the individual whiskers. Some examinations have indicated the presence of a zone adjacent to the whiskers that is free of dislocations [47, 80].

Another important microstructural feature is that the fracture surface of a superplastically deformed $SiC_w/2124Al$ specimen is quite different from those of specimens that fail in a nonsuperplastic fashion. The chemical bonding between the SiC whisker and Al matrix is generally recognized to be good [66]. As a result, the fracture surface of a conventional SiC_w/Al sample normally exhibits localized ductile fracture of the Al matrix with only limited fiber pull out. This is the major characteristic noted on the fracture surface of a nonsuperplastic specimen. The fracture surface of a superplastically deformed specimen, however, tested at the same temperature often exhibits extensive fiber pull out, suggesting substantial debonding during deformation. This fracture surface appearance may be interpreted as a result of extensive interfacial sliding between the whisker and the matrix (but may also be interpreted as a result of the high stress-high strain rate deformation condition).

One of the most peculiar observations is the fact that HSRS is not universally observed in SiC_w/Al composites. For example, elongations-to-failure for a matrix of 6061Al with also 20 vol%SiC_w reinforcement were never observed to exceed 100%, even though the material had been extensively thermomechanically processed and tested under similar conditions to the $SiC_w/2124Al$ composite [65]. On the other hand, Huang et al. [30] demonstrated that a $SiC_w/6061Al$ prepared by a squeeze-cast technique can be superplastic (300%) at 550°C with a strain rate of $1.7 \times 10^{-1}\ sec^{-1}$; m is about 0.32. Apparently, both the matrix elements and thermomechanical processing routes can play an important role in determining whether or not superplasticity occurs in these composites.

5.2.2 Silicon Nitride Whisker Reinforced Aluminum Composites

In addition to SiC whisker-reinforced aluminum composites, HSRS has been further demonstrated in a number of Si_3N_4 (both in á and â structure) whisker-reinforced composites [32, 53]. These composites were also mainly prepared by powder-metallurgy. Table 11 lists the $Si_3N_{4(w)}$-reinforced composite combinations that are superplastic. Results from SiC_w-reinforced aluminum are included for comparison. For discussion purposes, the $Si_3N_{4(w)}$-reinforced composites listed in Table 11 are hereafter abbreviated as Si_3N_4/matrix alloy, e.g. 20 vol% $\beta\ Si_3N_4/2124Al$ becomes $\beta\ Si_3N_4/2124$.

Table 11 Summary of HSRS Whisker-reinforced Aluminum Composites

Matrix	Whisker	HSRS superplasticity (strain rate, sec^{-1})
2124Al	β Si$_3$N$_4$	3×10^{-1}
6061Al	β Si$_3$N$_4$	5×10^{-1}
7064Al	α Si$_3$N$_4$	2×10^{-1}
2124 Al	α Si$_3$N$_4$	— (No)
2124Al	β SiC	$3 \times 3 \times 10^{-1}$
6061	β SiC	— (No)

From Table 11, it is apparent that simple criterion for HSRS based on the individual types of whisker or alloy matrix is not possible. This viewpoint is supported by the observation that β Si$_3$N$_4$/6061 is superplastic but β SiC/6061 is not, which illustrates the importance of the reinforced whisker selection. However, the α Si$_3$N$_4$/7064 is superplastic, but α Si$_3$N$_4$/2124 fails to exhibit superplasticity, illustrating the importance of the alloy matrix selection.

For those composites which exhibit HSRS, superplasticity is only observed at strain rates of about 0.1 to 0.2 sec^{-1}. The elongation of the β Si$_3$N$_4$/2124 composite, plotted as a function of strain rate, is given in Figure 21. The strong resemblance between Fig. 21 and the upper part of Fig. 19, which was obtained from a β SiC/2124 composite, is apparent. It should be noted that the morphology of the β Si$_3$N$_4$ whisker is similar to that of β SiC$_w$.
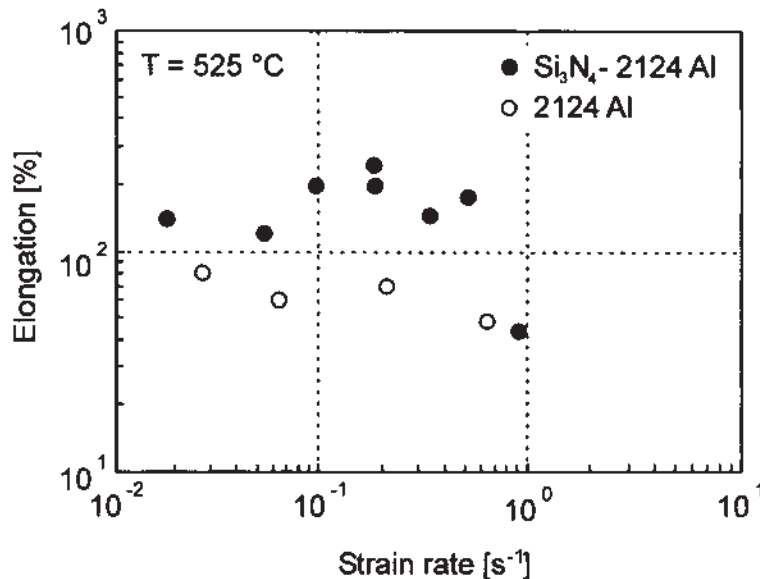

Figure 21 Elongations of 2124 Al and β Si$_3$N$_4$/2124Al composites, plotted as a function of strain rate. (From Ref. 67.)

Table 12 Grain size, Chemical Compositions, and Volume Fraction of Dispersoids in Various MA Alloys

		Dispersoids (wt%)	
Composition (wt%)	Grain size (μm)	Al_4C_3	Al_2O_3
IN9021 (Al-4.0%Cu-1.5%Mg-1.1%C-0.8%O)	0.5	4.1	1.2
IN9052 (Al-4.0%Mg-1.1%C-0.8%O)	0.5	4.1	1.2
IN905XL (Al-4.0%Mg-1.5%Li-1.2%C-0.4%O)	0.4	4.5	1.2

To demonstrate further the similarity of the deformation behavior, a comparison of the strain rates and flow stresses for β Si_3N_4/2124, β SiC/2124, and β SiC/6061 has been made based on the data set given in Tables 10 and 11. It is evident that the data from the two 2124Al metal-matrix composites not only agree qualitatively, but also the absolute values of stress and strain rate are consistent. The similarity of the two sets of data suggests that HSRS in these two composites is dominated by the behavior of the alloy matrix 2124Al and the reinforcement-matrix interfaces. Because the non reinforced 2124Al, 7064Al, and 6061Al matrix alloys were not superplastic at high strain rates, the above results indicate that the alloy matrix itself is not solely responsible for determining whether or not a composite is superplastic at high strain rates. The reinforcement-matrix interfaces play a more important role.

As is the case for SiC/Al composites, only a limited amount of microstructural information is available for the Si_3N_4/Al composite systems. High-strain-rate deformation of the Si_3N_4/Al composites results in a tangled dislocation network, whereas low-strain-rate deformation results in a low dislocation density [47]. The mechanistic implications of this observation are not yet obvious. Notably, the fracture surface of superplastic β Si_3N_4/2124Al is remarkably similar to the fracture surface of the superplastic β SiC/2124Al, suggesting that a similar deformation mode operates in both materials.

5.3 Mechanically Alloyed Al Composites

Several commercial mechanically alloyed (MA) materials have been observed to exhibit HSRS, including the Al-based alloys IN9021 [64], IN90211 (a modification of IN9021) [8], IN9052 [28], IN905XL [9], and MA SiC_p/IN9021 [29] and MA SiC_p/7075Al [55] composites. The MA materials usually contain oxide and carbide dispersions approximately 30 nm in diameter that have an interparticle spacing of about 60 nm. The grain size, chemical compositions, and volume fraction of dispersoids in IN9021, IN9052, and IN905XL are summarized in Table 12.

TEM investigations on the microstructures of mechanically alloyed Al composites gave evidence that their fineness and complexity of the microstructures is a prerequisite for their pronounced superplastic behavior at very high strain rates.

REFERENCES

1. *Aluminum Standards and Data.* 1984, Aluminum Association, Inc.
2. G. S. Ansell, T. D. Cooper and F. V. Lenel, "Oxide Dispersion Strengthening." *AIME Conference,* Vol. 47, 1966, Gordon and Breach, New York, London, Paris.

3. V. Arnhold and J. Baumgarten, "Dispersion Strengthened Aluminum Extrusions." *Powder Met. Int., 1985, 17(4), p. 168.*

4. M. Ashby, *Z. Metallk, 1964,* 55, p. 5.

5. J. S. Benjamin and M. J. Bomford, "Dispersion Strengthened Aluminum made by Mechanical Alloying". *Metall. Trans.* 1977, 8A, p. 1301.

6. H. W. Bergmann, "Mechanische Eigenschaften und Verformungsverhalten einer metallischen Matrix mit nichtmetallischen anorganischen Zusätzen". *Journal of Materials Technology*, 1976, 7(4), pp. 129–135.

7. H. W. Bergmann and G. Wassermann, *Z. Metallk.* 1978, 69, p. 126.

8. T. R. Bieler, T. G. Nieh, J. Wadsworth and A. K. Muhkerjee, "Superplastic-Like Behavior at High Strain Rates in a Mechanically Alloyed Aluminum." *Scr. Metall.*, 1988, 22, pp. 81–86.

9. E. A. Bloch, "Dispersion Strengthened Aluminum Alloys". *Metall. Rev.* 1961, 6(22), p. 193.

10. F. Boland, C. Colin, C. Salmon and F. Delannay, "Tensile Flow Properties of Al-Based Matrix Composites Reinforced with a Random Planar Network of Continuous Metallic Fibres". Acta Metall Mater. 1998, 46, pp. 6311–6323.

11. M. J. Bomford and J. S. Benjamin, Mechanically Alloyed Aluminum-Aluminum Oxide. 1974, U.S. Patent 3,816,080.

12. J. Broutmann and R. H. Krock, *Modern Composite Materials*, Vol. 1–5, 1967, Addison-Wesley, Reading, MA.

13. J. Cadek and V. Sustek, "Comment on 'Steady State Creep Behavior of Silicon Carbide Particulate Reinforced Aluminum Composites.'" *Scr. Metall. Mater.* 1994, 30, pp. 277–282.

14. R. M. Christensen, *Mechanics of Composite Materials*, 1979, Wiley, New York.

15. D. Cratchley, *Powder Metallurg.* 1963, 11, p. 59.

16. G. Frommeyer and G. Wassermann, *Acta Metall.* 1975, 23, p. 619.

17. G. Frommeyer, K. van Oldenburg and St. Beer, "Microstructure and Mechanical Properties of Mechanically Alloyed Intermetallic Mg$_2$Si-Al Alloys. *Z. Metallkde.* 1994, 5, pp. 372–377.

18. N. Furushiro, S. Hori and Y. Miyake, "High Strain Rate Superplasticity and its Deformation Mechanism in Aluminum Alloys." in *International Conference on Superplasticity in Advanced Materials (ICSAM-91)*, (S. Hori, M. Tokizane and N. Furushiro, eds), 1991, The Japan Society for Research on Superplasticity, Osaka, Japan, pp. 557–562.

19. P. S. Gilman, "The Physical Metallurgy of Mechanically Alloyed Dispersion Strengthened Al-Li-Mg and Al-Li-Cu Alloys." *Aluminum-Lithium Alloys II* , (T. H. Sanders and E. A. Starke, eds), 1984, Monterey, CA, April 1983, p. 485.

20. C. G. Goetzel, Dispersion Strengthened Alloys: "Properties and Applications of Light Metals." *J. Met.* 1959, 11, p. 276.

21. C. G. Goetzel, Dispersion Strengthened Alloys: "The Possibilities for Light Metals." *J. Met.* 1959, 11, p. 189.

22. G. Gonzales-Doncel and O. D. Sherby, "High Temperature Creep Behavior of Metal Matrix Aluminum-SiC Composites." *Acta Metall. Mater.*, 1993, 41, pp. 2797–2805.

23. G. Gonzales-Doncel, S. D. Karmarkar, A. P. Divecha and O. D. Sherby, "Influence of Anisotropic Distribution of Whiskers on the Superplastic Behavior of Aluminum in a Back-Extruded 6061 Al-20% SiC$_w$ Composite." *Comp. Sci. Technol.*, 1989, 35, pp. 105–120.

24. G. Gregory and N. J. Grant, "High Temperature Strength of Wrought Aluminum Powder Products." *Trans. AIME 200*, 1954, p. 247.

25. D. Gualandi and P. Jehenson, "Physical Metallurgy of Al-Al$_2$O$_3$ Composites (SAP) for Nuclear Applications." *Mod. Dev. Powder Metall.* 1966, 3, p. 36.

26. N. Hansen, "Dispersion Strengthened Aluminum Powder Products for Nuclear Applications." *Powder Metall,* 1967, 10(20), p. 94.

27. R. W. Hertzberg, F. D. Lemkey and J. A. Ford, *Trans. AIME 233*, 1965, p. 342.

28. K. Higashi, T. Okada, T. Mukai and S. Tanimura, "Positive Exponent Strain-Rate Superplasticity in Mechanically Alloyed Aluminum IN 9021." *Scr. Metall. Mater.*, 1991, 25, pp. 2503–2506.

29. K. Higashi, T. Okada, T. Mukai, S. Tanimura, T. G. Nieh and J. Wadsworth, "Superplastic Behavior in a Mechanically-Alloyed Aluminum Composite Reinforced with SiC Particulates." *Scr. Metall. Mater.*, 1992, 26(2), pp. 185–190 and pp. 761–766.

30. X. Huang, Q. Liu, C. Yao and M. Yao, "Superplasticity in a SiC_w-6061 Al Composite." *J. Mater. Sci. Lett.*, 1991, 10, pp. 964–966.

31. T. Imai, G. L'Esperance and B. D. Hong, "High Strain Rate Superplasticity of AIN Particulate Reinforced Aluminum Alloy Composites." *Scr. Metall. Mater.*, 1994, 31, pp. 321–326.

32. T. Imai, M. Mabuchi and Y. Tozowa, "Superplasticity in Si_3Ni_4 Whisker-Reinforced 7064 Al Alloy Composite." in *International Conference on Superplasticity in Advanced Materials (ICSAM-91)*, (S. Hori, M. Tokizane and N. Furushiro, eds.), 1991, The Japan Society for Research on Superplasticity, Osaka, Japan.

33. T. Imai, M. Mabuchi, Y. Tozowa and M. Yamada, "Superplasticity in β-Silicon Nitride Whisker-Reinforced 2124 Al Composite." *J. Mater. Sci. Lett.*, 1990, 9, pp. 225–257.

34. G. Jangg, "Powder Forging of Dispersion Strengthened Aluminum Alloys." *Metal Powder Rep.* 1980, 35(5), p. 206.

35. G. Jangg, F. Kutner and G. Korb, "Production and Properties of Dispersion-Hardened Aluminum." *Aluminum* 1975, 51, p. 641.

36. G. Jangg, F. Kutner and G. Korb, "Dispersion Hardening of Aluminum with Al_4C_3." *Powder Met. Int.* 1977, 9(1), p. 24.

37. R. H. Johnson, "Superplasticity." *Metall. Rev.* 1970, 15, pp. 115–134.

38. S. K. Kang, "The Microstructure of Mechanically Alloyed Aluminum Alloys." Presentation at ASM Conference, Philadelphia, October 2–6, 1983.

39. A. Kelly and G. J. Davis, *Metallurg. Rev.* 1965, 10, p. 1.

40. T. J. Koppenaal and N. M. Parikh, *Trans. AIME 224*, 1962, p. 1173.

41. K. G. Kreider, (ed.) *Metallic Matrix Composites*, 1974, Academic Press, New York.

42. R. H. Krock, in *Plansee Proceedings 1964—Metals for the Space age*, (F. Benesovsky, ed.), 1965, Metallwerk Plansee AG, Reuttle/Tyrol, p. 256.

43. R. H. Krock, in *Modern Developments in Powder Metallurgy*, Vol. 3 (H. Hausner, ed), 1966, Plenum, New York, p. 105.

44. R. H. Krock and L. A. Shepard, *Trans. AIME 227*, 1963, p. 1127.

45. E. Kubasta, "Deformation Behavior of Dispersion Hardened Al-C Alloys." *Strength of Metals and Alloys*, 5th International Conference, Munich, Vol 3, 1979, p. 1687.

46. W. Kurz and P. R. Sahm, *Gerichtet erstarrte eutektische Werkstoffe*, 1975, Springer, Berlin, Heidelberg, New York.

47. G. L' Esperance and T. Imai, "Superplastic Behavior and Microstructural Characterization of a β-Si_3N_4 Whisker Reinforced 2124 Aluminum Composite." in *International Conference on Superplasticity in Advanced Materials (ICSAM-91)*, (S. Hori, M. Tokizane and N. Furushiro, eds), 1991, The Japan Society for Research on Superplasticity, Osaka, Japan, pp. 379–384.

48. F. D. Lemkey, R. W. Hertzberg and J. A. Ford, *Trans. AIME* 233, 1965, p. 636.

49. F. V. Lenel, *Powder Metallurgy (Principles and Applications)* 1980, Metal Powder Industries Federation, Princeton, NJ.

50. R. W. Lund and W. D. Nix, "High Temperature Creep of Ni-20Cr-$2ThO_2$ Single Crystal." *Acta Metall.*, 1976, 24, pp. 469–481.

