Handbook of Crystal Growth

Fundamentals: Thermodynamics and Kinetics

VOLUME I, Part A

Second Edition

Editor-in-Chief and Volume Editor

Tatau Nishinaga

The University of Tokyo

Hongo, Bunkyo-ku, Tokyo, Japan

ELSEVIER

Amsterdam • Boston • Heidelberg • London
New York • Oxford • Paris • San Diego
San Francisco • Singapore • Sydney • Tokyo

Handbook of Crystal Growth

Fundamentals: Transport and Stability

VOLUME I, Part B

Second Edition

Editor-in-Chief and
Volume Editor

Tatau Nishinaga

The University of Tokyo

Hongo, Bunkyo-ku, Tokyo, Japan

ELSEVIER

Amsterdam • Boston • Heidelberg • London
New York • Oxford • Paris • San Diego
San Francisco • Singapore • Sydney • Tokyo

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
225 Wyman Street, Waltham, MA 02451, USA

First edition 1993

Second edition 2015

The crystals on the front cover are high-quality synthetic diamond crystals grown by Sumitomo Electric Industries, LTD by a temperature gradient method under high pressure and high temperature. The yellowish crystals are ordinarily synthetic diamonds containing nitrogen impurity of several tens of ppm (type Ib). The colorless ones are high-purity synthetic diamonds free of impurities (type IIa). The size of the largest high-purity crystal (lower left) is about 12 mm in diagonal length. The photo was provided by Dr. Hitoshi Sumiya.

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

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

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

SET Volume ISBN: 978-0-444-56369-9
Volume 1, Part A–978-0-444-63322-4
Volume 1, Part B–978-0-444-63323-1

General Preface

The history of crystal growth is long as those of the universe and the earth. Meteorites contain pyrites and olivine crystals which indicate these crystals were grown when the planets were born. Crystals naturally produced are used as gems from the early time of the human history. In Exodus, it is written that breast-piece was decorated by ruby, emerald, sapphire, amethyst, and other gems.

There are a lot of crystals around us. As examples, we can find snowflakes falling down from the sky, ice crystals in a lake in winter, salt and sugar crystals in the pots of our kitchen. But, it was after the invention of point contact and junction transistors, respectively in 1947 and 1948, that the industry paid a great interest on the crystal growth. Without the growth of high purity and highly perfect single-crystal semiconductor, at that time of Ge, the invention of the transistors will never happen.

It is well known that the modern information society will be not realized without electronic and optical devices. One finds large-scale integrated circuits of Si in every computer from laptop to super computers. For high speed and mass transmission of information, compound semiconductor devices are indispensable.

These devices are fabricated almost all by using single crystals of semiconductors and oxides. When we look into the history of the devices, we always see that an invention of crystal growth technique makes it possible to bring out new device. As we saw, the invention of transistor was possible only after the growth of high-quality Ge single crystal. The growth of large-diameter dislocation-free Si crystal has enabled the production of large-scale integrated circuit. Due to the invention of liquid-phase epitaxy, it became possible to realize light-emitting diode (LED) and laser diode (LD) in real use. Drastic technological improvement in highly lattice mismatch heteroepitaxy made it possible to realize blue $\sim$ ultraviolet LED and LD and it can be said that the success in the growth of high-quality nitride semiconductor gave the blue light all over the world. Hence, we should understand that new technology of crystal growth has always created new electronic and optical devices.

It is extremely good news for the community of crystal growth that 2014 Nobel Prize in Physics was awarded to Professors Isamu Akasaki, Hiroshi Amano and Shuji Nakamura for the invention of efficient blue light-emitting diodes which has enabled bright and energy-saving white light sources. This invention is basing on the growth of nitride semiconductors employing a low-temperature buffer layer on sapphire substrate in heteroepitaxy. We are happy that Professor Hiroshi Amano, one of the winners, is contributing to this Handbook as an author of Chapter 16 in Vol. IIIA.

The first edition of the Handbook of Crystal Growth was edited by D.T.J. Hurle. This Handbook was composed of three volumes and published in 1993–1994. The present second edition of the Handbook also consists of three volumes. Each volume was edited by separate editors. Volume I is edited by T. Nishinaga and the volume covers the basic aspects of crystal growth. In Volume IA, fundamentals and kinetics of crystal growth are described and in IB, advanced problems of transport and stability are discussed. Volume II is edited by P. Rudolph and this volume covers bulk crystal growth. Volume IIB presents basic technologies of bulk growth and IIB does growth mechanism and dynamics. Volume III was edited by T. F. Kuech and the volume covers thin film growth and epitaxy. Volume IIIA discusses basic techniques and IIIB does growth mechanisms and dynamics.

Present Handbook project was created in March, 2011 and six advisors were appointed. They are T. F. Kuech, G. B. Stringfellow, J. B. Mullin, J. J. Derby, R. Fornari, and K. H. Ploog. I am very much grateful for their important and valuable suggestions.

Finally, all editors would like to express their sincere thanks to Shannon Stanton, Elsevier, for her strong and well cared support to this work.

Tatau Nishinaga
(Editor in Chief)

Preface to Volume I

Crystal growth has three faces. One is the face of science, the second is that of art, and the final one is of technology. For a long time in the human history, people have wondered how snowflakes grow in such beautiful shapes and tried to understand the mechanism. The curiosity is the driving force for the growth science. Crystal growth is an important tool to obtain useful crystals for human life and industry use. The growth of ruby is one of the examples. People have developed the art of the growth and the art was improved by many workers and finally Verneuil has arrived at an elegant growth method called as Verneuil method.

For a long time, science and art of crystal growth have been developed separately such as in school and in small factory in town. However, after the crystal growth has been employed to fabricate advanced electronic and optoelectronic devices, the art based on science is strongly required. For the control of the accurate structure and dimension of such devices, the growth should be carried out with deep understandings of growth science.

A half century ago, nobody thought that the real-time observation of 2D nucleation is possible in molecular beam epitaxy (MBE) during the growth of quantum well laser by reflection high-energy electron diffraction (RHEED). It was only in recent years that growth spirals were found on the surface of GaN grown by metal organic chemical vapor deposition (MOCVD) to be used for the fabrication of blue light-emitting diode (LED) and laser diode (LD). It was shown that the relationship between growth rate and surface supersaturation of GaN MOCVD is explained very well by classical Burton–Cabrera–Frank (BCF) theory. In this experiment, a mask epitaxy using photolithography was employed.

In the history of crystal growth, there has been no age when the art and the science of crystal growth are so closely combined like today. However, there is still strong contribution of art required in the crystal growth technology. For instance, to grow the advanced devises by MBE, one should be very skillful to manipulate the machine, however, with advanced knowledge of growth science. This is what is different from the old art of crystal growth.

The purpose of the Volume I is to show the recent advances in the growth science and to give scientific bases for the technologies to be developed in the following Volumes II and III, which are devoted to the bulk crystal growth and thin film growth and epitaxy, respectively.

Volume Ia describes thermodynamics and kinetics and Volume IIb does the transport and stability. The first chapter of Volume Ia gives a historical introduction of the crystal growth especially for the beginners. This chapter is followed by those of phase equilibria,

defect thermodynamics, and stoichiometry. Then, Chapter 5 discusses the equilibrium shape of crystal and Chapter 6 does rough–smooth transition of step and surface. Both chapters aim at giving the true picture of the crystal surface. Chapters 7 and 8 will cover the most fundamental and basic aspects of crystal growth, nucleation and growth kinetics, respectively. Chapter 9 is devoted to explain the structure of melt and liquid alloys. To understand the growth from the melt, one should have the knowledge about the atomistic structure of the melt. Next three chapters discuss the simulation of crystal growth employing classical and quantum mechanically calculated potentials. The final chapter presents the colloid crystal growth, which provides the experimental modeling for the crystal growth.

The first chapter of Volume Ib gives a general introduction to morphological stability that is followed by Chapter 15, in which the modern theory of morphological stability, i.e., phase-field model, is explained and applied to solidification to understand microstructure formation processes. The next two chapters describe the experiments related to the morphological stability. In Chapter 16, the detailed theoretical and experimental studies of dendritic growth are presented. On the other hand, in Chapter 17, grain growth in the melt is discussed and it is demonstrated that the dendritic growth is often observed in grain growth. Nanocrystal growth is one of the rapidly expanding fields. Growth of nanocrystals from vapor is discussed as an example in Chapters 18. Crystal growth of protein and other biological molecules are studied very extensively to facilitate the advancement of life science. Chapters 19 and 20 are devoted to this subject. The following two chapters discuss the problems which one encounters in producing medicine. Chapter 21 describes the fundamental growth process of pharmaceutical crystallization, which is exactly the same as ordinary crystal growth. Chapter 22 discusses the growth of chiral molecules. Selective growth of one type is especially important for pharmacy production.

Chapters 23–25 describe in situ observation of crystal growth in vacuum, solution, and melt. Chapters 24 and 25 are devoted some parts to the growth in space. The final chapter describes the growth of quasicrystal which shows symmetries forbidden in ordinary crystallography. It is possible to grow single-grained quasicrystals and their properties were studied in detail.

The present editor wishes to acknowledge deeply all authors of Volume I for their excellent articles. The mails of the request were sent in the autumn of 2012 and the deadline was the end of October 2013. But, the most of authors only could send their manuscript by March of 2014 and some did in the beginning of July, 2014. I would like to thank all authors for sacrificing many hours of their important official and private time.

Tatau Nishinaga
Editor of the Volume I
July, 2014

List of Contributors

Noriko Akutsu Faculty of Engineering, Osaka Electro-Communication University, Neyagawa, Osaka, Japan

Pedro Cintas Depto. Química Orgánica e Inorgánica, Universidad de Extremadura, Badajoz, Spain

Gérard Coquerel Normandie Université, Université de Rouen, Mont-Saint-Aignan, Cedex, France

Sam R. Coriell National Institute of Standards and Technology, Gaithersburg, MD, USA

Can Cui Department of Physics, Center for Optoelectronic Materials and Devices, Zheijiang Sci-Tech University, Hangzhou, China

T.L. Einstein Department of Physics and Condensed Matter Theory Center, University of Maryland, College Park, MD, USA

Giuseppe Falini Dipartimento di Chimica "G. Ciamician," Alma Mater Studiorum Università di Bologna, Bologna, Italy

Robert S. Feigelson Geballe Laboratory for Advanced Materials, Stanford University, Stanford, CA, USA

Kozo Fujiwara Institute for Materials Research, Tohoku University, Sendai, Miyagi, Japan

Takashi Fukui Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan; Research Center for Integrated Quantum Electronics (RCIQE), Hokkaido University, Sapporo, Japan

Yoshinori Furukawa Institute of Low Temperature Science, Hokkaido University, Sapporo, Hokkaido, Japan

Juan Manuel García-Ruiz Laboratorio de Estudios Cristalográficos, Instituto Andaluz de Ciencias de la Tierra, CSIC-Universidad de Granada, Granada, Spain

Jean-Pierre Gaspard Solid state physics, interfaces and nanostructures, University of Liege, Sart-Tilman, Belgium

George H. Gilmer Mechanical Engineering Department, Colorado School of Mines, Golden, CO, USA

Martin E. Glicksman Department of Mechanical and Aerospace Engineering, Florida Institute of Technology, Melbourne, FL, USA

Jaime Gómez-Morales Laboratorio de Estudios Cristalográficos, Instituto Andaluz de Ciencias de la Tierra, CSIC-Universidad de Granada, Granada, Spain

Yoshikazu Homma Department of Physics, Tokyo University of Science, Shinjuku, Tokyo, Japan

Tomonori Ito Department of Physics Engineering, Mie University, Kurima-Machiya, Tsu, Japan

Yoshihiro Kangawa Research Institute for Applied Mechanics, Kyushu University, Kasuga-Koen, Kasuga, Japan

Detlef Klimm Leibniz Institute for Crystal Growth, Berlin, Germany

Xiang Yang Liu Department of Physics and Department of Chemistry, National University of Singapore, Singapore

Geoffrey B. McFadden National Institute of Standards and Technology, Gaithersburg, MD, USA

Wolfram Miller Leibniz Institute for Crystal Growth, Numerical Modelling, Max-Born-Str. 2, 12489 Berlin

Christo N. Nanev Rostislaw Kaischew Institute of Physical Chemistry, Bulgarian Academy of Sciences, Acad. G. Bonchev Str. Bl. 11, 1113 Sofia, Bulgaria

Mathis Plapp Physique de la Matière Condensée, École Polytechnique, CNRS, Palaiseau, France

François Puel LAGEP UMR CNRS 5007, Université de Lyon, ESCPE Lyon, Villeurbanne Cedex, Lyon, France

Robert F. Sekerka Department of Physics, Carnegie Mellon University, Pittsburgh, PA, USA

Talid Sinno Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA, USA

Katsuhiro Tomioka Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan; Research Center for Integrated Quantum Electronics (RCIQE), Hokkaido University, Sapporo, Japan; Japan Science and Technology Agency (JST), PRESTO, Kawaguchi, Japan

An-Pang Tsai Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Sendai, Japan; National Institute for Materials Science, Tsukuba, Japan

Katsuo Tsukamoto Graduate School of Science, Tohoku University, Aramaki Aoba, Sendai, Japan

Satoshi Uda Institute for Materials Research, Tohoku University, Katahira, Aoba-ku, Sendai, Miyagi, Japan

Makio Uwaha Department of Physics, Nagoya University, Chikusa-ku, Nagoya, Japan

Stéphane Veesler CINaM-CNRS, Aix-Marseille Université, Marseille, France

Peter G. Vekilov Department of Chemical and Biomolecular Engineering, and Department of Chemistry, University of Houston, Houston, Texas, USA

Cristóbal Viedma Depto. de Cristalografía y Mineralogía, Universidad Complutense, Madrid, Spain

Takao Yamamoto Division of Pure and Applied Science, Faculty of Science and Technology, Gunma University, Kiryu, Gunma, Japan

Luis A. Zepeda-Ruiz Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA

Tian Hui Zhang Center for Soft Condensed Matter Physics and Interdisciplinary Research, Soochow University, Suzhou, China; School of Physical Science and Technology, Soochow University, Suzhou, China

1

Crystal Growth through the Ages: A Historical Perspective

Robert S. Feigelson

GEBALLE LABORATORY FOR ADVANCED MATERIALS, STANFORD UNIVERSITY, STANFORD, CA, USA

Excerpt from Pliny the Elder's, *Natural History*, AD 77, "*… I have prefaced these volumes with the names of my authorities. I have done so because it is, in my opinion, a pleasant thing and one that shows an honourable modesty, to own up to those who were the means of one's achievements…*"

Trans. H. Rackham

CHAPTER OUTLINE

1.1 Introduction **2**
1.2 Evolution of Crystal Growth Theories **4**
1.2.1 Early Developments (Before the Nineteenth Century) 4
1.2.2 The Nineteenth Century 5
1.2.3 The Twentieth Century 6
1.3 Crystal Growth Methods **16**
1.3.1 Melt Growth 17
1.3.1.1 The Nineteenth Century and the Verneuil Process 18
1.3.1.2 The Twentieth Century 20
1.3.1.3 The Czochralski Crystal Pulling Method 20
1.3.1.4 Bridgman–Stockbarger/Gradient Freeze Methods 25
1.3.1.5 Nacken–Kyropoulos Methods 28
1.3.1.6 Stöber/Heat Exchanger Methods 30
1.3.1.7 Kapitza's Method 30
1.3.1.8 Zone Melting 31
1.3.1.9 Shaped Growth 33
1.3.1.10 Skull Melting 37
1.3.2 Solution Growth **38**
1.3.2.1 Introduction 38

Handbook of Crystal Growth. http://dx.doi.org/10.1016/B978-0-444-56369-9.00001-0

1.3.2.2 Aqueous Solution Growth 38
1.3.2.3 Growth of Biological Macromolecules 41
1.3.2.4 Growth from Gels 43
1.3.2.5 Nonaqueous Solution Growth 44
1.3.2.6 High Temperature Solution (Flux) Growth 44
1.3.2.7 Hydrothermal Growth 49
1.3.2.8 Electrochemical Crystal Growth 52
1.3.3 Vapor Growth 53
1.3.3.1 Introduction 53
1.3.3.2 Bulk Growth 55
1.3.3.3 Vapor Phase Epitaxy 58
1.3.3.4 The VLS Method 63
1.3.3.5 Artificial Epitaxy (Graphoepitaxy) 64
1.3.4 Synthetic Diamond Crystals 65
1.3.5 Solid State Recrystallization 70
1.4 Epilogue **72**
Acknowledgments **72**
References **73**

1.1 Introduction

Most scientists and engineers are aware that the basic tools they use in their daily research activities were developed by groups of researchers working in series or parallel over decades, in some cases centuries. It was often in an incremental way, one study building on another. As time went on, these earlier ideas became more refined and practical, providing future workers with a more through understanding of the physics and chemistry involved in different materials systems and leading to innovative new processes for making materials and devices that have affected everyone's lives. They helped define the world we live in and used their newly gained knowledge to stoke the technological revolution.