51. J. P. Lyle, "Aluminum Powder Metallurgy Products." *Proceedings, Metal Powder Association*, 1956, p. 93.

52. M. Mabuchi and T. Imai, "Superplasticity of Si_3N_4 Whisker Reinforced 6061 Al Composite at High Strain Rate." *J. Mater. Sci. Lett.* 1990, 9. pp. 763–765.

53. M. Mabuchi, T. Imai and K. Higashi, "Production of Superplastic Aluminum Composites Reinforced with Si_3N_4 by Powder Metallurgy." *J. Mater. Sci.*, 1993, 28, pp. 6582–6586.

54. M. W. Mahoney and A. K. Ghosh, "Silicon Carbide Reinforced Aluminum Structure." *Air Force Contract*, AFWAL-TR-82-3051. 1982, Rockwell International.

55. K. Matsuki, H. Matsumoto, M. Tokizawa, N. Takatsuji, M. Isogai and Y. Murakami, "Effect of SiC Particulate Content on Stress-Strain Rate Relationship during Hot Compression of SiC_p/7075 Al Composites." *J. JIM*, 1993, 8, pp. 876–883.

56. *Military Standardization Handbook*. 1976, MTL-HDBK-5C.

57. R. S. Mishra and A. B. Pandey, "Some observations of the High-Temperature Creep Behavior of 6061 Al-SiC Composites." *Metall. Trans.*, 1990, 21A, pp. 2089–2090.

58. *Modern Developments in Powder Metallurgy*, Vol. 1–3, (H. H. Hausner, ed), Plenum, New York.

59. Y. Murakami, in "Mechanical Properties of Metallic Composites", (S. Ochiai, ed), 1994, p. 382.

60. G. H. Narayanan, B. L. Wilson and W. E. Quist, "P/M Aluminum-Lithium Alloys by the Mechanical Alloying Process." *Aluminum-Lithium Alloys II* , (T. H. Sanders and E. A. Starke, eds), 1984, Monterey, CA, April 1983, p. 517.

61. H. C. Neubing, "Production and Properties of Aluminum Powder for Powder Metallurgy." *Powder Met. Int.* 1981, 13(2), p. 74.

62. P. Neumann and P. Haasen, *Phil. Mag.* 1971, 23, p. 285.

63. T. G. Nieh and J. Wadsworth, "Effects of Zr on the High Strain Rate Superplasticity of 2124 Al." *Scr. Metall. Mater.*, 1993, 28, pp. 1119–1124.

64. T. G. Nieh, P. S. Gilman and J. Wadsworth, "Extended Ductility at High Strain Rates in a Mechanically Alloyed Aluminum Alloy." *Scripta Metall.*, 1985, 19, pp. 1375–1378.

65. T. G. Nieh, C. A. Henshall and J. Wadsworth, "Superplasticity at High Strain Rate in SiC-2124 Al Composite." *Scr. Metall.* 1984, 18, pp. 1405–1408.

66. T. G. Nieh, J. Wadsworth and D. J. Chellman, "In Situ Auger Fractographic Study of a SiC-Reinforced Al Alloy", *Scripta Metall.*, 1985, 19, pp. 181–184.

67. T. G. Nieh, J. Wadsworth and O. D. Sherby, *Superplasticity in Metals and Ceramics*. 1997, Cambridge, p. 28ff.

68. A. B. Pandey, R. S. Mishra and Y. R. Mahajan, "Steady State Creep Behavior of Silicon Carbide Particulate Reinforced Aluminum Composites." *Acta Metall. Mater.,* 1992, 40, pp. 2045–2052.

69. J. Pilling, "Superplasticity in Aluminum Base Metal Matrix Composites." *Scr. Metall.* 1989, 23, pp. 1375–1380.

70. J. P. Poirier, "On Transformation Plasticity", *J. Geophys. Res.*, 1982, 87, pp. 6791–6797.

71. S. W. Porembka, "Progress on USAEC/AECL Cooperative Program." Report No. BMI—X-266, Battelle Memorial Institute, AEC Contract No W-7405-eng-92. 1963.

72. S.A.P. *Techn. Info. Bulletin AIAG*, 1962, Neuhausen, Switzerland.

73. A. Sakamoto, C. Fujiwara and T. Tsuzuku, "Development of Fabrication Process for High Strength Metal Matrix Composites." in *Proceedings of the 33rd Japan Congress on Materials Research*, 1990, The Society of Materials Science, Japan, pp. 73–79.

74. R. D. Schelleng and S. J. Donachie, "Mechanically Alloyed Aluminum." *Metal Powder Rep*. 1983, 38(10), p. 575.

75. R. D. Schelleng, P. S. Gilman and S. J. Donachie, ''Aluminum-Magnesium-Lithium-Forging Alloy made by Mechanical Alloying.'' *Overcoming Materials Boundaries*, 1985, NSTC 17, SAMPE, Covina, CA, p. 106.

76. A. Seeger, D. Wolf and H. Mehrer, ''Analysis of Tracer and Nuclear Magnetic Resonance Measurement of Self-Diffusion in Aluminum.'' *Phys. Status Solidi*, 1971, 48, pp. 481–486.

77. R. F. Singer, W. C. Oliver and W. D. Nix, ''Identification of Dispersoid Phases created in Aluminum during Mechanical Alloying.'' *Metall. Trans.* 1980, 11A: 1895.

78. M. Stern, ''Mechanical Ingoting of Aluminum and Magnesium Turnings.'' *Iron Age*, 12. August, 1943, 152, p. 90.

79. K. R. VanHorn, ''Properties , Physical Metallurgy, and Phase Diagrams.'' *Aluminum,* Vol. 1. *1967, American Society for Metals*, p. 347.

80. J. Wadsworth, C. A. Henshall, T. G. Nieh, A. R. Pelton and P. S. Gilman, ''Extended Ductility in Whisker Reinforced Aluminum Alloys and Mechanically Alloyed Aluminum Alloys.'' in *High Strength Powder Metallurgy Aluminum Alloys II*, (G. J. Hildeman and M. J. Koczak, eds), 1986, TMS-AIME, Warrendale, PA, pp. 137–154.

81. Z. Wei, B. Zhang and Y. Wang, ''Microstructure and Superplasticity in a Stir-Cast SiC_p/2024 Aluminum Composite'', *Scr. Metall. Mater.*, 1994, 30, pp. 1367–1372.

82. M. Y. Wu, J. Wadsworth and O. D. Sherby, ''Superplasticity in a Silicon Carbide Whisker Reinforced Aluminum Alloy.'' *Scr, Metall.*, 1984, 18, pp. 773–776.

18

Environmental and Toxicological Effects

GILBERT F. BOURCIER

Old Dominion Engineering Services Company, Midlothian, Virginia, U.S.A.

1 INTRODUCTION

This chapter will cover the major environmental and toxicological effects resulting from the production of aluminum. The coverage extends from the exploration of mining sites for bauxite to the recycling of the various products after completion of their useful life. There will necessarily be some aspects of safe practices and safety included, since many of these are either the result of US Federal and State Environmental regulations or Health and Safety regulations addressed by the US Occupational Safety and Health Administration (OSHA).

The author will not directly address the secondary effects of aluminum production, such as coal mining, production of oil or natural gas, electric power generation, or peripheral services to operating facilities. If significant, some discussion of these impacts will be presented, with more detailed information available through references or recommended reading from the listed titles in Section 16.

Since the environmentally related regulations are not the same in all countries, there will be a discussion of Best Management Practices and references to European Union and/or ISO 14000 standards to identify areas of potential concern. The reader may determine the degree of significance to applicable site-specific regulations. However, US regulations will be featured, and multi-jurisdictional applicability will be discussed, if known.

Where country-specific information is desired, contact with the appropriate embassy is a good starting point. Alternatively, contact World Bank representatives, since it is often involved in the financing of major activities, particularly in developing countries. Both contacts should be in the Washington D.C. US or in the major seats of government and/or financial centers.

United States environmental laws, with some summaries or abstracts, and their significance, are included in Section 14. It is important to know what materials are used, processed or stored at a given site and if they are subject to any governmental regulations. If so, applicable regulation(s) should be identified and the effects of that regulation determined. Materials leaving a site are usually in a solid, liquid or gaseous form. These materials must be identified and quantified to determine their regulatory status.

The US Environmental Protection Agency (US EPA) has listed a hierarchy of waste management activities, starting with Pollution Prevention, and continuing with Process Changes, Reuse/Recycling and ending with Landfill as the last, and least favored option. Pollution Prevention may be an economical solution to a number of site or facility related environmental problems, given the costs associated with conformance to the increasingly complex regulations and the penalties for noncompliance.

In increasing instances of regulated activity, monitoring to verify conformance to the specific emission limits is integral to the regulation, as is a mandatory requirement for reporting to the regulatory authorities. There are a number of published standards relating to monitoring and testing of the environmental emissions and conditions. These have been developed and published by ASTM, and are readily available. In part, because of international participation in the writing of these standards, ASTM standards reflect the input of many of the criteria relating to the worldwide environmental and regulatory climate (refer to Section 13).

Additionally, ASTM (Committee D-34, on Waste Management) is working closely with EPA on continuously updating an extensive group of analytical standards, published by EPA in SW 846.

The individual stages of aluminum production, will be presented as sections containing production-specific environmental and toxicological information. General environmentally related regulations that are applicable to all process steps, such as operating permits under the Clean Air Act, will be covered later in other chapters.

2   EXPLORATION AND MINING

One of the first considerations in aluminum production is to assure a defined source of raw materials. The major raw material is bauxite for the production of alumina, which is then used to produce primary aluminum. Scrap aluminum is for use in the production of secondary aluminum. Scrap usage will be addressed later in this chapter.

Large quantities of other raw materials that are required include limestone and caustic soda for alumina production and electric power, carbon, fluoride salts (e.g. cryolite), and refractories for the production of aluminum. Shipping requirements for the large quantities of raw materials to be processed require the capability to receive ocean-going vessels or railroad freight service. Building of either new or continuing use of existing shipping facilities involves control of dust emissions, maintenance of fuel storage tanks or bunkers, control of rainfall runoff, disposal of ships ballast and sanitary wastes and the impact of noise and facility emissions on the nearby community.

Each of these various raw material sources involves a number of environmental considerations. In the commercialization of these sources, the environmental aspects in new developments or utilizing existing sources are interrelated with the principles of workplace safety. This means adequate resources should be devoted to the minimization of personnel exposure to physical hazards or toxic substances. Personal protective equipment should be utilized, as required. References to the OSHA regulations involving health and safety are included in Section 14.

2.1 Bauxite Sources

Aluminum is the third most common element in the earth's crust (about 7.3 wt%), with only oxygen and silicon found in greater abundance. However, most of the aluminum found in the crust is not amenable for transformation into alumina, and subsequently, aluminum, due to technologic and/or economic reasons. Bauxite is the most common raw material for the production of alumina. While bauxite is globally abundant, a deposit's quality, logistical setting and geographical location are major factors in determining the economic feasibility of development.

Bauxite is formed from the weathering of aluminous protoliths. It contains hydrates of aluminum and iron, oxides of silicon and titanium as well as trace amounts of other elements, depending on the composition of the parent rock and the degree of weathering. The alumina in the bauxite is a combination of monohydrate ($Al_2O_3 \cdot H_2O$-Boehmite or Diaspore) and trihydrate ($Al_2O_3 \cdot 3H_2O$-Gibbsite) minerals. Usually the bauxite either contains within it or is underlain by clays (hydrous aluminum silicates) that are not desirable feedstocks for the Bayer process.

Since bauxite is a product of intense weathering, deposits are globally distributed in tropical to subtropical areas or areas that have been exposed to these conditions in the geologic past. Further, the deposits were formed in cratonic areas with little tectonic activity, which would have eroded these deposits. Bauxite deposits are found near the surface or may be buried under a certain amount of other material (overburden) depending on the geologic events occurring after their formation (personal communication—S. T. Angely, 1998).

Alumina has been produced from other natural resources, such as clays, alunite, anorthosite, and nepheline syenite. However, the processing of these raw materials are far less economic for the production of alumina than bauxite [1, 2].

Some researchers have experimented with recovery of alumina from fly ash that had been captured from coal burning power station boilers. Such processing has been done at the laboratory bench or as a small scale pilot operation. However, the need for a sintering step increases the energy inputs, indicating this method would be more costly and complex than bauxite processing [3].

2.2 Site Impact

The exploration and development of a bauxite deposit has an impact on the site and to the access routes. During exploration to locate and evaluate a deposit, the impacts are usually minimal. Access roads and camp facilities are usually rudimentary. The deposit is evaluated by drilling, pitting, or trenching on various grid spacings, but the actual surface area disturbed is usually very small in comparison to the area

being evaluated. The disposal of refuse from humans and the use of machinery, and impacts on water quality are the main environmental concerns at this juncture.

During mining, the impact on the site is more extensive. Developing the orebody requires more substantial access routes, land clearing, the excavation and stockpiling of topsoil, overburden and ore. Also, buildings for housing, power generation, shops, processing facilities and shipping facilities are constructed. The major environmental concerns now are with regards to soil erosion, siltation of waterways, water quality, dust and the disposal of industrial wastes.

In remote areas, tropical diseases, insects, animals etc., can present a health hazard to the workforce.

Many governmental jurisdictions require that a detailed environmental impact documentation of the proposed project, along with site restoration plans, be submitted for approval prior to the expected starting date of development.

Bauxite deposits can be found near the surface or they may be buried by varying amounts of overburden, sometimes as far as 50 or more below the surface. Mining is generally conducted by open pit methods. In some locales, underground mining is carried out on deep ore bodies, where either economics or the characteristics of the ore make this type of mining commercially feasible.

Topsoil and overburden are stockpiled for future site restoration. In many cases today, restoration begins when mining is completed on part of the deposit. This helps to minimize the surface area exposed to accelerated erosion and more rapidly brings the mined out area into a new and productive state.

The International Primary Aluminum Institute (IPAI) has developed a data base on the mining of bauxite and site restoration activities. The preliminary survey showed reforestation or agricultural use for more than 90% of the restored mine area, which is apparently the best possible use for these sites at this time [4].

2.3   Dust

Because most mines are in remote locations, any particulates carried by wind normally settle harmlessly in nearby areas and cause no problems. However, the unloading or storage of some grades of bauxite at the alumina refinery could result in the generation of particulate matter, which could be transported by surface winds to adjacent property. The impact of this particulate material on the facility personnel, or the downwind jurisdictions, depending on whether there are other mineral species present, could create a problem. The containment of particulates, depending on their regulatory status, may require a costly control measure.

In mid-1997 the US EPA proposed regulations on particulate matter, specifying the control of emissions of particulate matter in size ranges down to 2.5 microns ($PM_{2.5}$). Major concerns in the aluminum industry are the technology required and the associated cost of meeting these limits. There is even concern about the ability to accurately monitor the generation or emissions of particulates this small. This regulation, however it appears in final form, is expected to be applicable to all phases of aluminum production. Other sections, or titles, of the Clean air Act Amendments of 1990 are applicable to the various types of facilities in the aluminum industry and are phased in over a scheduled implementation program. Actual dates for these facility program starting dates are available through EPA and Federal Register sources.

2.4 Mine Water and Stockpiled Overburden Runoff

Materials exposed during mining may contain minerals or compounds that could react with atmospheric oxygen, bacteria, moisture, light, heat etc., to form acidic mine water. Facility personnel in contact with some these effluent compounds may require protective clothing or respiratory protection. Unauthorized discharge of some of these effluents into waterways or wetlands, whether inadvertent or by intent, could result in legal action being taken by the affected governmental authority. For this reason, treatment or management of all discharges is very important.

Silt traps are often used to control siltation. Water treatment facilities are often constructed to handle water from the industrial site and to treat sewage generated by the workforce.

Control of water discharges is an individual case consideration, based on what the monitoring of the water indicates and what the permit requires. When mining operations are completed, post-closure monitoring and water treatment can be an expensive, long-term activity. (C. McIlwaine, personal communication, 1995.)

During the mining operation, the stockpiled overburden will usually require collection and treatment of the rainfall that runs off this material. Fortunately, this runoff can use the same treatment process steps that the mine water receives, as it is representative of the sidewall strata uncovered. Silt curtains would be a prudent, and often required control system for minimizing the transport of some of the potentially toxic materials present to downstream locations. In some jurisdictions these silt curtains are mandatory.

2.5 Discussion

No matter which mineral species is to be recovered or the location of the source, certain constraints apply to the mining operations. The exploration, core drilling and other tasks necessary to locate and confirm an ore body involve some impact on the site as well as on access routes. These impacts can be as little as the proper location of the residues of the drilling or as large as the building of access roads and shipping facilities, vegetation and tree removal, overburden storage, soil erosion and siltation of waterways. Many governmental jurisdictions require that a detailed environmental impact documentation of the proposed project, along with site restoration plans, be submitted for approval prior to the expected starting date.

3 ALUMINA PRODUCTION – BAYER PROCESS

The production of one metric ton of alumina requires about two to three metric tons of bauxite, depending on the percentage of available alumina present. The bauxite is normally crushed at the mine site with primary, and possibly secondary, crushers. At the refineries, the bauxite is ground with concentrated caustic soda to a fine consistency (100 mesh or finer) in rod or ball mills.

To prevent subsequent scaling of heat exchanger surfaces, the ground bauxite slurry is normally desilicated in atmospheric reactors before entering the high temperature digestion section. The caustic concentration and temperature used

for digestion are dependent on the type of alumina bearing compounds in the bauxite. For gibbsitic bauxite, a low temperature (130–150°C), low caustic concentration (150–160 gpl NaOH, as Na_2CO_3, is used. For boehmitic and gibbsitic bauxite, a higher temperature (200–250°C), medium to high caustic concentration (160–380 gpl NaOH, as Na_2CO_3, is used. For diaspore, with boehmite and gibbsite, a high temperature (230–280°C), high concentration (230–380 gpl) NaOH, as Na_2CO_3, is used [5].