The crystal growth field (a branch of materials science, physics, chemistry and crystallography) has a rich historical background that goes back at least several millennia. It basically deals with understanding the underlying mechanisms involved in the crystallization process and the technology to produce a single crystal from some medium in a controlled fashion. One of the earliest written accounts documenting work on methods for preparing crystals was given by the Roman Pliny the Elder in 77–79 AD [1]. His collected work was a summary of knowledge going back to even more ancient times. It is probable that even prehistoric man engaged in the recrystallization of materials like salt. Among other topics, Pliny discussed the preparation of Vitriol (iron,

copper and zinc sulfate hydrates). The process involved evaporating mine or spring waters obtained from the Mediterranean region. About 1500 years later, in medieval times, Pliny's work was referred to by both Biringuccio [3] and Agricola [2]. They concentrated on preparing crystals of these compounds for medicines, dyes, fluxes and acids. The various methods employed generally began with the purification of mineral deposits, followed by recrystallization of the remaining solutions by evaporation.

The field of crystal growth encompasses a wide spectrum of scientific disciplines and includes (1) experimental and theoretical studies of crystallization processes, (2) the growth of crystals under controlled conditions for both scientific purposes and industrial applications and (3) crystal characterization. It also covers almost all classes of materials, i.e., inorganic and organic compounds, elemental materials as well as biological macromolecules. Many methods have been developed over the years for producing single crystals, the size range for which varies from the nanometer to meter scale. These crystals have in common an atomic ordering that persists throughout their bulk and without the presence of grain boundaries. The two principal scientific pillars upon which the field of crystal growth depends are thermodynamics and kinetics. The thermodynamic properties of a system describe how solid, liquid and gaseous phases behave with respect to state variables such as temperature, pressure and composition. They provide a road map, so to speak, which crystal growers use to plan growth strategies. For the preparation of crystals of a size, purity and composition required for a specific application, one needs to know what material phases will exist under various conditions of temperature, T, and pressure, P, etc., and how these phases will form under dynamic solidification processing conditions. Kinetic factors, on the other hand, influence our ability to produce a crystal at a desired growth rate and with a degree of perfection and uniformity suited to the intended application. We will explore below how interface stability and segregation behavior are influenced both by thermodynamic and kinetic factors.

In the beginning, crystal growth was not the well-defined field it is today. Work was carried out by chemists, physicists, etc., and research results were reported in various conferences and journals of these societies. The first conference to concentrate on the topic was at a Faraday Society meeting in 1949, held in Bristol, England. In spite of the growing importance of crystals for solid-state electronic applications in the early 1950s, it was almost a decade later before a second meeting concerning issues in crystal growth arose. That conference, held in Cooperstown, New York in 1958 [4], gathered together some of the most eminent crystal growth researchers to discuss a wide range of topics of interest to the crystal growth community. Conferences were also started in the Soviet Union (Moscow) as early as 1956. However, the major consolidation of the field into a viable entity was the formation in 1966 of the International Organization of Crystal Growth (IOCG) and under their aegis, the subsequent International Conferences on Crystal Growth (ICCG). These conferences have been held every three years since 1966.

The local organizers of the first ICCG conference held in Boston, Massachusetts immediately founded the American Association for Crystal Growth (AACG) under the joint chairmanship of Doctor Robert (Bob) Laudise and Doctor Kenneth (Ken) Jackson.

The AACG held their own national conferences soon afterward and other national groups formed around the world. The *Journal of Crystal* was established in 1967 under the leadership of Professor Michael Schieber, along with Sir Charles Frank and Dr Nicholas Cabrera as co-editors. Although papers on crystal growth topics are published elsewhere as well, the *Journal of Crystal Growth* has remained the major venue for papers on crystal growth theory, practice and characterization and has published related proceedings of conferences focused on various aspects of the field.

Many of the topics discussed in this introductory history are covered in much more detail in various chapters in this comprehensive, updated version of the *Handbook of Crystal Growth*. This treatment is designed to focus mainly on their historical context.

1.2 Evolution of Crystal Growth Theories

Although crystals can be grown by purely empirical means, control of their rate of growth, perfection, dimensions, composition and physical properties is greatly facilitated by having a good grasp of the fundamentals underlying crystal growth processes. Over the past century, a sound theoretical foundation has been built up through the efforts of many different scientists and engineers working in materials-related fields such as chemistry, physics and crystallography. The approach is generally two-fold: first to understand the nature of material systems (crystal structure and morphology, phase equilibrium, etc.), and second, to determine the factors that affect the crystallization process (nucleation, growth kinetics, segregation behavior, interface stability, heat and mass transport, etc.). Although remarkable progress has been made, the complex nature of the field and its changing emphasis on newer materials and structures keeps providing a constant source of challenges to our understanding of crystallization processes.

1.2.1 Early Developments (Before the Nineteenth Century)

The earliest scientific studies important to the field of crystal growth were made by natural scientists trying to understand the morphologies of mineral crystals. One of these early pioneers was the Swiss naturalist Conrad Gesner (1516–1565) who in 1564, after studying different crystals, reported that one crystal differs from another by its angles and form [5]. Later in the sixteenth century, Andreus Caesalpinus (1519–1603) wrote in "De Metallicis" [6] that the shape of crystals grown from water solution (e.g., salt, sugar and alum) were a characteristic of the material. Ichiro Sunagawa [7] proposed, however, that the science of crystal growth started with the treatise of N. Steno. Nicolas Steno, also known as Niels Stensen, (1638–1686) was a well-known Danish scientist specializing in the fields of geology and anatomy. He was also one of the founders of crystallography. In his treatise, published in 1669 [8], he observed that, although quartz crystals differ in appearance from one to another, the angles between corresponding faces are always the same. In addition, he noted that they grew by an inorganic hydrothermal process rather than through the action of bacteria [7]. Years later, Steno's law of *constant interfacial*

angles in crystals was confirmed, first by the Italian Domenico Guglielmini (1665–1710) [9] who asserted, like Casealpinus, that every salt has it own particular shape. A century later, the Frenchman Jean Baptiste Romé de l'Isle (1736–1790) [10], concluded from his study of many hundreds of different crystals that every crystalline substance with a specific composition had a similar and particular crystal shape (1772) See Figure 1.1(A)). He found six different fundamental forms from which all others could be derived. Although the above work, and that of other researchers not mentioned, set the stage for our improved understanding of the nature of crystals, it was not until much later that attention turned seriously to the question of how crystals grew and which mechanisms were involved. Figure 1.1(B) shows the internal structure of a lithium niobate crystal revealed by partial melting.

1.2.2 The Nineteenth Century

French physicist Auguste Bravais (1811–1853), building on l'Isle's previous work, determined in 1848 that there are 14 unique "Bravais" lattices comprising three-dimensional crystalline systems [11]. This work provided the basis for understanding symmetry, crystal morphology and crystalline anisotropy. The morphology of a crystal is influenced by (1) external factors, e.g., the surrounding nutrient phase and (2) internal features, e.g., cell dimensions, atom sizes, positions, and bond energies.

Contemporary quantitative crystal growth science originated with the thermodynamic studies of the American scientist J. Willard Gibbs (1839–1903). Gibbs studied how various phases behaved in heterogeneous systems under the influence of state variables such as temperature and pressure. His seminal work, *On the Equilibrium of*

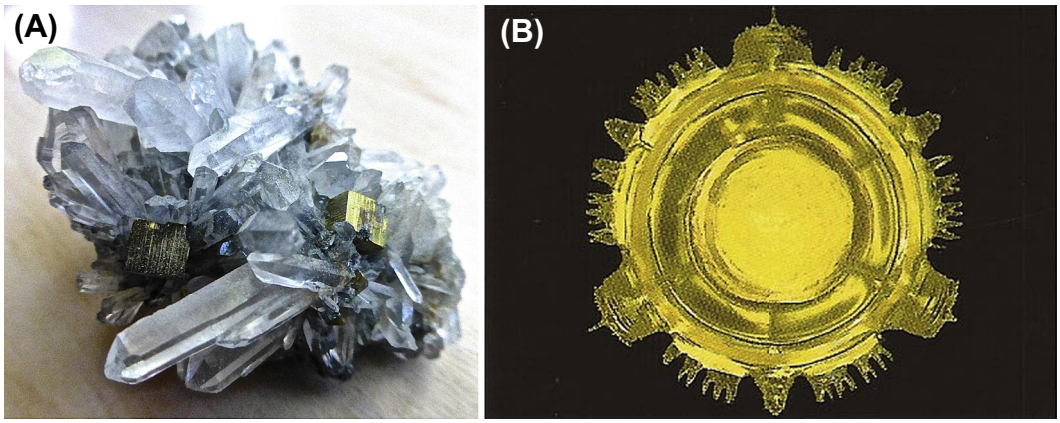


FIGURE 1.1 (A) Naturally occurring crystals of quartz (SiO_2) interspersed with pyrite (FeS_2) crystals. Their different morphologies reflect their internal crystal structures, trigonal and cubic respectfully. (B) The bottom side of a c-axis Czochralski grown lithium niobate crystal that was rapidly heated to cause it to separate from the melt surface. The resulting dendritic-like structure reveals the internal three-fold symmetry along the axis of this rhombohedral crystal.

Heterogeneous Substances (1876 and 1878) [12], included both the first and second laws of thermodynamics and thermodynamic reaction tendencies in any thermodynamic system. His graphical representations, the earliest phase diagrams, have been expanded through the years to include numerous chemical systems of both academic and industrial interest. These were derived largely by experimental studies but also in recent years by numerical simulations. These "phase diagrams" are vital data sources for the crystal grower, enabling him or her to select the most appropriate growth method and produce a crystal with the desired composition and properties.

To form a crystal, a nutrient phase (i.e., liquid, gas or solid) must be in a metastable state. In other words, the free energy (at constant volume) or the Gibbs potential (at constant pressure) of this phases must exceed that of the crystal. This excess is the driving force for crystallization. This metastability is accomplished by either supercooling a melt or supersaturating a solution or vapor phase. During crystallization, latent heat is evolved. Among other remarkable contributions made by Gibbs was that nucleation phenomena resulted from heterophase fluctuations in metastable homogeneous phases. Nucleation can be either homogeneous (from within the pure matrix phase) or heterogeneous (on a foreign substance such as particles or substrates within the matrix phase or the container walls). The maximal amount of supercooling or supersaturation required depends on the thermodynamic properties of the material system, various external forces such as mechanical vibrations, and of course, the nature of the crystal surface, etc. In practice, the initial nucleation stage is often bypassed by using oriented seed crystals.

1.2.3 The Twentieth Century

In the opinion of K. Jackson [13], our modern understanding of crystal growth processes began with the research work of Harold Wilson (1874–1964) [14] and Martin Knudsen (1871–1949) [15]. Wilson's work in 1900 concerned the velocity of solidification and viscosity of supercooled liquids, whereas Knudsen's work involved kinetic molecular theory that much later played an important role in molecular beam epitaxy. One of the most important early growth theories was proposed in 1921 by the German physical chemist Max Volmer (1885–1965) and his student Immanuel Estermann (1900–1973) [16]. Their adsorption-layer theory (i.e., layer-by-layer growth) was deduced from measuring the tangential growth rate of plate-like mercury crystals from the vapor state at low temperatures. The proposed adsorption-layer lies between the crystal and nutrient phase, with the crystallizing species losing only part of their latent heat, while maintaining some surface mobility in the layer parallel to the crystal surface. The species are incorporated into the crystal lattice at the edges of the incomplete atomic layers (steps on the growing crystal face). Volmer was also the first to consider the role of ad-atoms (or molecules) and holes on the crystal surface under equilibrium and nonequilibrium conditions.

Walther Kossel (1888–1956) [17], a German physicist known for his theory of chemical bonding, proposed in 1928 an atomistic view of crystal growth (kinetic theory), as

opposed to a "continuum" thermodynamic interpretation. It was similar to that independently proposed by Iwan N. Stranski (1897–1979), [18] a Bulgarian physical chemist, and was based on earlier diffusion theories concerning mass transport of the crystallizing species to the growth interface with the distinction that what went on in the interface region (how the species found an appropriate lattice site) was *not* a negligible effect. Their work is often linked together as the Kossel–Stranski model. They both concluded from early work on the rock salt structure that no other planes but the cubic ones are possible and that other planes (110, 111, etc.) are not present on the surface as complete planes but are made up of alternating (001) and (100) faces several atoms thick (kinetic roughening). This work led to what is commonly referred to as the TLK (terrace-ledge-kink) model where Kossel [19] suggested that incorporation of an atom required that the steps spread laterally across the surface. Somewhat later came the work of Stranski's younger colleague Rostislav Kaishev (1908–2002) linking the equilibrium crystal shape, i.e., the facets making this shape, with the average work required to detach a molecule from that facet, and thereby accounting for different structural positions on that facet and its edge. Stranski and Kaishev founded the famous Bulgarian school of nucleation and crystal growth (see Ref. [20]). Much of their work was on low-temperature aqueous solution growth and the crystallization of metals at room temperature in electrolyte solutions. An extensive discussion of Kossel and Stranski's work, together with other contemporaries, is given in Buckley's book *Crystal Growth* [21] and numerous other more recent publications.

The goal of scientific studies is the development of effective models that can explain observable physical phenomena and direct practical crystal growing via generalized predictive relationships. These activities were both based on scientific inquisitiveness and to provide guidelines for practitioners to produce material for the benefit of mankind. Basic studies on nucleation and crystal growth have greatly expanded over the years. Older theories and concepts have been refined and new concepts proposed and tested. Basic understanding has greatly benefited from important advances in crystal characterization technologies. They have provided direct evidence of crystal perfection and growth behavior down to the atomic scale. Two examples are the transmission electron microscope and in situ atomic force microscopy. The former technology makes possible the imaging of atomic structures of real crystals, allowing a study of their perfection and the nature of their imperfection. Atomic force microscopy can be used, to great effect, both to study the formation and kinetics of growth layers during solution growth (particularly biological macromolecules) and how they change upon post-growth heat treatments (surface reconstruction). Reflected beam electron microscopy has also been very useful.

The discovery of crystalline imperfections such as edge and screw dislocations, stacking faults, point defects and inclusions in an otherwise uniform crystal lattice, has had a strong impact on our understanding of crystal properties, on the one hand, and crystal growth mechanisms on the other. They are also of great technological importance for the influence they have on the electronic and mechanical properties of a material.

In 1934, Sir Geoffrey Taylor (1886–1975), a noted British physicist and mathematician, proposed that the *plastic* deformation of ductile materials could be explained in terms of the theory of dislocations developed by Vito Volterra in 1905. Some years later, the subject of dislocations occupied the thoughts of Sir Charles Frank (1911–1998), an eminent British crystallographer who spent much of his career at Bristol University. His fundamental contributions to the field of crystal growth include the laws governing dislocation branching, the existence and properties of dislocation networks, and in 1950, the Frank-Read mechanism for the generation of dislocations. In a well-documented account, the idea for this latter mechanism occurred to both Sir Charles and to W.T. Read (an American working at the General Electric Co.) independently and at the same time. Frank had shown the year before [22] that two-dimensional nucleation theory failed significantly to explain observed high crystal growth rates at low supersaturation. This discrepancy could, however, be readily reconciled if the growth face contained a screw dislocation outcrop. That this dislocation should lead to continuous step generation in the form of a "growth spiral" step on the growth face was immediately validated by experimentally observed growth spirals formed on actual crystals (i.e., Refs [23,24]).