In the digestion system, the ground, desilicated bauxite slurry and concentrated sodium aluminate liquor streams are heated countercurrently in a series of heat exchangers before entering continuous digester vessels. The digested slurry is flashed to atmospheric pressure in a series of flash tanks to recover heat. The flash vapor is used as the heating medium for the bauxite slurry and caustic liquor streams en route to the digester vessels. Live steam is normally used in the last heater stages or steam may be added directly to the digester vessels. Some refineries employ tubular digestion systems, with other heat transfer media, for this operation.

The caustic soda dissolves the aluminum compounds in the bauxite, forming a solution of sodium aluminate. A portion of the caustic soda reacts to form sodium alumino silicate compounds, normally referred to as desilication product (DSP), which lowers the extraction efficiency. The iron oxides, titanium compounds and other undissolved materials form an insoluble sludge, which is separated in settling tanks. The settler underflow slurry is washed in a countercurrent decantation system to recover more of the caustic soda and dissolved alumina. The countercurrent washing system consists of thickeners in series, sometimes with a final filtration stage. Normally, the overflow stream from the first stage of the countercurrent washing system is causticized with slaked lime in a separate reactor. This recovers the active solvent, sodium hydroxide, from dissolved sodium carbonate, which forms during the process from organic compounds in the bauxite or from carbon dioxide in the atmosphere.

The sodium aluminate liquor stream overflowing the settlers is filtered and cooled en route to precipitation. This supersaturated sodium aluminate solution (''green'' liquor) is ''seeded'' with fine sized aluminum hydroxide crystals to precipitate the aluminum hydroxide. Older plants use batch precipitation tanks, while the more modern facilities use a continuous precipitation process. The continuous process consists of agglomeration and growth stages, often with intermediate flash cooling to maximize yield and heat recovery. The precipitated aluminum hydroxide slurry is classified into product and seed fractions. The product fraction is washed, usually on pan or belt filters, prior to calcination. The mother, or ''spent'' liquor from the precipitation and clarification process is used to cool the green liquor in countercurrent heat exchangers and then concentrated in evaporators before recycling to digestion.

The filtered and washed aluminum hydroxide is then calcined at 1000°C to 1200°C to remove residual free moisture and water of hydration, forming aluminum oxide (alumina), a white granular material. Older plants use rotary kilns, while modern facilities use fluid bed or flash calcination. The calcined alumina product is shipped to smelters for electrolytic reduction to aluminum metal. In some refineries, a portion of the aluminum hydroxide is washed and dried for sale as chemical grade aluminum trihydrate. Ceramic grade alumina products are also produced at some refineries by using special calcining conditions.

3.1 Bauxite Residue

The insoluble sludge, or bauxite residue (normally referred to as "red mud"), after washing to recover reagents, is pumped as an alkaline slurry to an impoundment. In a number of refineries, the supernatant liquid is recycled as process water for mud washing. In the instances where there is an excess of evaporation loss relative to rainfall, the top layer of the bauxite residue may dry to the point where windblown dust must be controlled. The surface layers of such air dried impoundments, up to several centimeters deep, may typically contain about 90% solids. However, the lower layers usually still contain a substantial quantity of liquid. The solids content of bauxite residues can vary widely between facilities, ranging from 25% in the case of some US refineries, to the range of 60% for Western Australian plants. The residual liquor portion of the slurry contains sodium carbonate and sodium aluminate, along with other sodium salts that were not recovered in the counter-current mud washing system.

In the case where there is an excess of rainfall relative to the rate of evaporation, there is a potential for overflow from the impoundment. Under these conditions, water management becomes very important. The bauxite residue impoundment should be designed to accommodate the planned residue storage quantity, while minimizing surface area in order to mitigate the quantity of collected rainfall. Sufficient operating freeboard in the impoundment or the associated process water reservoirs should be provided to accommodate reasonably high rainfall events on the refining process area. Preparation of the impoundment area for either rainfall condition, as well as for the unusual rainfall occurrences, can be a low cost method of pollution prevention.

An example of the range of compositions of the residue solids from the processing of Caribbean and South American bauxites is as follows: 35–40% Fe_2O_3, 15–20% Al_2O_3, 7–9% SiO_2, 6–7% TiO_2, 5–6% CaO, 4–6% Na_2O; with the balance water of hydration and minor quantities of other minerals. These compositions typify analyses over an extended time frame (a number of years) and are reported as metallic oxide equivalents. Particle size is related to the criteria in the crushing and grinding circuit and can be as fine as 10–20 microns. The residual liquid associated with the bauxite is alkaline, with a pH that can exceed 12. (M. Chaubal, personal communication, 1991.) Western Australian bauxite residue solids can range much higher in Fe_2O_3 (upwards of 60%), with other elements in the area of 13–14% Al_2O_3, 4–6% SiO_2, 6% TiO_2 and 2–3% CaO, with the composition of other metallic oxide equivalents generally at a lower level.

One of the major environmental impacts of the alumina production process involves the management of the bauxite residue. When an impoundment has reached the physical capacity limits, various closure options may be considered. Current government regulations may require some type of capping and a vegetative cover. Capping may be necessary to minimize or eliminate water infiltration. The containment of alkaline run-off or leachates is one of the prime environmental management considerations.

Depending on the country and locale however, many companies are experimenting with alternative closure methods in which amendments would be added to the bauxite residue to promote vegetative growth. One of these experimental approaches used treated sewage sludge to cover the impoundment,

after which vegetation was planted. (T. Ballou, personal communication, 1996.) It now is growing into a ground cover that should substantially reduce, or even eliminate, the wind blown fine dust that has sometimes occurred on windy days.

A number of different experimental approaches have been used to evaluate recovering the metallic elements present in the bauxite residues. One example, from Bharat Aluminum Co., involved the pilot plant level recovery of titanium dioxide, from high titanium content (15–18% TiO_2) bauxite residues that contained a potentially economically recoverable quantity of this material [6]. The current relative cost of titanium bearing ores and the processes used for titania recovery make this processing option uneconomical for most bauxite residues, as TiO_2 content is relatively low.

Numerous researchers have also tried to recover iron from the iron-rich material, but the current relative cost of iron ore makes use of this concept also too costly.

Sintering the bauxite residues with limestone and caustic soda or soda ash can be used to recover additional sodium aluminate, if the processing economics are favorable. This sintering procedure was formerly practiced in the US and may still be used in some locations in the former Soviet Union when processing high silica domestic bauxites. Other researchers have evaluated processing high silica content bauxite with limestone and soda ash in a sintering operation, followed by a leach operation using a sodium carbonate solution [7,8]. The output of this leaching operation was a sodium aluminate solution and a di-calcium silicate residue (also known as brown mud or brown sand). The US systems formerly using this lime–soda sintering process found the fuel cost to be too high for economic competitiveness and have been discontinued.

3.2 Discussion

The production of alumina is a chemical process, using caustic soda, lime, water, steam, electricity and fuels. Certain of these inputs requires equipment controls and containment practices to minimize leakage that could contaminate adjacent soils and groundwater. Here, pollution prevention procedures can be substantially less costly than clean-ups. In the processing areas of the facility where large numbers of tanks containing caustic soda and sodium aluminate solutions are located, spill prevention and containment measures are practiced. These practices include padded and curbed areas, as well as storm water runoff controls. Good facility design must be capable of containing, recovering and recycling spills of process liquids from tanks or lines in the curbed areas. (In some site-specific cases, storm water monitoring and reporting are a requirement in the United States.)

Recovery and recycling of process side streams, such as the countercurrent mudwasher underflow liquid, involves the proper handling of materials with a relatively high pH. The benefits of reuse of expensive reagents and the elimination of waste streams make these procedures environmentally sound and economically advantageous.

The processing of bauxite requires use of a variety of fuels to provide a supply of process steam as well as for use in firing calcining units. Most plants employ cogeneration of power and steam. Boilers are typically fired with coal, oil or natural gas. Calcining equipment uses natural gas or oil. The air emissions, primarily

particulates, must be controlled in order to conform with applicable regulations. Reducing process energy use is a desireable cost saving and pollution prevention measure.

4 REDUCTION

The electrolytic reduction of alumina, with carbon as the reductant, has the potential to result in a number of major impacts on the environment. A brief description of the reduction process, the equipment and materials involved and the reactions taking place follows, along with a discussion of these environmental impacts and measures taken to minimize them. (D. Strahan, personal communication, 1996) [9].

4.1 Electrolytic Cells

The Hall–Heroult process for reducing the alumina to metallic aluminum in an electrolytic cell is used worldwide. Typically, the reduction cell dimensions, particularly the most current designs, are in the range of 10–15 m long and 5 m wide, with a depth of 1–2 m. All of the dimensions given may vary, depending on the design parameters of the specific facility.

The reduction cell is part of a "pot line" typically consisting of 100 or more of these reduction cells (pots), joined electrically in a series circuit. The cells operate at about 950°C, with an applied voltage in the range of 5 or less volts dc. Line amperages run in the range of under 100 kA and to more than 300 kA [10,11].

The normal cell design calls for a carbon cathode, both on the walls of the cell and on the bottom of the cell, where the metallic aluminum is collected as the alumina is reduced. The carbon cathode is typically calcined anthracite, which, when heated, gives off the volatile organic compounds (VOCs) peculiar to the temperature at that point, air infiltration and the binders used to hold the carbon in place as a monolithic structure. These binders can be derived from coal tar (pitch) and can contain hundreds of organic compounds that on heating are released as volatile organic compounds. Some of these are classified as hazardous air pollutants (HAPs) by the US EPA, and must be controlled.

Construction of these pots typically begins with the steel shell being lined with insulating refractories and prebaked carbon cathode blocks. A mixture of the calcined anthracite and a pitch binder is rammed into the bottom and the sidewalls of the pot, covering the refractory. An electrical connection to the bus bar is made from this cathode. The anode is installed in the upper section of the pot and is supported by appropriate structural elements. Power from the incoming bus bar is delivered to the anode through steel or iron pins that are cast into the prebake anodes, or in the case of Soderberg anodes, through steel or iron pins in the carbon/pitch paste that was poured into the cavity on the superstructure. This paste then moves into the cell, is baked in situ as the anode is consumed. The industry is moving rapidly into use of prebake pots to reduce the emissions of PAHs [12,14,15].

Additionally, the anode is also made from carbon, although the "prebake" design anode is produced by mixing carbon with a binder and baking at a high temperature to bind the carbon together, as well as to drive off most of the VOCs. During the baking process then, the prevention of release of the VOCs into the atmosphere is an operational requirement. The other anode variant, used in

"Soderberg" cells, uses a mixture of carbon and pitch binders that are fed as a viscous mixture into pots in the reduction cell, where they are baked "in situ" prior to contact with the pot bath. This "in situ" anode baking releases substantially more VOCs than the prebake anode, as one might expect. There are some advantages to the use of Soderberg reduction cells, although most new pot lines are of the prebake anode design. The carbon used for the anodes is typically from petroleum sources, and is produced in facilities designed for carbon production [12].

Typically, the layout of a pot line, which can consist of one or more pot rooms, is in a building with a high ceiling. This facilitates overhead cranes for installing new or rebuilt pots, moving pots, and for the conveying systems for delivery of alumina. This high ceiling also accommodates the increasingly complex air handling and emission control systems, but aids in the control of the relatively high ambient temperature resulting from the operating pots.

The alumina is dissolved in cryolite (sodium aluminum fluoride – Na_3AlF_6) called the pot bath. Other fluoride compounds, including aluminum fluoride (AlF_3), calcium fluoride (CaF_2), sodium fluoride (NaF) and lithium fluoride (LiF), are sometime added to the bath to achieve certain efficiency improvements and control predetermined operating criteria. The cryolite contains about 4–6% dissolved alumina, and is recharged with fresh alumina on a regular schedule [10,11].

4.2   Power

Electric power is brought in using a large bus conductor and is delivered to the cells through the carbon anodes, usually at several points in the upper half of the cell. As is common in facilities with the electrical bus bar installations located where they could be contacted by the operating personnel, employee safety is stressed through intensive training programs. There is an electromagnetic field generated in the reduction cells, so wristwatches and credit or other cards bearing a magnetic strip can be adversely affected by these fields and are best left outside the working area.

For those interested in the sources of electric power, the present average worldwide electrical power sources used by the aluminum industry for primary smelting has a component of 57% hydro, 33% coal, 5% nuclear, 4% gas and 1% oil [13]. This figure will vary between locations and the availability of these sources of power. The energy components will also vary for downstream operations, as these facilities generally depend on nearby sources of power and fuels. A number of hydro sources were made economically feasible in past years by the primary aluminum plants and other chemical process industry facilities. By providing a large base load for the power output of these dams, these industries subsequently enabled other consumers to enjoy the benefits of this low cost, nonpolluting power supply.

The chemistry of the reduction process is the reduction of alumina (aluminum oxide) with carbon, forming metallic aluminum and carbon dioxide. The alumina dissolved in the cryolite forms trivalent aluminum ions and bivalent oxygen ions. Under the impressed current the aluminum ions migrate to the cathode and are reduced to aluminum metal. The carbon in the anode reacts with the oxygen in the alumina to form CO_2, which is released from the pot. Some of the carbon dioxide is reduced to carbon monoxide by contact with the aluminum metal in the cell, which

is subsequently oxidized, forming alumina. This back reaction reduces cell efficiency and is constantly being monitored to keep it to a minimum. The carbon monoxide burns as it escapes the cell and could result in a health problem if not carefully controlled.

Theoretically, production of 1 k of aluminum requires 1.9 k of alumina, about $\frac{1}{3}$ k of carbon and just over 5 kW hours. Operating efficiencies change that to ½ k of carbon and about 6 kW hours in the most efficient pot lines. (D. Strahan, personal communication, 1996).

One of the major benefits of efficiency improvements is reduced emissions per kilo of aluminum produced, which, in today's regulatory climate could lead to reduced unit operating costs for the facility. Some of the major contributions to pollution prevention involve the improvements in operating efficiency in power utilization, which was in the 80–85% range in the mid-1940s and is now at about 92–95% in today's latest installations [13].

4.3 Environmental Impacts

Since the calcined anthracite used for lining the pots is typically bonded with a coal tar pitch (although a petroleum based pitch could be used), the 950°C temperature of the pot during operation, coupled with a varying quantity of air infiltration, can result in a number of undesirable PAHs being formed and released. As the pot ages and the more volatile emissions decrease as the result of continuous high temperature exposure, generation of these PAHs decreases [15].

The other particulate cell emissions of importance are alumina, and fluorides, in the form of cryolite. Gaseous emissions include carbon dioxide (CO_2), carbon monoxide (CO), carbon tetrafluoride (CF_4) and sulfur dioxide (SO_2) from pitch binders in the anode and cathode. Hydrogen fluoride (HF) and carbon hexafluoride (CF_6) are perfluorocarbon gases (PFCs) that result from particulate fluorides volatilized from the pot bath. While CO_2 has been implicated by its "global warming potential" as a greenhouse gas, the perfluorocarbon gases are much more aggressive in this respect. Fortunately, the perfluorocarbon gases can be minimized by careful control of the alumina content of the pot bath and the position of the anode relative to the cathode, pointing to a natural application of computerized control on the feed conveying and metering system. The US EPA and the domestic primary aluminum producers have a jointly established program (Voluntary Aluminum Industrial Partnership, or VAIP) to reduce the generation of the PFCs by using the best available control systems and bath chemistry on the pot lines [16,17].

The pot line emission control systems, by virtue of being required to control the worst case scenario, must be sized accordingly. Current technology for removal of these emissions includes the dry scrubbers, in which alumina is used to entrain particulate material, as well as to adsorb some gaseous emissions. This alumina is later recycled into the pots. Baghouses also play a part in capture of the particulate emissions. Here again is a waste material that may be classified as a hazardous waste were it to be discarded, based on the TCLP testing of the material. Simple economics dictates recycling this valuable source of alumina back into the pots [13].

Electrostatic precipitators, both wet and dry, are sometimes used to reduce the gaseous emissions from the pot lines. Since the captured material is from the reduction cells, this recovered material is recycled into the pots.

4.4 Applicable Regulations

The regulations that are most prominent in the reduction facility picture are those dealing with air and airborne emissions. Refer to Section 14 on Environmental Regulations for more information on the applicable regulations for air, water and solid waste discharges or emissions.

5 CASTING, ROLLING AND EXTRUSION

The generally used mode for transport of primary aluminum from a reduction plant is by rail, although truck and barge transport is used in some locations. The form in which primary aluminum is shipped is usually a sow of 500–1000 k, although T-ingots of varying sizes are sometimes produced and shipped. The casting facilities in reduction plants are usually capable of producing sheet ingot or extrusion billets as well as primary ingot for foundries or other converters. These plants are normally casting as much metal as possible into the form required by fabricating plants and eliminating the environmental impacts of remelting sows or T-ingot.

5.1 Casting

The initial operation in the conversion of primary ingot into product is to calculate the remelt furnace charge. The conventional furnace is an open hearth unit, generally rectangular, and fired with gas or oil burners. However, circular furnaces are in use in some facilities. Improved fuel efficiency and better refractory life are among the expected benefits of a circular furnace. The burners fire into the furnace chamber, heating the roof refractory to temperatures in the range of 1700°C. The radiant heat from this refractory and the radiant heat from the burner flame then heat the bath. The burners convert natural gas or #2 oil into heat, with CO_2 and water vapor as co-products of combustion. Other substances in the fuel, whether it be gas or oil, are oxidized. Where sulfur is present, SO_2 is generated. In the case of fuel oils, particularly from some western hemisphere sources, the presence of vanadium or other metals can result in an oxide of that metal being carried out in the flue gases or being incorporated into the molten bath. If there is a measureable quantity of heavy metals in the furnace emissions, these must be included in the facility emissions inventory and reported as such.

There is considerable interest in oxygen enriched burner mixtures, which have the effect of producing fewer products of combustion, but a much hotter flame than a conventional burner can provide. The elimination of some NO_x from the system has a positive impact on reducing the emissions and their effect on facility permitting [18].

For most casting operations, the furnace charge also contains in-plant run-around scrap, or home scrap, that comes from plant fabrication operations. Purchased scrap, whether from "old" or post-consumer sources or "new" scrap from outside fabricators, is also a common charge component. These scraps can contain contaminants that range from paper to paint, from oils to other metals or even plastics. Some of the older plastics contain lead additives. Some paints formerly used as primers contained cadmium compounds, which can remain in the dross or in the bag house dust, and may exceed regulatory limits in some cases. When these scraps are charged into a furnace and melted, the non-melting constituents

add to the furnace emissions and in cases where large quantities are charged, the products of melting and combustion should be accounted for in the plant's emissions inventory. There are empirical formulae used to calculate the emissions from the combustion of most fuels, and other materials as well, all of which may be generally applicable to parameters specified in operating permits.

In the case of adding alloying elements and grain refiners, some can contain beryllium, cadmium, chromium, or other heavy metals. Some volatilization of these metals could be expected, so personnel protection is advised. Some of these metals may not be completely taken into solution, leaving the potential for residuals entering into the dross or being absorbed in the refractory walls. TCLP testing on refractories at the time of furnace rebuilding would be advisable if there is a history of either alloying or grain refining with additions of metals that are on the EPA's listing of toxic substances or hazardous wastes (see Section 14).

Another type of melting furnace available, although not in common use, is an electric induction furnace. These units have no emissions from the fuel, but may generate some emissions from non-aluminum materials included in the furnace charge.

The remelting process includes degassing and removal of impurities from the molten metal with a fluxing and/or filtering step. There are a number of fluxes, with the most common incorporating chlorine and an inert carrier gas, typically argon, although other gases are sometimes used [19].

Other flux gases can include chlorine, and mixtures of chlorine, CO and argon or nitrogen. The presence of CO mandates controls to prevent escaping flux gases from becoming a health risk. Solid, or ''pill'' fluxes, such as sulfur hexafluoride, may result in emissions that could be a health hazard.

The handling of chlorine and the emission of the chlorine or chlorine compounds during the actual fluxing operation require precautions for operator safety and control measures to keep fugitive emissions to a minimum. Free chlorine escaping can combine with moisture in the air to form HCl, which must be controlled.

Melting of aluminum generates a slag, or as called in the aluminum industry, dross. The dross consists of aluminum oxide, carbides, nitrides and entrained aluminum. The amount of dross generated, and subsequently removed is highly dependent on operator skill in removing this dross from the surface of the bath after the meltdown, alloying and fluxing operations have been completed. Where other alloying elements (i.e. magnesium, silicon, zinc, copper, chromium, etc.) have been used, some traces of those elements or their oxides are usually present in the dross. Because drosses typically amount to as much as several percent of the material in the furnace and often contain as much as 50% entrained aluminum, they are sent to a dross recycler, where the metallic aluminum is recovered. The oxide fraction is sometimes recycled into abrasives, fiberglass or products that utilize aluminum oxide as a component.

Some melting furnace designs have an external well that is connected to the hearth by a submerged arch in the refractory wall. These are generally used to remelt scrap and will be discussed later in this chapter. Where scrap is charged into a furnace sidewell, the presence of oil, grease, etc. can produce a substantial amount of smoke, which must be captured and passed through a baghouse. Sometimes the smoke particles can contain adsorbed cadmium, chromium or other materials.

These other materials can be organics, such as PCBs from scrap sources. The US EPA's Toxic Characteristic Leaching Procedure (TCLP) analysis of this collected baghouse dust could contain substances exceeding the limits specified in EPAs list for Characteristics of Toxicity, or in the compilation of listed wastes from non-specific sources.

Since most melting furnaces use natural gas for a heat source, the pollution prevention principles most commonly used are the use of low NO_x burners and some method of heat recuperation. Furnace designs are also used to improve melting efficiency, as the minimizing of the unit release of any pollutant is a credit in the application of life cycle analysis to a particular product. We also suggest that higher productivity, or faster metal throughput, can reduce the generation of dross, which has a time-rate function as part of its formation/generating cycle [20].

After casting and before rolling, some ingots are homogenized, and scalped. The scalping operation will often use a soluble oil as a cutting lubricant/coolant. Management of this oil, its storage, filtration and disposal would parallel that done to rolling lubricants and will be addressed later in this chapter. Baling or otherwise consolidating these scalper chips can potentially carry liquid into the remelt furnace, a safety concern. This necessitates a scrap pretreatment or preheating operation prior to charging into a furnace.

5.2 Rolling

Following the casting of rolling ingots, a heating operation preceding the hot rolling is used to bring the ingots to a suitable temperature. For some products, an atmosphere control for the ingot heating furnace may be used. Ammonium fluoborate is one compound which has been used in the past to minimize surface oxidation and may still be used in some facilities for that purpose.

5.2.1 Hot Rolling

Hot rolling uses a soluble oil, generally in about a 5–10% emulsion, along with additives that inhibit fungus formation and stabilize the emulsion. Controlling the steam and entrained oils that result from hot rolling is a challenge, due to the large volumes of air/steam that must be handled. This air/steam/oil aerosol is typically sent through a condenser for recovery and reuse of the liquid phase. However, since no equipment is 100% efficient, some of this material is emitted to the atmosphere. Identification and quantification of the material emitted is part of the facility permitting requirements.

Disposal of spent rolling oils is usually carried out by breaking the emulsion and shipping the recovered oil to a recycler. Management of the remaining water, which contains some oil, is best done by an in-house water treatment system. Discharges of this contact water are regulated, with most jurisdictions having specific permit limits on the analysis of discharged water. Some bearing lubricants contain heavy metal compounds, some of which can enter the oil system. These contaminants can add complexity to the recycling of spent rolling solutions.

5.2.2 Cold Rolling

Cold rolling typically uses a rolling lubricant/coolant consisting primarily of n-alkanes (C_{12}- C_{22}). This oil has a narrow distillation range (300–330°C), to ensure

virtually complete distillation during annealing operations. This oil also contains additives to improve lubricity and rolling characteristics. During rolling, this oil can form a mist that is collected in a closed system. Other similar rolling lubricants are also in use in some facilities. Fugitive emissions can become a factor in the use of this type of oil in today's high speed rolling mills and limitations of the control systems.

The cost of rolling oils virtually mandates their recovery and reuse, with air pollution regulations specifying the limits of releases from these facilities [22].

The distillation of the oil during annealing could present a problem with minimizing its release to the atmosphere, or release to the inside of the manufacturing facility where personnel are present. Capture of these vapors could be a first choice for control, however, incineration of this distillate is also an option if facility emission permits or operating permits will allow this.

Finishing operations, including heat treatments with water quenching, slitting, levelling, stretching and packing can involve use of lubricants that should be kept from being spilled on the ground. In general, rolling lubricants are contained in mill sumps and process tanks located inside a building. These are not regulated as either above ground storage tanks (ASTs) or underground storage tanks (USTs). Most other tanks associated with the outside storage of new or used oils fall into the area where a regulation of some type is applicable. This determination is usually site specific, depending on the usage. No matter what, leaking tanks can result in release of the contents into the environment, whether it be into soil, water or air. Not only that, current UST regulations in the USA required overfill protection, leak detection, corrosion protection and spill containment by December 1998.

The conversion of ingot into sheet by use of some method of continuous casting has the potential to introduce into the atmosphere the vapors or mists from the cooling water, which may contain lubricants or proprietary constituents. The individual processes and practices used should identify all input materials and conduct a material balance to determine their fate. This then will be useful in establishing the mode of emission control.

5.3 Extrusion

The extrusion process uses billets, in a number of diameters, that are generally cast using input ingot and scrap in a manner similar to rolling ingots. Remelt furnaces and other casting practices parallel those for rolling ingot production. Billet heating typically uses a gas fired heater, which has a defined range of emissions, depending on the fuel used.

Some facilities produce their own dies, which may require heat treatment of the steel, with resulting waste materials from salts or quenchants that could present disposal problems.

Dies are usually cleaned with caustic (NaOH) to remove traces of metallic aluminum that interferes with proper extruded product quality, and requires neutralizing and removal of aluminum and other materials from these solutions prior to disposal [22].

6   PRINTING, COATING AND ANODIZING

6.1   Printing

Printing is generally conducted on the web of a very thin sheet of aluminum foil, or a laminated structure containing foil, with paper and/or plastic. The cleaners used for surface preparation, as well as the adhesives and inks, are typically proprietary materials. However, the latest advanced technologies that are being implemented in some facilities have replaced most of the solvent based inks with water based products, greatly reducing the emissions resulting from the thermal curing of the coated product. In other facilities, thermal oxidizers are used to combust the vapors coming from the coating and curing of the printed material.

Disposal of the waste materials from the printing process, typically solvents, is generally through disposal by a hazardous waste management contractor experienced in handling these materials.

6.2   Coating

Coating of aluminum, generally in coil form, is typically done in a continuous processing line, beginning with an alkaline washing solution, with surfactants and detergent used to remove residual oil. This is followed by the application of a conversion coating that uses a hexavalent chrome solution to produce a trivalent chromate on the surface of the strip. A chromic acid rinse is subsequently used to neutralize the residual conversion chemical. The resultant coating allows paint to be applied and adhere properly. Without that pretreatment, painted aluminum would be subject to relatively easy peeling of the paint with resultant low market acceptance. There are a number of other pretreatment chemicals used, including chromate/phosphate or zinc phosphate. Cleaning prior to conversion coating and neutralization of the treatment chemicals is done in similar fashion for coating lines. There are currently a number of alternative cleaning technologies that do not involve chlorinated or other VOC generating materials [22].

The application of paint, which is predominantly water based, is done with rollers or sprays in a coating room that allows control of fugitive emissions. The most common control mechanism for these emissions is thermal oxidation (combustion). The resulting dispersions of combustion by-products from curing ovens are included in the emissions listed in the operating permit for the facility. Obviously, implementing pollution prevention, by eliminating the solvent based paints and volatile solvents that may be present in some water based coatings is a big step in reducing the cost of permit maintenance.

Powder coating, with low levels of solvents, is being introduced in certain applications. These coatings are generally cured with infra-red heaters or convection ovens, with all volatiles sent to a combustor or used in the combustion air supply to the burners, where they are destroyed.

The residual chrome solutions from the conversion coatings are neutralized and treated with lime to convert the chrome into a readily disposable solid waste [22,24].

6.3 Anodizing

There are a number of anodizing processes, with the most commonly used involving a cleaning step, with alkaline/detergent cleaners, and several rinses. For high reflectivity anodizing, a bright dip, containing sulfuric, nitric and phosphoric acids is used to level the surface. The residual smut is removed with a dilute nitric acid solution. Because of the cost of the reagents used, and the hazardous materials involved in these processes, regeneration of the acids as part of the process is a common operation. It is still pollution prevention, with an economic incentive.

The actual anodizing is done in a sulphuric acid solution, using dc current. This operation also necessitates regeneration of the acid solution as a part of the process, for the same reasons as for the pretreatments. There are a number of other acids/reagents used for anodizing to obtain certain characteristics, such as colors or preparation for the use of teflon or other compound impregnation of the oxide film. Sealants for the anodic oxide include plain water or another compound, nickel acetate being one of many in common use.

The effluents from all types of finishing lines can be either liquid or gaseous, so treatments must be used that will keep these within the limits prescribed by regulatory authorities. Since water quality limits for discharges may vary between locations and between countries, careful adherence to the limits specified and used in these control systems is required.

The fume recovery systems, with appropriate hooding, are used to keep fugitive emissions of the acids under control, and to preclude any health related problems. Air pollution regulations for emission of acid vapors are stringent, as are those for the reagents used in the anodizing lines or in the sealing processes [24].

7 THERMAL TREATMENTS

7.1 Annealing

From the time ingots or billets are cast, a thermal treatment is usually part of the fabricating practices. Where a protective atmosphere is used, it is usually an inert atmosphere, either nitrogen or an exothermically produced and dehydrated mixture of CO and H_2, typically containing 95% N_2 with the total combustibles in the range of 5%. This CO/H_2 atmosphere has the capability of causing health problems if kept in a confined space. Venting to the atmosphere formerly was a common practice. This would now require inclusion in the facility emissions inventory, with application to the operating permit.

Current Best Management Practice (BMP) for annealing or heat treatments at temperatures higher than about 200°C uses nitrogen, dehydrated to at least −40°C, to prevent discoloration from oil residues and to minimize surface oxidation, particularly from alloys containing magnesium.

7.2 Solution Heat Treatment

Solution heat treatments can be batch based, with individual sheets hung on a structure that is raised into the furnace for the required cycle and then rapidly lowered into the quench tank. This water, or quenching solution, will become contaminated with aluminum and must be treated prior to discharge. Contact water

in most jurisdictions is a regulated discharge, so periodic analytical work and control procedures are recommended precautions against exceeding discharge limits. The same constraints apply to heat treatments of any kind carried out in a continuous processing line.

7.3   Aging, Stabilizing and Stress Relieving

These thermal treatments can be carried out in the absence or presence of a protective atmosphere and can be on dry or oily material. Depending on the temperature and cycle criteria, some emissions may result, necessitating controls.

8   RECYCLING

The positive environmental impacts associated with aluminum recycling have been a major factor in the growth of this activity. The aluminum beverage can recycling story has been one which exploited the energy and material savings resulting from the recycling programs that have proliferated since the first cans were recycled in the late 1950s. The growth of this package and the associated recycling has reached to all states in the USA as well as countries where the aluminum beverage can has been marketed [25–27].

   The use of a "life cycle" type analysis in the early years of recycling established the energy savings of 95% for recycled aluminum, in comparison with primary sources of aluminum. The attendant savings in material being mined and transported materially reduced the environmental impact of producing aluminum. These savings were applicable to the production of alumina and primary aluminum, which is where the largest proportion of energy for producing aluminum is consumed.

   Some of the conditions to be prepared for when using scrap purchased on the open market include the presence of paint, which, in the case of non-food aluminum applications, can contain cadmium-based primers or chromate conversion coating residues (which go into the smoke delivered to the baghouse or into the dross, neither of which is a desireable event as it could cause these to become classified as a hazardous waste). Some scrap can be contaminated with oil, which can cause excessive smoke during remelting. Some types of scrap recovered from very old automobile transmission can have adhering transmission fluid that may contain PCBs. This material can adhere to the baghouse dust, potentially causing disposal problems. Many facilities have baghouse operating procedures that call for the addition of pulverized limestone to the baghouse to coat the bags after the bags have been shaken down to remove accumulated dust. Limestone may not give sufficient pH control to avoid TCLP failure, but hydrated lime may provide the corrosion protection for the steel structure and be sufficiently alkaline to allow passing the TCLP test.

   Should multiple leaching procedures for TCLP tests ever become a reality, the baghouse bag coating/corrosion protection system would probably have to be revisited [28,29].

   Another particularly troublesome contaminant in scrap, not frequently seen, but worthy of mention due to the consequences of it being present, is radioactive material that could, if present in large quantities, cause serious health problems and subsequent disposal problems when cleaning up the resulting contamination.

Common sense practices in many forward thinking facilities are to install radiation detection equipment outside of the receiving plants so that any detected loads can not get inside the plant; these loads should be directed to the proper facilities for disposition. The steel industry makes extensive use of radiation detectors, particularly where scrap is processed.

Some types of scrap, including old electrical wiring can contain vinyl insulation. Some of these vinyls contain lead stabilizing compounds, which, after removal, can be classified as a hazardous waste, with recordkeeping, storage and expensive disposal as management requirements [29].

9   FUELS AND LUBRICANTS

The environmental considerations associated with fuels are primarily directed at petroleum motor fuels and the various oils used for heat generation. The motor fuel issues revolve around storage and handling of gasoline, diesel and propane. Process lubricants and fuels cover the range from light rolling lubricants (kerosene-like, parafinnic or napthenic oils), furnace fuels (No. 2 fuel oil) and boiler fuels (Nos. 4 or 6 fuel oil). In some facilities, the use of PCBs in power transformers is still permitted and monitoring these transformers, filing annual reports and proper disposal of wastes is the focus of very specific, strict regulatory oversight (refer to Section 14).

The storage tanks for petroleum products must conform to the regulations requiring containment, spill and overfill protection, tank integrity, registration with state or federal agencies, monitoring for leaks and periodic inspections. Containment measures to prevent contamination of surface waters is required in certain installations, as is disposal of spent lubricants, filter media and petroleum containing sludges.

10   PROCESS WATER AND CHEMICALS

For a facility that has transformers or rectifiers on site containing a coolant with more than 50 parts per million (PPM) of Polychlorinated Biphenyls, prescribed regulations are applicable. Regular inspections, with detailed record keeping and quarterly or annual reporting requirements, are mandatory. Clean up of spilled PCBs is a regulated activity, which in most cases requires immediate reporting of such spills to the regulatory authorities. Failure to comply with specified clean up measures can result in penalties.