Some of the important work on crystal symmetry in modern times was done by Donnay and Harker in 1937 [25] and later by Hartman and Perdok [26]. Hartman and Perdok's theory [26] classified different types of faces, with only one type forming crystal facets. For ionic crystals, they defined the energy released during growth of a layer as E(hkl) and were able to generate growth forms by assuming that E(hkl) was proportional to the growth rate. These calculated forms were similar to natural or manmade crystals such as zircon, garnets, etc. Many researchers before and since have also observed variations from predicted or expected crystal morphologies due to impurity adsorption on a growth face. That led to methods to alter the morphologies for a specific application, one example being the purposeful poisoning of a fast-growing needle axis to make a more equiaxed crystal. Other notable contributions to our understanding of growth shapes include those of Sunagawa (1960) [7] and Bennema (1980) [27].

A major effort to control the purity and dopant uniformity in Si and Ge electronic devices was begun at Bell Laboratories in the early 1950s. The research team of Burton, Prim and Slichter came up with a relationship that described how impurities and dopants are distributed along an as-grown boule (the now well-known BPS equation). Their work was first reported in 1952, but not openly published until 1953 [28]. Measuring solute concentrations, solid–liquid distribution coefficients, diffusion coefficients and solute distributions in actual crystals, they derived equations describing what the concentration of a dopant or impurity would be in an as-grown crystal as a function of its initial melt concentration and growth rate R.

$$K_e = K_o/K_i + (1 - K_o)\exp(-R\delta_D/D) \tag{1.1}$$

where K_e is the effective segregation coefficient, K_o is the interface or equilibrium segregation coefficient, R the growth rate, δ_D is the diffusion boundary layer thickness and

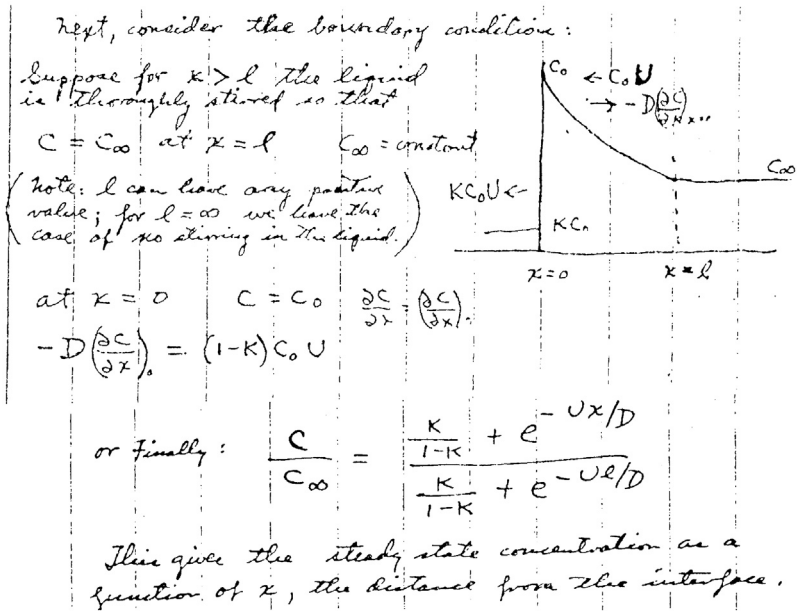


FIGURE 1.2 One of the diagrams in J.A. Burton's laboratory notebook (1951) explaining the concept of the BPS theory. The plot shows the variation of solute concentration in the solid and melt during unidirectional solification. The $x = 0$ position is the growth interface while $x = l$ defines the width of the solute boundary layer [29].

D is the solute diffusion coefficient. Figure 1.2 shows a sketch and notes from Burton's laboratory notebook illustrating the features of the now famous segregation relationship.

This research was part of a larger effort amongst various semiconductor scientists and engineers at Bell Laboratories [29,30]. Since then, BPS segregation theory has become a particularly valuable tool for crystal growth practitioners, allowing them to control the compositional uniformity in single- and multi-component material systems via control of growth velocity, melt composition, fluid convection, etc. While BPS theory is useful for well-behaved systems, it does not work in all cases. A critical analysis of the limitations of the BPS theory and later modifications by various scientists was given by Carruthers [31] (from the standpoint of the mother liquor hydrodynamics) and by Chernov [32] (from the standpoint of equilibrium and nonequilibrium processes at the growing interface). See also Handbook chapter "Segregation and Component Distribution" for a description of the limitations of BPS segregation theory.

As mentioned in the introduction, the Faraday Society in 1949 convened one of its meetings for the sole purpose of presenting and discussing papers on crystal growth. This was the first scientific conference devoted to this topic as a separate subject. Subsequently, the proceedings of this conference, entitled *Crystal Growth*, were published in the Discussions of the Faraday Society. During that meeting, Burton and Cabrera [33] presented their research on the influence of surface structure on the rate of

growth of a perfect crystal. They considered two types of surface structures, one ordered (atomically flat) and the other disordered (stepped, containing terraces, ledges and kinks with ledge heights of atomic dimensions). Burton and Cabrera made use of the analogy between the existence of these two surface structure types and two possible states in two-dimensional systems-ordered and disordered *phases*, transferring from one to another as the temperature/binding energy ratio changes. This concept followed the 1945 suggestion by Yakov I. Frenkel that the step should be disordered and possess a kink configuration at nearly each atomic site. As a result, these steps should grow much faster than a step-free terrace. Indeed, only at the kink configuration can an atom join the crystal lattice, and thereby reduce its Gibbs potential to that of any of the bulk atoms in the crystal. In the simple cubic system, for example, only the (110) and (111) faces are flat, all others are stepped [17]. The growth proceeds by the attachment of atoms at Kossel–Stranski kinks along the step ledges [34], but not on the terraces, which do not participate in the growth phenomena. Therefore, on the flat, ordered surface of a perfect crystal, growth will not proceed until a small island or cluster nucleates on the surface, thus producing a step loop that is kinked by thermal fluctuations. The stability of such a cluster is given by the Gibbs–Thomson relationship that describes the cluster's solubility. If the cluster reaches the critical size, it may expand, generating a new lattice layer. Thus the nucleation frequency (very low at low supersaturations) determines the ordered face propagation rate. If one considers a nonperfect lattice, where the surface contains defects, such as screw dislocations [22], twins, etc., growth can proceed without the necessity for surface nucleation due to the defect providing growth steps.

Burton and Cabrera also examined the kinetics of vapor phase growth on these surfaces, considering the diffusion of the adsorbed atoms across the close-packed crystal surfaces (terraces), where secondary nucleation is required. This was a refinement of the earlier two-dimenstional nucleation model proposed by Becker and Döring in which surface diffusion [35] was not taken into account. Combining their results with Frank's theory concerning the presence of spiral dislocations that can act as growth steps on otherwise atomically flat surfaces, they published together a seminal paper from which the well-known BCF theory derives [36]. In this theory, developed for vapor growth but later extended to solution growth, the boundary between the crystal and nutrient phase was considered to be sharp (interface of zero thickness), i.e., as proposed by Kossel-Stranski, rather than by Gibbs's finite layer thickness model. In this case, atoms or molecules belonged to only one or the other phase. The BCF theory of layer-by-layer growth of the crystal lattice on smooth surfaces was quantitatively confirmed in numerous studies of growth from solutions, including electrocrystallization.

In the ensuing years, interface structure and surface kinetics models have been refined to include more complex interfaces, including material systems such as biological macromolecules [37]. These and other crystals with large lattice spacings grown from room-temperature solutions have made it possible for in situ atomic force microscopy to capture spiral dislocation sources generating new layers during solution growth, as well as the important phenomena of step bunching, low kink density at steps,

etc. Nevertheless, this theory provided the crystal growth field with a more sound theoretical foundation together with a better understanding of experimental results. It formed an important base from which future studies could be built upon.

The roughening concept was employed in 1958 by Jackson [38] to consider the problem of why many melt-grown crystals of nonmetals had specific faceted, euhedral shapes, whereas metal crystals did not. He derived an elegantly simple theory for the solid–liquid interface structure that could successfully explain and predict experimental results. Jackson used a two-layer Bragg-Williams statistical model (rather than the BCF Onsager model) taking into account nearest neighbor bonds into the solid and lateral bonds within the solid–melt interface. The free energy for adding atoms to a singular (or atomically smooth) interface is calculated until a complete monolayer is formed. Starting with the change in excess free energy associated with randomly adding atoms to such a surface, Jackson found the following relationship for solid–liquid transitions.

$$\alpha = (L/KT_e)(\eta/\nu) \tag{1.2}$$

where L is the change in internal energy associated with the transfer of one atom from the bulk liquid to the bulk solid (latent heat), η is the maximum number of adatom nearest neighbors on the surface, ν is the total number of nearest neighbors of an atom in the crystal, and T_e is the equilibrium temperature for the phase change. This so-called Jackson "α factor" consists of two terms: the first is essentially the entropy of melting divided by the gas constant and is a materials parameter, and the second depends on the crystal structure and specific surface under consideration. The crystallographic term is maximum for close-packed planes, and always <1. It has values of 2/3 for a (100) simple cubic, structure and $^1/_2$ for (111) fcc and (110) bcc structures. Materials with $\alpha < 2$ grow with nonsingular interfaces, whereas materials with $\alpha > 2$ exhibit facets on the growing interface. The former are often metals, with simple centro-symmetric crystal structures, whereas the latter are materials with more complex crystal structures. Using transparent systems having different values of α, Jackson and Hunt [39] were able to demonstrate experimentally the efficacy of their model. Figure 1.3 shows the crystalline morphologies observed for high and low α factor materials. A comparison of the BCF and Jackson models was given by Woodruff [40]. In 2004, Chernov [37] discussed how interface growth kinetics has advanced during the past 50 years.

In the years following, interface structure and surface kinetics models were refined to include more realistic interfaces where each interfacial atom cannot be ascribed to one or the other phase. Instead, this disordered interface is viewed as a layer several atomic spacings thick, where all atoms move randomly and, on average over time, realizes continuous transition between the fully ordered crystal bulk and the disordered melt. This approach allowed for the prediction of a kinetic coefficient linearly connecting the supercooling ΔT at the rough crystal–melt interface to its growth rate V for simple liquids, like metals.

$$V = A\sqrt{kT/m}\Delta T/T_e. \tag{1.3}$$

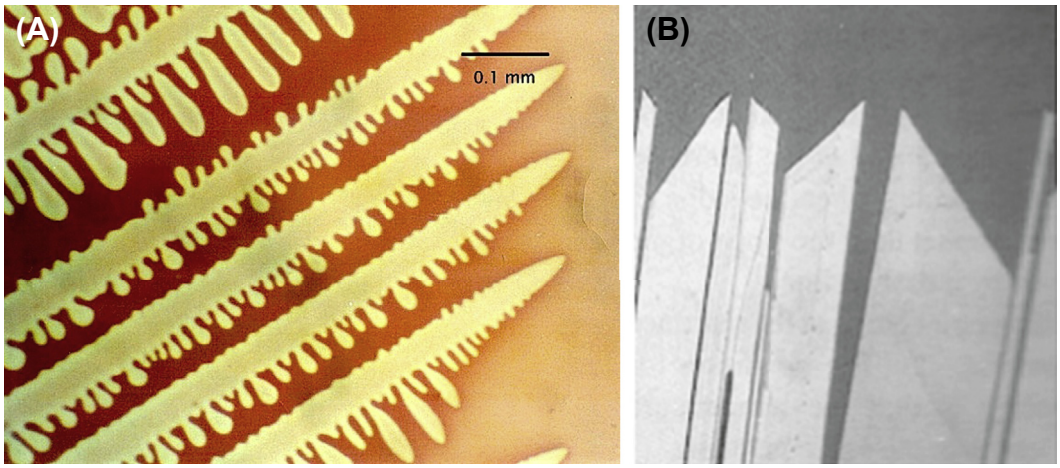


FIGURE 1.3 Comparison of crystal morphologies for (A) a transparent metal analog with an α-factor less than 2. This material grows with a dendritic structure and (B) a benzyl crystal with an α-factor greater than two showing well-developed facets.

Here m is the mass of the atom while the numerical coefficient A is determined by the structure of the liquid and is close to unity. All in all, the BCF theory and its further development provided the crystal growth field with a more stable theoretical foundation together with a better understanding of experimental results. It formed an important base from which future studies could be built upon.

In 1953, research to elucidate and quantify the nature of interface instabilities during crystal growth began with the work of Canadian metallurgists Rutter and Chalmers [41]. They postulated that the cellular (honeycomb-like) substructure that formed in solidifying metals containing a small concentration of impurities (as revealed by rapid melt decanting) was due to some type of instability at the growth interface. This led to the idea that a boundary layer containing rejected impurities develops at a growing solid–liquid interface, depressing the melting point of the liquid in that region so that it became supercooled, but at a higher temperature than the interface. The now well-known term "constitutional supercooling" was derived from Chalmers studies.[1] Shortly thereafter, William Tiller, observed banding in lead crystals arising from unintentional variations in the translation rate (hence growth rate). The structural banding was also found to be associated with the boundary layer composition. Professor Chalmers charged his group to develop a mathematical expression for what was happening at the interface to cause these interesting interfacial instabilities. Their discovery was published later the same year [42]. Their simple

[1]G.P. Ivantsov working independently in Russia in the late 1940s postulated the same concept, calling it "concentrational" supercooling (Dokl. Akad.Nauk. SSSR 81 (1951) 179).

relationship for constitutional supercooling provides one of the most useful tools in the crystal grower's arsenal. The relationship, Eqn (1.4) below, shows how the ratio of temperature gradient in the liquid (G) to the growth velocity (R) must remain above some critical value to achieve stable growth. That value depends upon the material properties of the growth system, i.e., the initial melt concentration C_o (far away from the interface), the slope of the liquidus curve (m), the segregation coefficient (k_o) and the diffusion coefficient. To maintain stable growth and avoid constitutional supercooling one requires that

$$G/R > mC_o(1/k_o - 1)/D \tag{1.4}$$

For the crystal grower this means that to produce a crystal without second phases and cellular structure one *must* either decrease the growth rate for a given temperature gradient or increase the temperature gradient. Faster growth rates are typically very desirable, and so many efforts were undertaken to build special furnaces, sometimes incorporating baffling, localized cooling, etc., to achieve steep thermal gradients.

The early roughening transitions models were two-dimensional models based on the Onsager (BCF) or Bragg-Williams (Jackson) models. It was found that computer modeling was needed to study the problem in more complex three-dimensional systems. Leamy and Gilmer [43] were the first to produce simulated computer images both above and below the surface roughening transition. They also determined the free energy (F) of a growth step for various values of Jackson's α-factor [44]. They showed that F for the step goes to zero at the roughening transition and therefore does not require an energy barrier for new layer formation.

Molecular dynamic simulations have provided detailed information about the process of crystal growth at the atomic level. Its use in morphological stability problems was taken up by numerous groups over the ensuing years which, coupled with experimental work, has led to a significantly greater understanding of the crystallization process (see Figure 1.4).