Where a facility has on site the reportable quantity of chemicals on EPAs list of hazardous materials, annual usage, reports of discharges or emissions are required by law. There are a number of forms that must be filled out and submitted to the regulatory authorities. These fall under Community Right to Know regulations (refer to Section 14).

Water that is collected as a runoff from a manufacturing operation must meet the applicable standards for the particular jurisdiction, whether it be under State or Federal regulatory authority. Most regulatory authorities regulate the waste water leaving a facility. Sanitary water, from locker rooms, rest rooms and from kitchen or cooking units are usually sent directly to a sanitary sewer for movement to a publicly owned treatment works (POTW), or to an on site treatment plant.

Contact water, which can be as relatively uncontaminated as water used to cool ingots during casting or heat treatment quench water, has to be treated to a permitted level of purity prior to discharge to the POTW or to a receiving water, such as a river.

Water that results from separation from soluble oils, must also be treated to attain a predetermined and permitted purity level. Likewise, some facility operations that use either acidic or alkaline solutions must have these neutralized prior to discharge. The water treatment facility that services a plant that uses chromate conversion coatings prior to painting strip or extrusions, or a can manufacturing plant, must treat their liquid process wastes to remove the heavy metals and then neutralize the effluent.

11   BY-PRODUCTS, CO-PRODUCTS AND WASTE DISPOSAL

In all facilities some waste materials will need proper, and economical disposition. Ideally, recycling or reuse of some of these wastes can minimize the overall disposal requirements. The concept of pollution prevention has numerous potential applications in most facility activities if sufficient forethought is used. Where a by-product, such as red mud from an alumina plant can be processed to recover another material, a reduction in waste disposal costs should result. Or, if concurrent to alumina recovery another material is recovered, wastes are reduced and additional material is available for marketing. This same concept applies to dross from a molten metal processing facility. Recycling the dross to recover metallics is commonly practiced due to the value of the metal.

One of the large sources of hazardous wastes is the spent pot liner from the reduction process. These had been classified as a listed hazardous waste (K-088) in the early 1980s, and were later delisted by EPA. The reason for the classification was the presence of cyanide, which was generated by heat, air and the pitch binders in the lining and fluorides from the pot bath.

Regulations in the mid 1990s classified this material as a hazardous waste, requiring treatment prior to landfilling in a permitted landfill. Of the many processes evaluated to treat this material, Reynolds Metals Company developed a process acceptable to EPA, which designated it as Best Demonstrated Applicable Technology (BDAT). This designation made the process the approved disposal method for spent pot liner.

12   TOXICOLOGICAL IMPACTS

Toxicity is defined by Webster as a subject that deals with the effects of poisons on living organisms. The capability to function as a poison is related to the material involved and the quantity ingested or applied. Almost every substance can have the effect of a poison, if taken in large enough quantities. For the purposes of this chapter, the toxicity of the materials involved in the production of aluminum, the production processes and their effect on the facility personnel, the areas surrounding a facility and of the aluminum product itself will be discussed.

Some of the more publicized observations of toxicity issues resulting from the production of aluminum involved the effect of fluorides on the teeth of cattle. The fluorides were airborne emissions from some primary reduction plants which

were eventually deposited on nearby pasture land. The nature of cattle food and its recirculation in their digestive tracts multiplied the exposure to this material, causing tooth loss. This is an example of too much versus the right quantity of a good thing, with fluoride at low concentrations in drinking water having a beneficial effect in dental cavity reduction. The aluminum industry has since reduced the emission of particulate and gaseous fluorides to very low levels, and concurrently returned a valuable material to the production process.

Of current interest is the impact of legislative and administrative actions pointed at packaging of all types. Where the effect is most visible on aluminum is the legislation originating with the Council of Northeastern Governors (CONEG) that limits hexavalent chromium, cadmium, lead and mercury in packaging, which includes not only the product, but all of the wrappings and shipping pallets. The purpose of this legislation was to reduce the presence of these four materials both in the leachate from landfills and from refuse burning incinerator emissions and ash. The total quantity of these four materials allowed in a product or package is 100 ppm, which for practical purposes is about 80–90 ppm in order to compensate for analytical uncertainty [31].

The impact of the eighteen states enforcing CONEG legislation on aluminum beverage can production relates to a large degree on the recycling of aluminum cans. In the United States today, recycled cans make up more that three quarters of the production of the sheet used for can manufacture. The CONEG jurisdictions have restrictions on the charging of metal containing the four restricted materials into furnaces. The most often cited source of these restricted materials, particularly lead, is before some individuals bring aluminum cans to a recycling facility they add pieces of iron, lead or other heavy material to the cans brought in for recycling in order to maximize the payment. The aluminum producers have utilized magnetic separation, air classifiers and intense scrutiny of incoming materials to combat this activity.

The CONEG legislative actions have forced the producers of aluminum as well as the beverage producers to conduct extensive research on the pick-up of the restricted materials into carbonated beverages, over short and extended time periods. These research studies also dealt with establishing relationships with both air emission and water quality standards [32].

Another legislative action that impacts the aluminum, and other metalworking industries as well, is California's Safe Drinking Water and Toxic Enforcement Act of 1986, commonly known as "Proposition 65". This law mandated hazard warnings for risks of cancer and reproductive toxicity from products, jobs and the environment. The US EPA listing of toxic chemicals referred to in this legislation includes most of the chemicals and materials in common use in the metalworking industry, including the aluminum industry. The presence of any of these listed materials requires a warning notice on the package or in the workplace. The implications of a warning about the presence of materials on EPA's list of toxic substances are obvious on the potential sale in items so labeled. The posting of warnings in the workplace where these materials are present also carries a message to the employees [31].

Proposition 65 specifies a warning is required for a lifetime risk of 1/100,000 for carcinogens over a lifetime (70 years specified) which makes the standard 70 times more stringent than it appears to a casual observer. For reproductive toxicants (lead, hexavalent chromium and cadmium), the risk cutoff and warning is the

amount of a chemical for which 1,000 times this exposure will have no observable effects.

One of the major effects of the Proposition 65 legislation affected the production of plumbing fixtures, such as faucets and valves, by the brass foundry industry. Some of the scrap copper used in this type of production contained lead, most likely from solders. The lead restrictions mandated use of carefully controlled scrap streams and analytical procedures on these products. Other areas involved the installation of copper plumbing, where the previously used lead/tin solders were replaced with lead-free solders, primarily pure tin.

Some of the states have enacted similar legislation, but most have some areas of language differences from the California law that mandate close attention to the applicable statutes. The presence of ''bounty hunter'' provisions in these laws allows outside parties to bring action if discrepancies are believed to be present.

For toxic materials in the workplace, some of the more publicized materials are asbestos and PCBs and sites where RCRA (superfund related activity) is underway. Since asbestos was a commonly used material in cast house and process furnace equipment, its presence in older facilities could require extraordinary precautions as well as monitoring to preclude release into the environment. Asbestos removal is best left to organizations that are trained and licensed to conduct such work.

Some current regulatory inquiries and investigations are pointing at placing controls or otherwise limiting the use of non-asbestos ceramic/refractory fibers in very small sizes.

Where PCBs are most commonly used, transformers and rectifiers, there are very stringent regulations concerning their use and storage. The spilling or leakage of very small quantities calls for reporting to the regulatory authorities and conforming to a prescribed clean-up procedure. Disposal of PCB containing products or wastes are confined to a few permitted sites, which should be identified to the facility personnel involved in managing the facility PCB programs.

13   STANDARDS

13.1   ASTM Standards

ASTM has a number of Committees that deal with environmentally related subjects. These are Committees D-18, on Soil and Rock; D-19, on Water; D-22, on Sampling and Analysis of Atmospheres; D-34, on Waste Management; D-35, on Geosynthetics, E-6, on Performance of Buildings E-35, on Pesticides; E-36, on Conformity Assessment (formerly Laboratory and Inspection Agency Evaluation and Accreditation); E-47, on Biological Effects and Environmental Fate; E-50, on Environmental Assessment and E-51, on Environmental Risk Management. Cumulatively, these Committees have issued nearly 1,000 Standards, in the format of Specifications, Guides Test Methods and Practices. More specific information is available from ASTM, 100 Barr Harbor Drive, West Conshohocken, Pennsylvania 19428–2959. The phone number is: (610) 832–9500; the fax number is: (610) 832–9555 and the e-mail number is: http://www.service@astm.org [33].

13.2 Best Management Practices

The concept of Best Management Practices (BMP) for application at a number of facilities is the common sense approach to compliance with applicable regulations in order to minimize adverse impacts on the environment as well as to minimize potential near and long term environmental liabilities. The BMPs should recognize, especially for a multi-national company, that similar activities in different countries may have to comply with widely different environmental regulations. Compliance with the applicable minimum criteria, while maintaining cost effectiveness, and adhering to the goal of environmental excellence should be the result of implementing BMP concepts. Monitoring and enforcing the BMPs is generally the function as well as the responsibility of the organizations' environmental management personnel. The development and use of BMPs should incorporate the criteria of the ISO, EU, ASTM or other standards in order to inform customers, regulators, government officials and other interested parties that the organization is at the leading edge of environmental performance.

13.3 European Standards

The environmental standard of importance in Great Britain is BS 7750. The European Union has the Environmental Management Audit Scheme (EMAS), which requires disclosure of facility or site audit results. An organization of multinational companies has formed an International Audit Protocol Consortium to evaluate compliance with Environmental, Health and Safety (E H & S) regulations.

13.4 International Standards Organization (ISO)

The ISO is headquartered in Geneva, with a number of Technical Committees, (TC) of which TC 207, on Environmental Management, is of interest to readers of this publication. The Technical Assistance Groups (TAG), which are composed of volunteer members assigned to a Subcommittee (SC), address the specific subjects, as seen in Table 1. The subcommittees are further broken down into working groups that develop the draft standards.

The American National Standards Institute (ANSI) is the United States member-body to ISO, with ASTM functioning as the Administrator, or Secretariat, to ISO TC 207. This US TAG is responsible for formulating the US positions on the ISO 14000 series of standards. Additional information on participation can be obtained from ASTM, American National Standards Institute, (ANSI), American Society for Quality (ASQI), or the National Science Foundation (NSF).

One of the concepts of the ISO standards is to involve multinational groups, representing a number of countries. These participants normally bring their individual countries' perspective into the consensus standards being developed. The concept of self-auditing the conformance of the facilities or organizations with a given standard differs from the EU standards, which utilize an outside auditor to gage compliance with EU standards. This self audit function differs from the EU audits in that disclosure of audit results is not required, only a statement of conformance.

Table 1 lists the status of ISO Environmental Standards, as of late 1997

Table 1 ISO 14000 Summary

ISO Standard Number	Status	Comments
ISO 14001—Environmental Management System	Published	ISO 14002, ISO 14004 also in this group
ISO 14010—Guidelines for Environmental Auditing	Published	ISO 14011, ISO 14012 and ISO 14015 are related standards
ISO 14020—Environmental Labels and Declarations	In draft stages	ISO 14021, ISO 14022, ISO 14023, ISO 14024 and ISO 14025 are related standards
ISO 14031—Evaluation of Environmental Performance	In second draft stage	
ISO 14040——Life Cycle Assessment	In final draft stage to be published in 1997	ISO 14041, ISO 14002 and ISO 14003 are related standards
ISO 14050—Environmental Management Vocabulary	In final draft stage	To be published in 1997
ISO Guide 64—Guide for Inclusion of Environmental Aspects in Product Standards	In final draft stage	To be published in 1997
Technical Report—Reference material for the implementation of 14001/14004 by forest organizations	In draft stages	

14 ENVIRONMENTAL REGULATIONS

This section lists the most relevant US Federal regulations and briefly summarizes the salient parts. For the full text of the given regulation, refer to the indicated title in the Code of Federal Regulations. Title 40, Protection of the Environment, contains the regulations of interest to the metalworking industry. The general guideline to consider is that any emission could be regulated, whether it be solid, liquid or gas. The degree of regulatory impact that would usually be applicable is related to what is being emitted and how much is involved. So the efforts to minimize emissions or discharges, if adhered to, and creatively managed, could result in a cost saving activity.

Since the Code of Federal Regulations is a lengthy and complex publication as well as being a constantly changing document, (updates are biannual) access to the latest update for a given rule is recommended. This source of information could be either staff experts, a consultant, or an attorney who are familiar with the specific regulation and its administration by the governing State or Federal authority.

The following brief summaries of the more important Title 40 regulations are designed to provide some information on their significance to the Aluminum Industry.

14.1 40 CFR Part 50—National Primary and Secondary Ambient Air Quality Standards (Clean Air Act Amendments—CAAA). This part lists the ambient primary air quality standards that must be met to protect public health. Secondary ambient air quality standards define levels of air quality judged necessary to protect public welfare from known or anticipated adverse effects of a pollutant. Section 112 of the CAAA requires facilities to implement a risk management program to assess hazards associated with accidental release of regulated substances. An emergency response system for incidents involving release is required. This emergency response system is currently being evaluated with its potential to incorporate all emergencies into one system, which would facilitate training personnel and notification of authorities when any emergency occurs. (Chlorine and propane or butane, above threshold quantities, are covered by this procedure). The New Source Review criteria involve Lowest Achievable Emission Rate (LAER) control technology in nonattainment areas and Prevention of Significant Deterioration (PSD) requirements in all other areas of the USA.

14.2 40 CFR Part 57—Primary Nonferrous Smelter Orders. This section deals with emissions from Primary copper, lead, zinc and molybdenum or other smelters operating before August 7, 1977. These smelters are generally reducing sulfide ores and the regulation applies to the emissions from the facilities.

14.3 40 CFR Part 61—National Emission Standards for Hazardous Air Pollutants (NESHAP). This part covers a number of pollutant emissions that are of interest to the aluminum industry; including beryllium, mercury, chromium, asbestos and fugitive emission leaks from equipment.

14.4 40 CFR Part 70 and 71—These parts address State and Federal Operating Permit Programs. The regulation provides for establishment of comprehensive State air quality permitting systems or Federal programs if the State does not have a final Title V program. This part also requires that fees be imposed on emissions from facilities and specifies quantities and types of various pollutants that define major emission sources.

14.5 40 CFR Part 117—Determination of Reportable Quantities for Hazardous Substances. The regulation applies to facilities with National Pollutant Discharge Elimination System (NPDES) permits, listing the quantities of a large number of pollutants, which must be reported to regulatory authorities if discharged in excess of the specified limits in a table in the regulation.

14.6 40 CFR Parts 122 to 131—These parts deal with the National Pollutant Discharge Elimination System (NPDES) for water quality regulations, particularly the Permitting Procedures, the Toxic Pollutant Effluent Standards and the Water Quality Standards. Permits and monitoring are required for facilities that have storm water run-off that could enter the waters of the USA, when it is in a point source or enters a separate storm drain sewer system rather than a combined storm and sanitary sewer system or as sheet run-off into a field. This regulation is probably best interpreted by local or state regulators.

14.7 40 CFR Parts 260 to 273—These parts deal with the identification, management, transportation and disposal of hazardous wastes and used oil.

Part 261 is important in that it contains the means of identifying what determines the characteristic of a hazardous waste and list of wastes deemed hazardous or acutely hazardous.

One of the first actions that should be taken in the facility is to determine if the facility is a hazardous waste generator. There are three classes: Large Quantity Generator (1000 or more kg per month of hazardous waste or 1 kg per month of acutely hazardous wastes or 100 kg per month of spill residues; Small Quantity Generators (more than 100 but less than 1000 kg per month or 1 kg per month of acutely hazardous wastes) and Conditionally Exempt Generators (less than 100 kg per month of hazardous wastes or less than 1 kg per month of acutely hazardous wastes). It is important for the facility operator to identify and characterize all wastes leaving the facility and to make sure the Employer's EPA Identification Number is on all manifests and other paperwork. Record keeping is critical to insure compliance with the regulations. In most cases, hazardous wastes are restricted from land disposal and so must be treated at the hazardous waste management/disposal site prior to disposal in order to conform with the regulations. Confirmation of that treatment is a vital record which must remain in the facility files for three or more years, along with a copy the shipment manifest that had been signed by the recipient.

Annual training, fully documented, must be provided to all personnel who have any contact with hazardous wastes. It is important to understand the distinction between reclaiming, which is a treatment of hazardous wastes (material is processed to recover a usable product, i.e. recovery of lead values from spent batteries) and must be permitted and recycling (i.e. remelting scrap to recover all materials introduced in the system) which is exempt from regulation.

14.8 40 CFR Parts 280 to 282—These parts deal with Underground Storage Tanks (USTs). In general, USTs containing petroleum products as well as those containing hazardous materials or hazardous wastes are regulated. In the USA, USTs are required to have leak detection, spill, overfill and corrosion protection by December 1998. There are regular monitoring and testing protocols as well as permitting and record keeping requirements. There are some minor exemptions to these regulations, but usually involve very small tanks, or those containing materials coming under another regulation, such as the Toxic Substances Control Act (TSCA).

14.9 40 CFR Parts 300 to 302—These parts deal with Oil and hazardous substance pollution contingency plans.

14.10 40 CFR Parts 303 to 307—These parts deal with the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) or Superfund.

14.11 40 CFR Parts 370, 372—These sections include the Emergency Planning and Hazardous Chemical Release Reporting (Community Right to Know). The most visible aspect of this regulation is in Section 313, the annual report required, listing releases of the regulated substances.

14.12 40 CFR Part 421—This part deals with the Nonferrous Metals Manufacturing – Point Source Category. The regulation addresses the effluent water discharges from bauxite refining, primary aluminum smelting and secondary aluminum smelting, and the criteria applicable to each. The use of Best Practicable Control Technology currently available (BPT) and Best Available Technology currently available (BAT).

14.13 40 CFR Parts 436, 440—These parts deal with mining and ore processing and specify the characteristics required for water effluents. Where mine drainage is involved, very large quantities of liquids must be treated or otherwise managed.

14.14 40 CFR Parts 446, 447 and 465—These parts deal with paints, inks and coil coating. The regulations list the water pretreatment standards and effluent limitations for virtually all painting, printing or cleaning operations. This is related to the NPDES permitting required.

14.15 40 CFR Parts 467 and 471—These parts deal with forming and metal powders. The regulations list the water pretreatment standards and effluent limitations for fabricating operations, including rolling, extrusion, drawing, stamping, etc. This is also related to the NPDES permitting required.

14.16 40 CFR Part 761—This part deals with Polychlorinated Bipheynls (PCBs), which were first regulated by the Toxic Substances Control Act of 1976 (TOSCA). PCBs are being phased out, although a number of transformers and capacitors still contain PCBs. Where these units contain less than 50 parts per million (ppm) they are not regulated. However, spills must be handled under strict guidelines, as specified in the regulation, depending on the type of spill that occurred. For PCB items in a facility, comprehensive record keeping, personnel training and annual reporting are required. Disposal of PCBs or PCB-contaminated items is closely regulated and paperwork, manifests, etc. are to be retained for a specified number of years.

14.17 40 CFR Part 763—This part deals with Asbestos, which has a number of very strict and comprehensive criteria relating to removal of Asbestos containing items from service. Such removal is best undertaken by a qualified contractor. Disposal of asbestos items is limited to a few TSCA permitted sites in the USA, which are similar for PCB items. Occupational Safety and Health Administration (OSHA) regulations concerning asbestos are found in 29 CFR 1926.

14.18 40 CFR Parts 792 to 799—These parts deal with analytical and laboratory activities. Should the facility plan to conduct analytical studies on hazardous or toxic materials, there are numerous criteria that must be adhered to in order to have regulatory authorities accept the test results. Outside laboratories, due to the extensive testing that is required for some materials, often have the skills and certifications necessary to produce acceptable data.

The next two references are from other sections of the Code of Federal Regulations that are applicable to aluminum production.

14.19 49 CFR Parts 171 to 179 deal with US Department of Transportation (DOT) regulations regarding transport of hazardous materials.

14.20 29 CFR Parts 1910 and 1911 deal with Occupational Safety and Health regulations, and cover the issues of toxics in the workplace.

15 TERMINOLOGY

The very large number of technical words and phrases found in the various regulations and standards are usually defined in the associated regulatory documents as well as in the applicable parts of the Code of Federal Regulations, since many of the words have meanings that are specific to that particular document. Others are to be found in the dictionary and should be retrieved from that type of publication. To include all of these would place an unnecessary burden on the reader.

ACKNOWLEDGEMENT

REFERENCES

1. R. J. Anderson, "Metallurgy of Aluminum and Aluminum Alloys," in *Handbook of Nonferrous Metallurgy*, (D. M. Liddell, ed.), 1945, McGraw-Hill, New York, pp. 1–43.
2. J. L. Bray, *Nonferrous Production Metallurgy*, 2nd Edn, 1953, John Wiley & Sons, Inc. New York, pp. 12–71.
3. Z. Konik, A. Derdacka-Grzymek, A. Stok and J. Iwanciw. "The Utilization of Fly Ashes from Various Countries for Alumina Production," *Proc. The Minerals, Metals and Materials Society*, 1994, pp. 23–27.
4. P. R. Atkins, "Bauxite Mining—A Worldwide Environmental Study," *Proc. The Minerals, Metals and Materials Society*, 1994, pp. 5–9.
5. J. J. Kotte, "Bayer Digestion and Predigestion Disilication Reactor Design," *Proc. The Minerals, Metals and Materials Society*, 1981, p. 49.
6. P. K. Maitra, "Recovery of TiO_2 from Red Mud for Abatement of Pollution and for Conservation of Land and Mineral Resources," *Proc. The Minerals, Metals and Materials Society*, 1994, pp. 159–165.
7. D. N. Banerjee, D. N. Dey, and P. K. Jena, "An Alternate Route for making Alumina by Soda Roasting of Bauxite Using Sintering Technique," *Proc. The Minerals, Metals and Materials Society*, 1988, pp. 131–135.
8. D. O. Yan and H. L. Li. "Discussion on Heat Consumption in the Manufacture of Alumina by Soda-Lime Sintering Process," *Proc. The Minerals, Metals and Materials Society*, 1990, pp. 157–160.
9. P. K. Maitra, "Effluents From the Integrated Aluminum Industry: Characterization and Possible Uses," UNEP Industry and Environment. January, February, March 1991, pp. 13–21.
10. N. E. Richards and H. Forberg, "The Effectiveness of Lithium Fluoride Modified Baths in Aluminum Production," Light Metal Age, August 1995, pp. 52–57.
11. W. Haupin and H. Forberg, "Impact of Bath Composition on Performance and Economics of Aluminum Smelters," Light Metal Age, December 1995, pp. 41–50.

12. G. Bouzat, J. C. Carraz and M. Meyer, "Measurements of CF_4 and C_2F_6 Emissions from Prebaked Pots," *Proc. The Minerals, Metals and Materials Society*, 1996, pp. 413–417.

13. R. Huglen and H. Kvande, "Global Considerations of Aluminum Electrolysis on Energy and Environment," *Proc. The Minerals, Metals and Materials Society*, 1994, pp. 373–380.

14. F. M. Kimmerle and L. Noel, "COS, CS_2 and SO_2 Emissions from Prebaked Hall-Heroult Cells," 1997, pp. 153–158.

15. U. Mannweiler and R. C. Perruchoud, "Reduction of Polycyclic Aromatic Hydrocarbons (PAH) by Using Petroleum Pitch as Binder Material: A Comparison of Anode Properties and Anode Behavior of Petroleum Pitch and Coal Tar Pitch Anodes," *Proc. The Minerals, Metals and Materials Society*, 1997, pp. 555–558.

16. P. Pinfold, Proc. 35th Annual Conference of Metallurgists of C.I.M., Light Metal Age, February 1997, pp. 74–83.

17. M. J. Gibbs and C. Jacobs, "Reducing PFC Emissions from Primary Aluminum Production in the United States," Light Metal Age, February 1996, pp. 26–34.

18. L. T. Yap and M. Pourkashanian, "Low NOx Oxy-Fuel Flames for Uniform Heat Transfer," *Proc. The Minerals, Metals and Materials Society*, 1996, pp. 655–660.

19. I. Ventre, J. Rolland, "Replacing Ceramic Foam Filtration by an ALPUR S1000 in line Treatment: A Winning Bet at Cableries De Lins," *Proc. The Minerals, Metals and Materials Society*, 1995, pp. 1245–1251.

20. R. Abernathy, J. McElroy and L. T. Yap, "The Performance of Current Oxy-Fuel Combustion Technology for Secondary Aluminum Melting," *Proc. The Minerals, Metals and Material Society*, 1996, pp. 1233–1239.

21. W. P. L. Carter, D. Luo and I. L. Malkina, "Investigation of Atmospheric Ozone Formation Potential of C_{12} – C_{16} N-Alkanes," Draft Report to Aluminum Association, May 1996.

22. T. Short, "The Application of Finishes to Aluminum Extrusions," Light Metal Age, August 1997, pp. 20–44.

23. R. Knipe, "'Green' Cleaning Technologies," Advanced Materials and Processes, August 1997, pp. 23–25.

24. A. I. Nussbaum. "Coil Coating Technology for the Nineties," Light Metal Age, August 1993, pp. 7–34.

25. Y. Bertaud, R. Guillermet, H. Lemaire, J. Ph. Cael, G. Nijhof and H. Rossel, "Packaging and Environment in Europe (Eureka Pack-ee) Automated Sorting of Aluminum from Domestic Waste," *Proc. The Minerals, Metals and Materials Society*, 1996, pp. 1041–1047.

26. A. I. Nussbaum, "Recycle II – Alcan Rolled Products $23 Million Aluminum Can Recycling Facility, Oswego, NY," Light Metal Age, June 1997, pp. 10–15.

27. C. Miller, "Profiles in Garbage: Aluminum Packaging," Waste Age, October, 1997, pp. 91–92.

28. W. B. Steverson, "MRFs and UBCs: A Concern Yet an *Opportunity*," *Proc. The Mineral, Metal and Materials Society*, 1995, pp. 1303–1307.

29. W. Stevens and F. Tremblay, "Fundamentals of UBC Decoating/Delacquering for Efficient Melting," *Proc. The Minerals, Metals and Material Society*, 1997, pp. 709–713.

30. G. Bourcier, "Non-Ferrous Metal Recovery," *Proc. Ninth International Conference on Municipal Solid Waste Combustor Ash Utilization, Resource Recovery Report*, November 1996, pp. 35–47.

31. W. K. Viscusi, "Economic Analysis of Proposition 65 and Related Warnings Measures," Duke University, September 1991.

32.   "Toxics in Packaging Clearinghouse. Model Toxics in Packaging Legislation: An Evaluation of its Provisions, Administration and Impact," Coneg Policy Research Center, Inc., Washington, DC, October 1994.
33.   K. Morgan, "Checking the Pulse of ISO 14000," ASTM Standardization News, September 1997, pp. 16–19.

The following items are recommended for additional information on the environmental and toxicology chapter.

Mining, Alumina, Reduction

L. W. Lewis, Shim You, W. Pederson and E. W. Black. "Vegetation of Thickened Red Mud Tailing Deposits without the use of Soil Capping Techniques," *Proc. The Minerals, Metals and Materials Society*, 1995, pp. 31–34.
A. Pinoncely and K. Tsouria, "FCB Flash Calciner Technology: Ten Years Performances," *Proc. The Minerals, Metals and Materials Society*, 1995, pp. 113–120.
Daekyoo Hwang, "Value Engineered Cap Designs," *Proc. Haztech International Conference*, Houston, TX, 16 pages, May, 1990.
Daekyoo Hwang, "Optimizing Remedies in Feasibility Study," *Proc. HMCRI (formerly Superfund) Conference*, December 1991.
R. A. Roberts and P. J. Ramsey, "Evaluation of Fluorocarbon Emissions from the Aluminum Smelting Process," *Proc. The Minerals, Metals and Materials Society*, 1994, pp. 381–388.
R. P. Pawlek, "Health and Safety Aspects when Relining Aluminum Electrolysis Cells," *Proc. (abstract only) The Minerals, Metals and Materials Society*, 1994. Complete article, Journal Of Metals.

Casting, Rolling

S. G. Epstein, "Summary of Ten Years of Molten Aluminum Incident Reporting," *Proc. The Minerals, Metals and Materials Society*, 1995, pp. 885–894.
P. Wallach, "Personal Protective Clothing: From Fundamentals to a Global Strategy of Protection in the Casthouse Environment," *Proc. The Minerals, Metals and Materials Society*, 1996, pp. 909–918.
G. J. Binczewski, "Molten Aluminum Plus Water – A Different Point of View," *Proc. The Minerals, Metals and Material Society*, 1997, pp. 737–744.
S. G. Epstein, "Update on Molten Aluminum Incident Reporting," *Proc. The Minerals, Metals and Materials Society*, 1997, pp. 887–897.
R. T. Richter, D. D. Leon, and T. L. Levendusky, "Investigations of Coatings Which Prevent Molten Aluminum/Water Explosions – Progress Report," *Proc. The Minerals, Metals and Materials Society*, 1997, pp. 899–904.
H. Spoel and W. A. Zebedee, "The Hot Aluminum Dross Recycling System," *Proc. The Minerals, Metals and Materials Society*, 1996, pp. 1247–1250.
D. J. Roth, "Recovery of Aluminum from Rotary Furnace Salt Cake by Low Impact Rotary Tumbling," *Proc. The Minerals, Metals and Materials Society*, 1996, pp. 1251–1253.
R. C. Bost and L. D. White, "Untangling the Web of Spill and Release Reporting," Pollution Engineering, October 1997, pp. 50–53.
S. M. Bassett, "Cutting the Red Tape: EPA Tailors NSR (New Source Review)," Pollution Engineering, June 1996, pp. 56–60.
S. Quesnel, S. Afshar and C. Allaire, "Corrosion of Refractories at the Bellyband of Aluminum Melting and Holding Furnaces," *Proc. The Minerals, Metals and Materials Society*, 1996, pp. 661–667.

Appendixes

Appendix 1 Alloy Equivalents

International Alloy Designation	Equivalent or similar alloy	Country of Origin
A199	1200	Austria
A199,5	1050	
E-AI	1350	
AlCuMgl	2017	
AlCuMg2	2024	
AlCuMgO,5	2117	
AlMg5	5056	
AlMgSiO,5	6063	
E-AlMgSi	6101	
AlZnMgCul,5	7075	
990C	1100	Canada
CB60	2011	
CG30	2117	
CG42	2024	
CG42 Alclad	2024 Alclad	
CM41	2017	
CN42	2018	
CS41N	2014	
CS41N Alclad	2014 Alclad	
CS41P	2025	
GM31N	5454	
GM41	5083	
GM50P	5356	
GM50R	5056	
GR20	5052	
GS10	6063	
GS11N	6061	

Appendix 1 *Continued*

International Alloy Designation	Equivalent or similar alloy	Country of Origin
GS11P	6053	
MC10	3003	
S5	4043	
SG11P	6151	
SG121	4032	
ZG62	7075	
ZG62 Alclad	7075 Alclad	
A5/L	1350	France
A45	1100	
A-G1	5050	
A-G0.6	5005	
A-G4MC	5086	
A-GS	6063	
A-GS/L	6101	
A-ML	3003	
A-M1G	3004	
A-U4G	2017	
A-U2G	2117	
A-U2GN	2618	
A-U4G1	2024	
A-U4N	2218	
A-U4SG	2014	
A-S12UN	4032	
A-Z5GU	7075	
E-A1995	1350	Germany
AlCuBiPb	2011	
AlCuMg0.5	2117	
AlCuMg1	2017	
AlCuMg2	2024	
AlCuSiMn	2014	
AlMg4.5Mn	5063	
AlMgSi0.5	6063	
AlSi5	4043	
E-AlMgSi0.5	6101	
AlZnMgCu1.5	7075	
lE	1350	Great Britain
91E	6101	
1H14	2017	
1H19	6063	
1H20	6061	
L.80, L.81	5052	
L.86	2117	
L.87	2017	
L.95, L.96	7075	
L.97, L.98	2014	
2L.55, 2L.56	5052	

Continued

Appendix 1 *Continued*

International Alloy Designation	Equivalent or similar alloy	Country of Origin
2L.58	5056	
3L.44	5050	
5L.37	2017	
6L.25	2218	
N8	5063	
N21	4043	
P-AlCu4MgMn	2017	Italy
P-AlCu4.5MgMn	2024	
P-AlCu4.5MgMnplacc.	2024 Alclad	
P-AlCu2.5MgSi	2117	
P-AlCu4.4SiMnMg	2014	
P-AlCu4.4SiMnMgplacc.	2014 Alclad	
P-AlMg0.9	5657	
P-AlMg1.5	5050	
P-AlMg2.5	5052	
P-AlSi0.4Mg	6063	
P-AlSi0.5Mg	6101	
Al99.5E	1350	Spain
L-313	2014	
L-314	2024	
L-315	2218	
L-371	7075	
Al-Mg-Si	6101	Switzerland
Al1.5Mg	5050	
Al-Cu-Ni	2218	
Al3.5Cu0.5Mg	2017	
Al4Cu1.2Mg	2024	
Al-Zn-Mg-Cu	7075	
Al-Zn-Mg-Cu-pl	7075 Alclad	
Al99.0Cu	1100	ISO
AlCu2Mg	2117	
AlCu4Mg1	2024	
AlCu4SiMg	2014	
AlCu4MgSi	2017	
AlMg1	5005	
AlMg1.5	5050	
AlMg2.5	5052	
AlMg3.5	5154	
AlMg4	5086	
AlMg5	5056	
AlMn1Cu	3003	
AlMg3Mn	5454	
AlMg4.5Mn	5083	
AlMgSi	6063	
AlMglSiCu	6061	
AlZn6MgCu	7075	