Modern concepts of interfacial and morphological stability are largely based on the 1963 work of Mullins and Sekerka [45]. Whereas previous researchers knew that various perturbations during growth such as mechanical vibrations, temperature fluctuations, etc., could lead to interface instabilities such as cells and dendrites, they were unable to explain the dynamic mechanisms that were responsible. Mullins and Sekerka developed a mathematical theory of linear morphological interface stability. This was based on small perturbations (sinusoidal ripples) on the growth plane in an unstirred melt that either decay or grow with time. Their analysis led to a more refined relationship that considered the destabilizing effect of the diffusion field and the influence of surface free energy on the boundary conditions. Their results extend the constitutional supercooling criterion described in Eqn (1.3), with several extra terms affecting interface stability [42]. Linear stability theory proves that constitutional supercooling is the correct criterion in the limit of disturbances with small wavenumbers (long wavelengths). The important

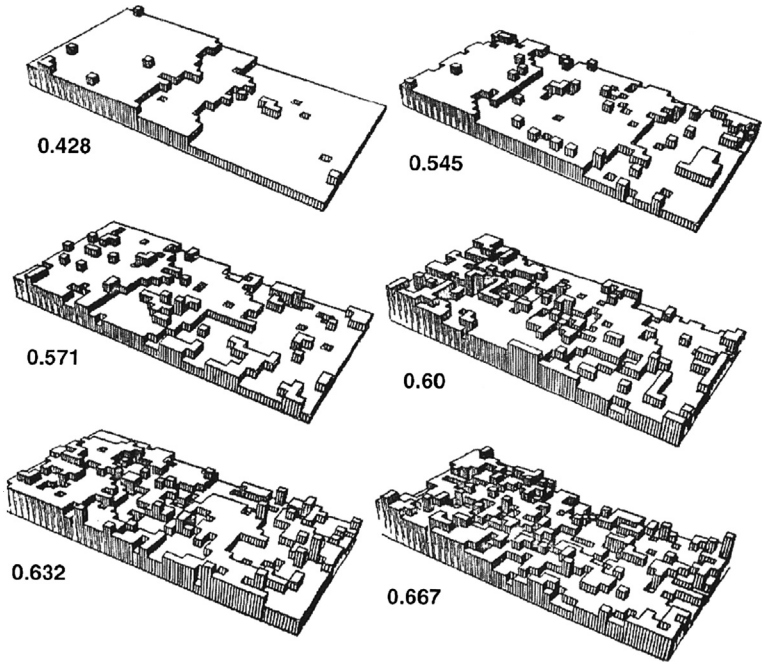


FIGURE 1.4 Monte Carlo simulations of equilibrium surface structures (microstates) for a simple cubic crystal as a function of temperature (as KT/ε) [44]. The surface orientation is the (20,1,0) and at the lowest temperature, $KT/\varepsilon = 0.428$, the step edge is clearly rough. As the temperature increases, the roughness increases. At a critical value, the thermal roughening transition, the steps become indistinguishable.

crystal growth/materials parameters affecting interface stability are the temperature gradient normal to the growing interface, the slope of the liquidus from the phase diagram, the growth velocity, latent heat, mass and interfacial energy densities and the thermal conductivities of the liquid and solid. It was also found that stability depends on whether the thermal conductivity of the melt is greater or less than that of the solid. Other researchers expanded on this research to include molecular attachment kinetic effects, interface energy anisotropy, nonlinear effects, etc., among them V. Voronkov [46] whose independent investigation on mosaic and cellular structures actually predated that of Mullins & Sekerka. John Cahn [47] was the first to treat anisotropic surface tension and interface attachment kinetics (for a spherical geometry). Coriell and Serkeka [48] studied the same types of anisotropies for a planar interface. Chernov [49] treated the case of strong anisotropies and Hurle [50] analyzed the influence of melt convection.

Historically, understanding crystal morphology has provided much of the impetus driving theoretical crystal growth studies. Aside from the regular crystalline forms found in nature (e.g., quartz) or produced during solution or vapor growth in the laboratory, other more complex crystalline morphologies such as dendrites and multiphase eutectic systems have stimulated researchers to uncover the underlying mechanisms involved in their creation.

Dendrites are "tree-like" branched crystal structures that grow in various media under unstable growth conditions. The earliest humans were sure to have noticed and pondered the reason behind the large variety and symmetry of the beautiful snowflake. Snowflakes form when microscopic supercooled cloud droplets freeze and their morphology is dependent on the ambient conditions during their growth. Dendrites also can form during the crystallization of metals, inorganic and organic compounds and even biological macromolecules from melts and solutions. They are common in metals and alloys grown from the melt in shallow temperature gradients.

Dendrites typically contain a stem terminating in a tip and side branches along the stem (see Fig 1.3a). Growth proceeds by steady-state propagation of the tip and a time-dependent crystallization of secondary and tertiary side branches. In 1947, G.P. Ivantsov [357] was the first to identify these self-reproducing crystal shapes—like paraboloids, the basis of the dendrite tip. In pure materials, growth is controlled by diffusion of latent heat away from the advancing growth interface, and in impure systems and alloys it is driven by solute buildup at the interface and where chemical diffusion dominates over thermal transport. In 1960, Temkin [51], and shortly afterward Bolling and Tiller [52], described the role of thermodynamic and kinetic driving forces in the dendritic growth of pure materials. From that time onward, theoretical and experimental dendritic growth studies have proceeded, relying on newer mathematical and computational approaches. Hamilton and Seidensticker [53] examined the role of twin planes in the rapid dendritic propagation of germanium crystals on the basis of re-entrant corner nucleation. In 2004, Glicksman and Lupulescu [54] reviewed 40 years of progress toward understanding the mechanisms involved in the dendritic growth of pure materials including low gravity experiments. An update on this subject is provided in the Handbook chapter "Dendritic Growth."

Growth of polyphase alloys or compounds by unidirectional solidification has also been the subject of much interest to crystal growth researchers. These structures can be produced from eutectic ($L \rightarrow \alpha + \beta$), monotectic ($L_1 \rightarrow \alpha + L_2$) and peritectic ($L + \alpha \rightarrow \beta$) three-phase melt systems. A eutectic crystal can contain four types of structures within a matrix phase (1) parallel lamellar, (2) parallel rods, (3) globular particles of regular shape and (4) irregularly shaped particles. Researchers were interested in the relationship between growth velocity on lamellar spacing and interface undercooling. R. Vogel [55], in 1912, was the first to postulate that growth occurred by both phases growing simultaneously. Eutectic growth theory, however, remained largely qualitative until 1957 when Tiller [56] introduced his diffusion model of eutectic growth. This development was based on the earlier theory on eutectoid growth by Clarence Zener (1905–1993). Tiller's work was used as a basis for Jackson and Hunt's model of 1966 [57], a well-known model and one often used as the basis for later papers. Readers interested in this topic are directed to Glicksman's book [58] that provides, among other crystal growth topics, an excellent review on progress in eutectic solidification.

The transport of heat and mass during crystal growth is of great importance in the design of a growth process and in understanding the resulting features found in the

crystals produced. In melt growth, the dominant factor is often heat transport, whereas in solution and vapor growth, mass transport normally dominates. As to which transport mechanism dominates, it is a matter of degree and an important consideration is what happens in the boundary layer near the interface. As a crystal grows, latent heat is evolved and the allowable growth speed depends on its removal. Therefore, the geometry of the system, the thermal properties of the crystal, the ambient atmosphere and the growth rate all comprise important factors. In addition, the growing crystal needs fresh nutrient to sustain its growth and the rates at which various species reach the interface will partly determine the maximum allowable growth rate and crystal perfection. The concentration of dopant and/or impurity species are often different in the interface region than in the bulk medium, thus influencing mass transport. Instabilities in heat and mass flow can lead to defects such as striations and interface breakdown. The degree and nature of melt convection will strongly affect both the growth process itself and the crystal produced. Many processes, for example Czochralski growth, use forced convection (crystal rotation) to enhance the growth rate and improve thermal and crystal homogeneity, whereas in other methods, for example, vertical and horizontal Bridgman growth, natural buoyant convection occurs from thermally and solutally induced density gradients. In Czochralski growth, the crystal is rotated and sometimes the crucible as well. W. Wilcox [59] and J. Carruthers and K. Nassau [60] studied the fluid dynamic behavior of such systems, as did many other researchers. The effect of fluid flow and flow instabilities are also important in other melt growth processes such as unidirectional solidification, vapor deposition and solution growth. See also Handbook chapter “Segregation and Component Distribution.”

Defects, inhomogeneities, segregation, and interface effects during crystal growth have all been the subject of numerous studies. Some useful reviews have been provided by D. Hurle and P. Rudolph [61] and C. Wang et al. [62].