Appendix 2   Aluminum Specifications

Specification	Alloy and Product or Process
FED. STD. 123	Marking for domestic shipment (civilian agencies)
FED. STD. 184	Item identification marking for aluminum products
FED. STD. 245	Tolerances for aluminum wrought products
QQ-A-200	General specification for extruded rod, bar, shapes, tube and wire
QQ-A-200/1	3003 extruded rod, bar, shapes, tube, and wire
QQ-A-200/2	2014 extruded rod, bar, shapes, tube, and wire
QQ-A-200/3	2024 extruded rod, bar, shapes, tube, and wire
QQ-A-200/4	5083 extruded rod, bar, shapes, tube, and wire
QQ-A-200/5	5086 extruded rod, bar, shapes, tube, and wire
QQ-A-200/6	5454 extruded rod, bar, shapes, tube, and wire
QQ-A-200/7	5456 extruded rod, bar, shapes, tube, and wire
QQ-A-200/8	8061 extruded rod, bar, shapes, tube, and wire
QQ-A-200/9	8063 extruded rod, bar, shapes, tube, and wire
QQ-A-200/10	6066 extruded rod, bar, shapes, tube, and wire
QQ-A-200/11	7075 extruded rod, bar, shapes, tube, and wire
QQ-A-200/13	7178 extruded rod, bar, shapes, tube, and wire
QQ-A-200/14	7178-T76, TI6510, T6511 extruded rod, bar, shapes, and wire
QQ-A-200/15	7075-TI6, T76510, T76511 extruded rod, bar, shapes and wire
QQ-A-200/16	6061-T6 extruded structural shapes
QQ-A-200/17	6162 extruded wire, rod, bar, shapes, and tube
QQ-A-225	General specification for rolled, drawn, or cold finished wire, rod, bar and special shapes
QQ-A-225/1	1100 rolled, drawn, or cold finished wire, rod, and bar
QQ-A-225/2	3003 rolled, drawn, or cold finished wire, rod, and bar
QQ-A-225/3	2011 rolled, drawn, or cold finished wire, rod, and bar
QQ-A-225/4	2014 rolled, drawn, or cold finished wire, rod, and bar
QQ-A-225/5	2017 rolled, drawn, or cold finished wire, rod, and bar
QQ-A-225/6	2024 rolled, drawn, or cold finished wire, rod, and bar
QQ-A-225/7	5052 rolled, drawn, or cold finished wire, rod, and bar
QQ-A-225/8	6061 rolled, drawn, or cold finished wire, rod, bar, and special shapes
QQ-A-225/9	7075 rolled, drawn, or cold finished wire, rod, bar, and special shapes
QQ-A-225/10	6262 rolled, drawn, or cold finished wire, rod, and bar
QQ-A-250	General specification for sheet and plate
QQ-A-250/1	1100 sheet and plate
QQ-A-250/2	3003 sheet and plate
QQ-A-250/3	Alclad 2014 sheet and plate
QQ-A-250/4	2024 sheet and plate
QQ-A-250/5	Alclad 2024 sheet and plate
QQ-A-250/6	5083 sheet and plate
QQ-A-250/7	5086 sheet and plate
QQ-A-250/8	5052 sheet and plate
QQ-A-250/9	5456 sheet and plate

Appendix 2 *Continued*

Specification	Alloy and Product or Process
QQ-A-250/10	5454 sheet and plate
QQ-A-250/11	6061 sheet and plate
QQ-A-250/12	7075 sheet and plate
QQ-A-250/13	Alclad 7075 sheet and plate
QQ-A-250/14	7178 sheet and plate
QQ-A-250/15	Alclad 7178 sheet and plate
QQ-A-250/18	Alclad one side 7075 sheet and plate
QQ-A-250/21	7178-Tl6 and $\sim$651 sheet and plate
QQ-A-250/22	Alclad 7178-T16 and T1651 sheet and plate
QQ-A-250/24	7075-776 and 77651 sheet and plate
QQ-A-250/25	Alclad 7075-776 and 77851 sheet and plate
QQ-A-250/26	7011 Alclad 7075 sheet and plate
QQ-A-250/28	7011 Alclad 7178 sheet and plate
QQ-A-250/29	2124-T851 plate
QQ-A-250/30	2219 plate and sheet
QQ-A-367	2014, 2018, 2025, 2218, 2219, 2618, 4032, 5083, 6061, 6066, 6151, 7049, 7075, and 7076 forgings
QQ-A-430	1100, 2017, 2024, 2117, 2219, 3003, 5005, 5052, 5056, 6053, 6061, 7050, and 7075 wire and rod for rivets and cold heading
QQ-A-1876	Aluminum foil
QQ-B-655	4045, 4343, 4145 and 4047 brazing filler metals
QQ-R-566	1100, 2319, 4443, 4047, 4145, 5183, 5356, 5554, 5556, 5654, welding rod and electrodes, and 242.0, 295.0, 355.0, 356.0, 712.0 welding rod
WW-C-00540	Rigid conduit
WW-P-402	Alclad 3004-H34 culvert pipe
WW-T-700	General specification for drawn tube, seamless
WW-T-700/1	1100 drawn tube, seamless
WW-T-700/2	3003 drawn tube, seamless
WW-T-700/3	2024 drawn tube, seamless
WW-T-700/4	5052 drawn tube, seamless
WW-T-700/5	5088 drawn tube, seamless
WW-T-700/6	6061 drawn tube, seamless
WW-T-700/7	7075 drawn tube, seamless
	Military Specifications
Specification	Alloy and Product or Process
MIL-STD-129	Marking for shipment and storage
MIL-STD-649	Preparation for storage and shipment
MIL-C-915	Alclad 5056-H392 wire
MIL-C-5541	Chemical films for aluminum and aluminum alloys
MIL-H-6088	Heat treatment of aluminum alloys
MIL-W-6712	1100 and 4043 metal spraying wire
MIL-T-7081	6061 hydraulic tube
MIL-A-8625	Anodic coatings for aluminum alloys
MIL-A-12545	1100, 2014, 6061, 6070, and 7075 impacts
MIL-E-16053	1100, 2319, 4043, 4047, 4145, 5039, 5183, 5356, 5554, 5556, and 5654 bare welding electrodes

Continued

Appendix 2 *Continued*

Specification	Alloy and Product or Process
MIL-F-17132	6061 tread plate
MIL-B-20148	4047 and 4343 brazing sheet
MIL-A-22771	2014, 2219, 2618, 6061, 6151, 7039, 7049, and 7075 forgings and rings
MIL-P-25995	3003, 6061 and 6063 pipe
MlL-A-45225	5083, and 5456 forged armor
MIL-A-46027	5083 and 5456 armor plate
MIL-A-46083	5083, 5456 and 2219 extruded armor
MIL-A-46104	6070 extruded rod, bar, shapes and tube
MlL-A-46118	2219 armor plate and forgings
MIL-T-50777	2024 seamless drawn tube
MIL-A-52174	Duct sheet
MIL-A-52242	7001 extruded rod, bar and shapes
MIL-C-60536	Anodic hard coatings for aluminum alloys
MIL-C-60539	Anodic coatings for aluminum alloys
MIL-A-81596	2024, 3003, 5052 and 5056 foil

American Society of Mechanical Engineers (ASME)

Specification	Alloy and Product or Process
SB 209	1060, 1100, 3003, Alclad 3003, 3004, Alclad 3004, 5050, 5052, 5083, 5086, 5154, 5254, 5454, 5456, 5652, 6061 and Alclad 6061 sheet and plate
SB 210	1060, 3003, Aced 3003, 5052, 5154, 6061, and 6063 drawn seamless tube
SB 211	2014, 2024 and 6061 rolled, drawn, or cold finished wire, rod and bar
SB 221	1060, 1100, 2024, 3003, 5083, 5086, 5154, 5454, 5456, 6061 and 6063 extruded wire, rod, bar, shapes and tube
SB 234	1060, 3003, Alclad 3003, 5052, 5454, and 6061 drawn, seamless tube for condensers and heat exchangers
SB 241	1060, 1100, 3003, Alclad 3003, 5052, 5083, 5086, 5454, 5456, 6061 and 6063 pipe and tube
SB 247	2014, 3003, 5083, and 6061 hand and die forgings
SB 308	6061 rolled or extruded standard structural shapes

American Society for Testing and Materials (ASTM)

Specification	Alloy and Product or Process
B 209	1060, 1100, 2014, Alclad 2014, 2024, Alclad 2024, Alcladed one side 2024, 2124, 2219, Alclad 2219, 3003, Alclad 3003, 3004, Alclad 3004, 3005, 3105, 5005, 5050, 5052, 5083, 5086, 5154, 5252, 5254, 5454, 5456, 5457, 5652, 5657, 6061, Alclad 6061, 7075, Alclad 7075, Alclad one side 7075, 7178, Alclad 7178 sheet and plate.

Continued

Appendix 2 *Continued*

Specification	Alloy and Product or Process
B 210	1060, 1100, 2011, 2014, 2024, 3003, Alclad 3003, 5005, 5050, 5052, 5083, 5086, 5154, 5456, 6061, 6063, 6262 and 7075 drawn, seamless tube
B 211	1060, 1100, 2011, 2014, 2017, 2024, 2219, 3003, 5052, 5056, Alclad 5056, 5154, 6061, 6262, and 7075 rolled, drawn, or cold-finished wire, rod and bar
B 221	1060, 1100, 2014, 2024, 2219, 3003, Aided 3003, 3004, 5052, 5083, 5086, 5154, 5454, 5456, 6005, 6061, 6063, 6066, 6262, 6351, 6483, 7005, 7075, and 7178 extruded wire, rod, bar, shapes and tube
B 230	1350-1119 wire
B 231	Aluminum conductor, concentric-lay-stranded
B 232	Aluminum conductor, steel reinforced, concentrically-stranded
B 233	1350 redraw rod
B 234	1060, 3003, Alclad 3003, 5052, 5454, and 6061 drawn, seamless tube for condensers and heat exchangers
B 236	1350 bus conductor
B 241	1060, 1100, 2014, 2024, 2219, 3003, Alclad 3003, 5052, 5083, 5086, 5254, 5454, 5456, 5652, 6063, 7075, and 7178 seamless pipe and seamless extruded tube
B 247	1100, 2014, 2018, 2218, 2219, 2618, 2025, 3003, 4032, 5083, 6151, 6061, 6066, 7049, 7050, 7075 and 7076 die and hand forgings
B 308	6061 rolled or extruded standard structural shapes
B 313	1100, 3003, 3004, and 6061 round welded tube
B 316	1100, 2017, 2024, 2117, 2219, 3003, 5005, 5052, 5056, 6053, 6061, 7050, 7075 and 7178 rivet and cold-heading wire and rod
B 317	6101 extruded rod, bar, structural shapes and pipe for electrical purposes
B 345	1060, 3003, Alclad 3003, 5083, 5086, 6061, 6070 and 8351 pipe for gas and oil transmission
B 373	1145 and 1235 toll for capacitors
B 306	5005-1119 wire for electrical purposes
B 397	Concentrically stranded 5005-H19 conductor
B 308	6201-T81 wire for electrical purposes
B 399	Concentrically stranded 6201-T81 conductors
B 400	Compact round, concentrically stranded,1350 conductors, hard-drawn
B 401	Compact round, concentrically stranded aluminum conductors, steel reinforced
B 404	1060, 3003, Alclad 3003, 5052, 5454, and 6061 seamless condenser and heat exchanger tube with Integral fins

Appendix 2 *Continued*

Specification	Alloy and Product or Process
B 429	6061 and 6063 extruded structural pipe and tube
B 483	1060, 1100, 1435, 3003, 5005, 5050, 5052, 6061, 6083 and 6262 drawn tube
B 491	1050, 1100, 1200, 1235, 3003 and 6063 extruded round coiled tube
B 531	5005 rolled rod
B 544	Round solid 1350 grade conductor
B 547	1100, 3003, Alclad 3003, 3004, Aced 3004, 5050, 5052, 5083, 5086, 5154, 5454 and 6061 formed and arc welded round tube
B 609	Aluminum 1350 round wire, annealed and Intermediate tempers, for electrical purposes
B 832	Aluminum-alloy rolled tread plate
	Aerospace Material Specifications (AMS)
Specification	Alloy and Product or Process
2201	Tolerances for rolled, drawn, cold-finished, and centerless ground wire, rod, bar and forging stock
2202	Tolerances for sheet and plate
2203	Tolerances for drawn tube
2204	Tolerances for rolled or extruded standard structural shapes
2205	Tolerances for extruded rod, bar, shapes, and tube
2468	Hard-coating of aluminum alloys
2469	Process and performance requirements hard-coating
2470	Chromic-acid anodizing of aluminum alloys
2471	Clear sulfuric acid anodizing of aluminum alloys
2472	Dyed sulfuric acid anodizing of aluminum alloys
2473	Chemical conversion treatments for aluminum alloys
2474	Law electrical resistivity chemical conversion treatments for aluminum alloys
2808	Identification marking of forgings
2815	Line code identification marking of welding wire
2816	Color code identification marking of welding wire
4000	1060-0 sheet
4001	1100-0 sheet and plate
4003	1100-1114 sheet and plate
4004	5052-11191 foil
4005	5056-11191 foil
4006	3003-0 sheet and plate
4007	2024-O foil
4008	3003-1114 sheet and plate
4009	6061-0 foil
4010	3003-1118 foil
4011	1145-foil
4012	Laminated shim stock, edge bonded
4013	Laminated shim stock, surface bonded

Continued

Appendix 2 *Continued*

Specification	Alloy and Product or Process
4014	2014-T651 plate
4015	5052-0 sheet and plate
4016	5052-1132 sheet and plate
4017	5052-1134 sheet and plate
4020	Alclad 6061-T651 plate
4021	Alclad 6061-O sheet and plate
4022	Alclad 6061-T4 sheet and plate
4023	Alclad 6061-T651 sheet and plate
4025	6061-O sheet and plate
4026	6061-T4 sheet and plate
4027	6061-T6 sheet and plate
4028	2014-O sheet and plate
4029	2014-T6 sheet and plate
4031	2219-O sheet and plate
4033	2024-T351 plate
4034	Alclad 2024-1351 plate
4035	2024-O sheet and plate
4036	Alclad one side 2024-T3 sheet and T351 plate
4037	2024-T3 sheet and 1351 plate
4038	7075-T651 plate
4039	Alclad 707-T651 plate
4040	Alclad 2024-O sheet and plate
4041	Alclad 2024-T3 sheet and T351 plate
4043	6061-T451 plate
4044	7075-O sheet and plate
4445	7075-T6 sheet and 1651 plate
4046	Alclad one side 7075-16 sheet and 1651 plate
4048	Alclad 7075-O sheet and plate
4049	Alclad 7075-T6 sheet and 1651 plate
4050	7050-T7451 (formerly 173651) plate
4051	Aided 7178-O sheet and plate
4053	6061-T651 plate
4054	Brazing sheet No. 21-O
4055	Brazing sheet No. 22-O
4056	5083-0 sheet and plate
4057	6063-T1323 sheet
4056	5083-T1343 sheet
4059	5083-T1321 sheet and plate
4062	1100-HI4 drawn tube
4063	brazing sheet No. 11-O
4064	Brazing sheet No. 12-O
4065	3003-O drawn tube, seamless
4066	2219-T851 drawn tube, seamless
4067	3003-T114 drawn tube, seamless
4068	3219-T351 drawn tube, seamless
4069	5052-O drawn tube, special tolerances, seamless
4070	5052-O drawn tube, seamless

Appendix 2 *Continued*

Specification	Alloy and Product or Process
4071	5052-O drawn, hydraulic tube, seamless
4077	Alclad one side 2024∼ sheet and plate
4078	7075-17351 plate
4079	6061-O drawn tube, special tolerances
4080	6061-O drawn tube
4081	6061-T4 drawn, hydraulic tube
4082	6061-T6 drawn tube
4083	6061-T6 drawn, hydraulic tube
4086	2024-T3 drawn, hydraulic tube
4087	2024-O drawn tube
4086	2024-T3 drawn tube
4094	Alclad 2219-181 sheet and 1851 plate
4095	Alclad 2219-T31 sheet and 1351 plate
4096	Alclad 2219-O sheet and plate
4101	2124-1851 plate
4102	1100-F rolled or cold-finished wire, rod and bar
4107	7050-T74 (formerly T736) die forgings
4108	7050-T7452 (formerly T73652) hand forgings
4110	2017-T451 roiled or cold-finished rod and bar
4111	7049-T73 forgings and forging stock
4112	2024-16 rolled, drawn, or cold-finished wire, rod and bar
4113	6061-16 structural shapes
4114	5052-F rolled or cold-finished wire, rod and bar
4115	6061-O rolled, drawn, or cold finished wire, rod and bar
4116	6061-T4 rolled or cold finished wire, rod and bar
4117	6061-T6 rolled, drawn, or cold finished wire, rod and bar
4118	2017-T4 rolled, drawn, or cold finished wire, rod and bar
4119	2024-T351 rolled or cold finished rod and bar
4120	2024-T4 rolled, drawn, or cold finished wire, rod and bar
4121	2014-T6 rolled, drawn or cold finished wire, rod and bar
4122	7076-T6 rolled or cold finished wire, rod and bar
4123	7075-T651 rolled or cold finished rod and bar
4124	7075-T7351 rolled or cold finished bar
4125	6151-T6 forgings
4126	7075-T6 forgings and 7075-F forging stock
4127	6061-T6 die forgings and rolled rings and 6061-F forging stock
4128	6061-T451 rod and bar, rolled or cold finished
4129	6061-T651 rod and bar, rolled or cold finished
4130	2025-T6 forgings
4131	7075-T736 lorgings
4132	2618-T61 forgings
4133	2014-T6 forginga and 2014 forging stock
4134	2014-T4 forgings
4135	2014-T6 forgings
4139	7075-T6 forgings
4140	2018-T61 forgings

Continued

Appendix 2 *Continued*

Specification	Alloy and Product or Process
4141	7075-T73 die forgings and forging stock
4142	2218-F forgings
4143	2219-T6 forgings
4144	2219-T852 forgings
4145	4032-T6 forgings
4146	6061-T4 die forgings and rolled rings and 6061-F forging stock
4147	7075-T7352 forgings
4150	6061-T6 extruded wire, rod, bar, shapes and tube
4152	2024-T3 extruded wire, rod, bar, shapes and tube
4153	2014-T6 extruded wire, rod, bar, shapes and tube
4154	7075-T6 extruded wire, rod, bar, shapes and tube
4156	6063-T6 extruded wire, rod, bar, and shapes
4160	6061-O extruded wire, rod, bar, shapes and tube
4161	6061-T4 extruded wire, rod, bar, shapes and tube
4162	2219-T8511 extruded rod, bar, shapes and tube
4163	2219-T3511 extruded rod, bar, shapes and tube
4164	2024-T3510 extruded rod, bar, shapes and round tube
4165	2024-T3511 extruded wire, rod, bar, shapes, and round tube
4166	7075-T73 extruded wire, rod, bar, shapes and tube
4167	7075-T73511 extruded wire, rod, bar and shapes
4168	7075-T6510 extruded wire, rod, bar, shapes, and round tube
4169	7075-T6511 extruded wire, rod, bar, shapes, and round tube
4172	6061-T4511 extruded rod, bar, shapes and tube
4173	6061-T6511 extruded rod, bar, shapes and tube
4180	1100-H18 wire for metal spraying
4182	5056-O wire
4184	4145 brazing metal
4185	4047 brazing metal
4186	7075-F wire, rod and bar; rolled, drawn or cold finished
4187	7075-O wire, rod and bar; rolled, drawn or cold finished
4190	4043-F welding wire and rod
4191	2319-F welding wire and rod
4192	2024-T361 sheet and plate
4193	2024-186i sheet and plate
4194	Alclad 2024-T361 sheet and plate
4195	Alclad 2024-T861 sheet and plate
4196	7011 Alclad 7075-0 sheet and plate
4197	7011 Aided 7075-T6 sheet and T651 plate
4201	7050-T7651 plate
4340	7050-T76511 extrusions
4341	7050-T73511 extrusions
4342	7050-T74511 (formerly 1736511)

Continued

Appendix 2 *Continued*

Specification	Alloy and Product or Process
	American National Standards Institute (ANSI)
Specification	Alloy and Product or Process B36.10 Standard dimensions for pipe
C60.5	Aluminum alloy rigid conduit
H35.1	Alloy and temper designation systems for wrought aluminum
H35.2	Dimensional tolerances for aluminum mill products

Appendix 3A Wrought Aluminum Chemical Specifications

Alloy	Si	Cu	Mn	Mg	Cr	Ni	Zn	Pb	Bi
1050	–	–	99.5% min aluminum			–	–	–	–
1060	–	–	99.5% min aluminum			–	–	–	–
1100	–	0.12	99.5% min aluminum			–	–	–	–
1145	–	–	99.5% min aluminum			–	–	–	–
1175	–	–	99.5% min aluminum			–	–	–	–
1200	–		99.5% min aluminum						
1230	–	–	99.5% min aluminum			–	–	–	–
1235	–	–	99.5% min aluminum			–	–	–	–
1345	–	–	99.5% min aluminum			–	–	–	–
1350	–		99.5% min aluminum			–	–	–	
2011	–	5.5	–	–	–	–	–	0.4	0.4
2014	0.8	4.4	0.8	0.5	–	–	–	–	–
2017	–	4	0.7	0.6	–	–	–	–	–
2018	–	4	–	0.7	–	2	–	–	–
2024	–	4.4	0.6	1.5	–	–	–	–	–
2117	–	2.6	–	0.35	–	–	–	–	–
2124	–	4.4	0.6	1.5	–	–	–	–	–
2618	–	2.3	–	1.6	–	1	–	–	–
3003	–	0.12	1.2	–	–	–	–	–	–
3004	–	–	1.2	1	–	–	–	–	–
3005	–	–	1.2	0.4	–	–	–	–	–
3105	–	–	0.5	0.5	–	–	–	–	–
4032	12.2	0.9	–	1.1	–	0.9	–	–	–
4043	5.2	–	–	–	–	–	–	–	–
4045	10	–	–	–	–	–	–	–	–
4343	7.5	–	–	–	–	–	–	–	–
5005	–	–	–	0.8	–	–	–	–	–
5050	–	–	–	1.4	–	–	–	–	–
5052	–	–	2.5	0.25	–	–	–	–	–
5056	–	–	0.12	5.1	0.12	–	–	–	–
5083	–	–	0.7	4.45	0.15	–	–	–	–
5086	–	–	0.45	4	0.15	–	–	–	–
5154	–	–	–	3.5	0.25	–	–	–	–
5252	–	–	–	2.5	–	–	–	–	–
5356	–	0.12	5	0.12	–	–	–	–	–
5454	–	–	0.8	2.7	0.12	–	–	–	–
5456	–	–	0.8	5.1	0.12	–	–	–	–
5457	–	–	0.3	1	–	–	–	–	–
5657	–	–	–	0.8	–	–	–	–	–
6003	0.7	–	–	1.2	–	–	–	–	–
6053	0.7	–	–	1.3	0.25	–	–	–	–
6061	0.6	0.21	–	1	0.2	–	–	–	–
6063	0.4	–	–	0.7	–	–	–	–	–
6151	0.9	–	–	0.6	0.25	–	–	–	–
6253	–	–	–	1.2	0.25	–	2	–	–
6262	0.6	0.27	–	1	0.09	–	–	0.55	0.55

Appendix 3A *Continued*

Alloy	Si	Cu	Mn	Mg	Cr	Ni	Zn	Pb	Bi
6463	0.4	–	–	0.7	–	–	–	–	–
6951	0.3	0.25	–	0.6	–	–	–	Ti	Other
7005	0.35	0.1	0.20–0.7	1.0–1.8	0.08–0.20	–	4.0–5.0	0.01–.08	0.05
7008	0.1	0.05	0.05	0.7–1.4	0.12–0.25	–	4.5–5.5	0.05	0.05
7049	0.25	1.2–1.9	0.2	2.0–2.9	0.10–0.22	–	7.2–8.2	0.1	0.05
7050	0.12	2.0–2.6	0.1	1.9–2.8	0.04	–	5.7–6.7	0.06	0.05
7075	–	1.6	–	2.5	0.3	–	5.6	–	–
7108	0.1	0.05	0.05	0.7–1.4	–	–	4.5–6.5	0.05	0.05
7178	0.4	1.6–2.4	0.3	2.4–3.1	0.18–0.28	–	6.3–7.3	0.2	0.05
8017	0.1	0.1–0.20	–	–	0.01–.05	–	0.05	–	0.03
8030	0.1	0.15–0.30	–	–	0.05	–	0.05	..	0.03
8176	0.03–0.15	–	–	–	–	–	0.1	–	0.05
8177	0.1	0.04	–	0.04–0.12	–	–	0.05	–	0.03

Sources: Appendix 3A and 3B.
Alcoa Aluminum Handbook, Alcoa Aluminum, Pittsburgh, 1972.
ASM Metals Reference, ASM, Metals Park, 1983.

Appendix 3B Cast Aluminum Chemical Specifications

Alloy	Si	Fe	Cu	Mn	Mg	Cr	Ni	Zn	Ti	Sn	Other Elements Each	Other Elements Total
34	2.0	1.2	9.2–10.8	0.5	0.20–0.35	–	0.5	0.8	0.25	–	–	0.35
35	4.5–6.0	0.6	0.6	0.5	0.05	0.25	–	0.5	0.25	–	–	0.35
38	0.7–1.2	0.8	4.0–5.0	0.3	0.03	–	–	0.3	0.2	–	0.05	0.15
39	0.7	0.8	3.5–4.5	0.35	1.3–1.8	0.25	1.7–2.3	0.35	0.25	–	0.05	0.15
303	10.5–12.0	1.0	3.0–4.5	0.5	0.1	–	0.5	2.9b	–	0.2	–	0.5
304	4.5–6.0	1.0	0.6	0.35	0.1	–	0.5	0.35	–	0.15	–	0.25
305	11.0–13.0	1.0	0.6	0.35	0.1	–	0.5	0.35	–	0.15	–	0.25
308	7.5–9.5	1.0	3.0–4.0	0.5	0.1	–	0.5	2.9b	–	0.2	–	0.5
309	9.0–10.0	1.0	0.6	0.35	0.45–0.6	–	0.5	0.35	–	0.15	–	0.25
310	0.25	0.4	0.25	0.1	0.50–0.65	0.40–0.6	–	5.0–6.5	0.15–0.25	–	0.05	0.2
312	0.2	0.6	0.2	0.40–0.60	1.9–2.4	0.20–0.40	–	4.0–4.5	0.25	–	0.05	–
313	0.15	0.4	0.35–0.65	0.05	0.65–0.8	–	–	6.0–7.0	0.25	–	0.05	0.15
314	0.35	1.2	0.35–0.65	0.05	0.30–0.45	–	–	6.0–7.0	0.25	–	0.05	0.15
315	0.25	0.8	0.40–1.0	0.6	0.25–0.50	0.35	0.15	7.0–8.0	0.25	–	0.1	0.25
320	0.35	0.4	0.15	0.35	3.6–4.5	–	–	0.15	0.25	–	0.05	0.15
321	11.0–12.5	1.1	0.50–1.5	0.35	0.8–1.3	–	2.0–3.0	0.35	0.25	–	0.05	0.15
322	4.5–5.5	0.6	1.0–1.5	0.50d	0.45–0.6	0.25	–	0.35	0.25	–	0.05	0.15
323	6.5–7.5	0.50	0.25	0.35	0.25–0.40	–	–	0.35	0.25	–	0.05	0.15
324	0.2	0.2	0.2	0.1	9.6–10.6	–	–	0.1	0.2	–	0.05	0.15
326	5.5–6.5	0.8	3.0–4.0	0.5	0.1	0.35	0.35	1	0.25	–	–	50
327	7.0–8.6	0.8	1.0–2.0	0.20–0.6	0.25–0.6	–	0.25	1.5	0.25	–	–	50
329	5.5–6.5	1	3.0–4.0	0.8	0.10–0.50	–	0.5	1	0.25	–	–	50
331	8.0–10.0	0.8	3.0–4.0	0.5	0.10–0.50	–	0.5	1	0.25	–	–	
332	8.5–10.5	1	2.0–4.0	0.5	0.6–1.5	–	0.5	1	0.25	–	–	50
334	11.0–13.0	0.8	1.8–2.8	0.5	0.8–1.3	–	1	1	0.25	–	–	50
335	4.5–5.5	0.15	1.0–1.5	0.05	0.45–0.6	–	–	0.05	0.2	–	0.05	0.15
336	6.5–7.5	0.12	0.1	0.05	0.25–0.40	–	–	0.05	0.20	–	0.05	0.15
380	2.5–3.5	1	3.5–4.5	0.5	0.1	–	0.35	1	0.25	–	0.05	0.15
382	0.1	0.1	4.0–5.2	0.20–0.50	0.20–0.55	–	–	–	0.15–0.55	0.40–1.2 (Ag)	0.05	0.1
383	9.5–11.5	0.6–1.0	2.0–3.0	0.5	0.1	–	0.3	2.9	–	0.15	–	0.50

Appendix 4A Typical Properties of Wrought Aluminum Alloys

Alloy	Temper	Thermal Conductivity at 25°C		Electrical Conductivity (%IACS)		Electrical Resistivity	
		cal/cm/ cm^{2}C/s	BTU/ in/ft^2/ °F/h	Equal Weight	Equal Volume	micro- ohm-cm	ohm- cir mil, ft
1060	O	0.56	1625	62	204	2.8	17
	H18	0.55	1600	61	201	2.8	17
1100	O	0.53	1540	59	194	2.9	18
	H18	0.52	1510	57	187	3	18
1350	All	0.56	1625	62	204	2.8	17
2011	T3	0.36	1050	39	123	4.4	27
	T8	0.41	1190	45	142	3.8	23
2014	O	0.46	1340	50	159	3.4	21
	T4	0.32	930	34	108	5.1	31
	T6	0.37	1070	40	127	4.3	26
2017	O	0.46	1340	50	159	3.4	21
	T4	0.32	930	34	108	5.1	31
2018	T61	0.37	1070	40	127	4.3	26
2024	O	0.46	1340	50	160	3.4	21
	T3	0.29	840	30	96	5.7	35
	T4	0.29	840	30	96	5.7	35
	T361	0.29	840	30	96	5.7	35
	T6	0.36	1050	38	122	4.5	27
	T81	0.36	1050	38	122	4.5	27
	T861	0.36	1050	38	122	4.5	27
2117	T4	0.37	1070	40	130	4.3	26
3003	O	0.46	1340	50	163	3.4	21
	H12	0.39	1130	42	137	4.1	25
	H14	0.38	1100	41	134	4.2	25
	H18	0.37	1070	40	130	4.3	26
3004	All	0.39	1130	42	137	4.1	25
3105	All	0.41	1190	45	148	3.8	23
4032	O	0.37	1070	40	132	4.3	26
	T6	0.33	960	35	116	4.9	30
4043	O	0.39	1130	42	140	4.1	25
5005	All	0.48	1390	52	172	3.3	20
5050	All	0.46	1340	50	165	3.4	21
5052	All	0.33	960	35	116	4.9	30
5056	O	0.28	810	29	98	5.9	36
	H38	0.26	750	27	91	6.4	38
5083	O	0.28	810	29	98	5.9	36
5086	ALI	0.3	870	31	104	5.6	33
5154	All	0.3	870	32	107	5.4	32
5252	All	0.33	960	35	116	4.9	30
5454	O	0.32	930	34	113	5.1	31
	H38	0.32	930	34	113	5.1	31

Appendix 4A *Continued*

Alloy	Temper	Thermal Conductivity at 25°C		Electrical Conductivity (%IACS)		Electrical Resistivity	
		cal/cm/ cm²C/s	BTU/ in/ft²/ °F/h	Equal Weight	Equal Volume	micro- ohm-cm	ohm- cir mil, ft
5456	O	0.28	810	29	98	5.9	36
6005	T5	0.4	1160	43	142	4	24
6053	O	0.41	1190	45	148	3.8	23
	T4	0.37	1070	40	132	4.3	26
	T6	0.39	1130	42	139	4.1	25
6061	O	0.43	1250	47	155	3.7	22
	T4	0.37	1070	40	132	4.3	26
	T6	0.4	1160	43	142	4	24
6063	O	0.52	1510	58	191	3	18
	T1	0.46	1340	50	165	3.4	21
	T5	0.5	1450	55	181	3.1	19
	T6	0.48	1390	53	175	3.3	20
	T83	0.48	1390	53	175	3.3	20
6151	O	0.49	1420	54	178	3.2	19
	T4	0.39	1130	42	138	4.1	25
	T6	0.41	1190	45	148	3.8	23
6262	T9	0.41	1190	44	145	3.9	24
6463	T1	0.46	1340	50	165	3.4	21
	T5	0.5	1450	55	181	3.1	19
	T6	0.48	1390	53	175	3.3	20
7072	O	0.53	1540	59	193	2.9	18
7075	T6	0.31	900	33	105	5.2	31

Sources: Appendix 4A and 4B.
Alcoa Aluminum Handbook, Alcoa Aluminum, Pittsburgh, 1972.
Aerospace Structural Metals Handbook, Volume II, Syracuse University Press, Syracuse, 1970.
ASM Metals Reference, ASM, Metals Park, 1983.

Appendix 4B Typical Properties of Cast Aluminum Alloys

Alloy	Density Lb/in^3	Specific Gravity	Coefficient of Thermal Expansion (in/in °F)		Approximate Melting Range °F
			68–212°F	20–100°C	
1060	0.098	2.7	13.1	23.6	1195–1215
1100	0.098	2.71	13.1	23.6	1190–1215
1350	0.098	2.7	13.2	23.8	1195–1215
2011	0.102	2.82	12.7	22.9	1005–1190
2014	0.101	2.8	12.8	23	945–1180
2017	0.101	2.79	13.1	23.6	955–1185
2018	0.101	2.8	12.4	22.3	945–1180
2024	0.1	2.77	12.9	23.2	935–1180
2117	0.099	2.74	13.2	23.8	1030–1200
2618	0.1	2.76	12.4	22.3	1040–1200
3003	0.099	2.73	12.9	23.2	1190–1210
3004	0.098	2.72	13.3	23.9	1165–1210
3105	0.098	2.71	13.1	23.6	1175–1210
4032	0.097	2.69	10.8	19.4	990–1060
4043	0.097	2.69	—	—	1065–1170
4045	—	—	—	—	1065–1110
4343	—	—	—	—	1065–1135
5005	0.098	2.7	13.2	23.8	1170–1210
5050	0.097	2.69	13.2	23.8	1155–1205
5052	0.097	2.68	13.2	23.8	1125–1200
5056	0.095	2.64	13.4	24.1	1060–1180
5083	0.096	2.66	13.2	23.8	1075–1185
5086	0.096	2.66	13.2	23.8	1085–1185
5154	0.096	2.66	13.3	23.9	1100–1190
5252	0.097	2.68	13.2	23.8	1125–1200
5356	0.096	2.64	—	—	1065–1180
5454	0.097	2.68	13.1	23.6	1115–1195
5456	0.096	2.68	13.3	23.9	1060–1180
5457	0.097	2.68	13.1	23.6	1165–1210
5657	0.098	2.7	13.1	23.6	1180–1215
6005	0.098	2.7	13.1	23.6	1180–1205
6053	0.097	2.69	12.8	23	1070–1205
6061	0.098	2.7	13.1	23.6	1080–1205
6063	0.098	2.7	13	23.4	1140–1210
6151	0.098	2.7	12.9	23.2	1090–1200
6262	0.098	2.12	13	23.4	1080–1205
6463	0.098	2.7	13	23.4	1140–1210
6951	0.098	2.7	13	23.4	1140–1210
7049	0.103	2.84	13	—	890–1175
7050	0.102	2.83	13.4	—	910–1165
7072	0.098	2.72	13.1	23.6	1190–1215
7075	0.101	2.8	13.1	23.6	890–1175
7178	0.102	2.83	13	—	890–1165
8017	0.098	2.71	13.1	—	1190–1215
8030	0.098	2.71	13.1	—	1190–1215
8176	0.098	2.71	13.1	—	1190–1215

Index