1.3 Crystal Growth Methods

Crystal growth technology is mainly an applications-driven field. In the last 60 years or so, the major applications have been in the fields of electronic and optical materials. Crystals, however, can be prepared from all types of materials including elements, alloys and inorganic, organic and biological compounds. The compounds can vary from simple binary mixtures to multicomponent systems having numerous components and complex molecular or crystal structures. As a result, crystal growth methods vary widely depending on the thermodynamic and kinetic properties of the system of interest. The starting point for developing a viable crystal growth process begins with a thorough knowledge of the phase relations of the system under investigation. For example, we need to know whether the compound melts congruently, has a phase transformation below its melting temperature, has a high vapor pressure, etc. The most appropriate strategy for producing a crystal depends on the size required, purity and an ability to

control its defect structure (either by elimination, neutralization or incorporation). Crystal dimension requirements (size and shape) are a very important issue in determining the methodology. Single crystals can be grown in bulk, thin film, particulate and fiber form and from the nanometer scale up to meter dimensions. During the last decade, nanoscale wires, whiskers and quantum dots have been found to have unique properties, and this has opened up the possibility for new and improved devices for advanced applications. Classical single crystal growth methods and newer techniques have been used to create a variety of desired nanostructures.

The number of crystal growth methods available to the crystal grower is quite large and varied. The simplest approach to categorizing them is by the nutrient phase from which the crystal is grown. Single crystals can be grown from (1) a liquid phase (melt or solution), (2) from a vapor phase (condensation, sublimation or reaction) or (3) from within a strained solid. Each method has certain advantages and disadvantages that depend on both the properties of the material system involved and the application requirements. Melt growth methods are generally preferred to other methods wherever possible, while solid-state growth methods are the least useful from a commercial point of view.

The growth of a crystal from any nutrient phase requires either a seed crystal or the creation of a solid interface within the growth medium by homogeneous or heterogeneous nucleation. Homogeneous nucleation requires additional energy in the form of supercooling in melt growth or supersaturation in solution and vapor growth methods. Wherever possible, however, the use of a seed crystal or a compatible substrate (as in thin film growth) is desirable. We will explore some of the strategies that have been employed by growers to prepare very high quality, high performance materials.

The theoretical studies mentioned above range from fundamental questions about the mechanisms involved in various crystallization environments to computer simulations of actual growth systems. Issues such as growth rate anisotropy, component segregation, interface faceting, stability and morphology, fluid dynamics, thermal stability and gradient effects, etc., have been extensively studied. During the last decade in particular, computer modeling has helped growers design and modify growth systems in a more systematic way to create thermal and fluid flow environments to enhance interface shape, stability and growth rates.

1.3.1 Melt Growth

When a material melts under nearly congruent conditions and has no low-temperature destructive phase transformations, it is usually desirable to prepare a single crystal of it directly from its melt. Often seed crystals are used to control the orientation and to take advantage of growth rate and thermal anisotropies (heat and expansion). The most useful methods include the Czochralski, Bridgman–Stockbarger, Kyropoulos, Verneuil (flame fusion), and float zone methods. There are innumerable variations to these

general techniques such as the heat exchanger method (an inverted Kyropoulos configuration) and the pedestal growth and micro pull-down techniques to name a few. The discussion of melt growth will begin with the oldest technique for growing large crystals from a melt: Verneuil's flame fusion method.

1.3.1.1 The Nineteenth Century and the Verneuil Process

Alchemists were not only trying to transmute base metals into gold, they were also attempting to grow gemstones in the laboratory. From the beginning of the nineteenth century, various researchers were attempting to grow crystals of diamond, emerald, ruby and sapphire by various techniques, particularly by melting various oxide mixtures. These early methods, however, only produced small crystallites. A really viable commercial process did not appear until the work of Auguste Victor Louis Verneuil (1856–1913) on the growth of large ruby crystals was made public in 1902. Actually, he developed the now well-know flame fusion process a decade earlier and spent the next decade improving the method before making it public knowledge. A very thorough description of the life and work of Verneuil was given by K. and J. Nassau [63]. Verneuil (see Figure 1.5(A) below) was a French "renaissance" man and well-beloved teacher, actively interested in music performance and art and whose accomplishments spanned many different areas of chemistry. He became interested in chemistry working in his father's photography shop (his father changed careers after meeting Mr Deguerre (Louis Jacques Maude, 1787–1851), the inventor of photography. In 1873, at age 17, Verneuil went to study in the chemical laboratory of the distinguished Professor Edmund Frémy

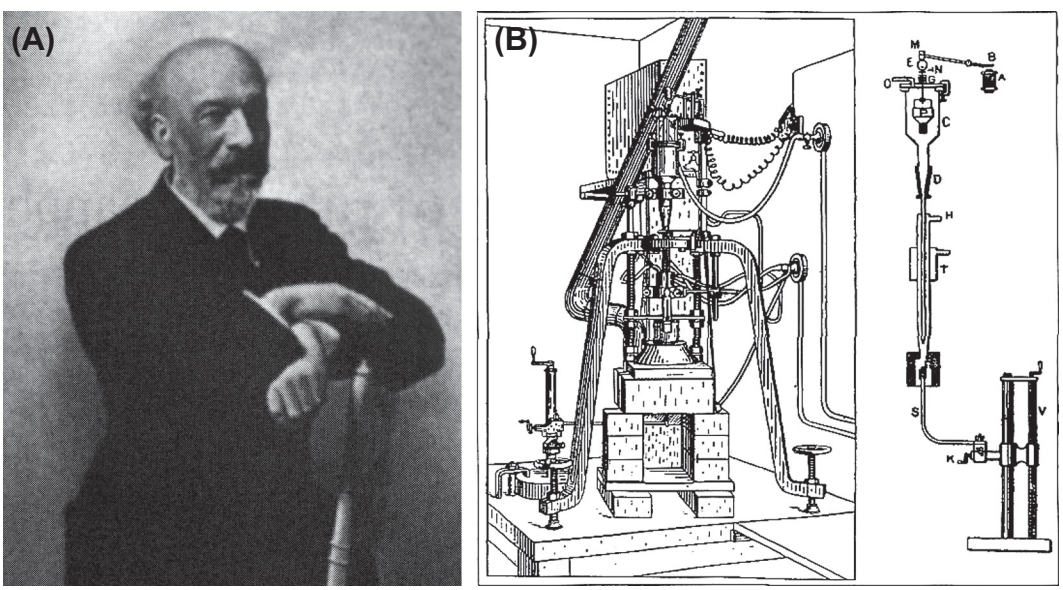


FIGURE 1.5 (A) A photograph of Dr. A.V.L Verneuil, and (B) a schematic diagram of Verneuil's crystal growth apparatus [63].

(1817–1894). He eventually earned his doctoral degree in 1886. In his earlier years in the Frémy laboratory, he participated in research on ruby crystal growth by a melting technique using porous alumina crucibles. He also became friends with Henri Moisson (1852–1902) who worked on diamond crystallization in Frémy's laboratory.

In 1886, an unknown group from Geneva started selling larger synthetic ruby crystals than were available elsewhere. It is now believed that these so-called "Geneva Rubies" were actually grown by an early version of the flame fusion process [63]. Verneuil was intrigued by these samples, and it stimulated him to develop the method for which he became famous. The mysterious group from Geneva ceased operation in 1905, not long after the Verneuil Process had gone into commercial production. Such groups have appeared from time to time trying to pass off synthetic crystals for natural stones that garner much greater value.

The flame fusion method was first developed to produce large, high-quality ruby for the gemstone market and also for watch bearings. The process, which is still in use today, involves passing a powder of the compound through a vertically aligned oxy-hydrogen flame. Molten droplets descend by gravity onto a rotating alumina pedestal containing the growing crystal and the crystal grows upward on the pedestal. The basic apparatus used by Verneuil is shown in Figure 1.5(B). Temperature gradients are steep, boules are prone to cracking and the early powder delivery systems were often unreliable. Important processing refinements were made by Verneuil over his lifetime to improve the process reliability and crystal quality. The first problem he solved was the severe cracking problem. He accomplished this by reducing the contact area of the boule with the pedestal. While preventing cracking, the boules after growth were still highly strained. This strain was relieved naturally when the boules split in half or were split by hand. Powder delivery was done by mechanical tapping mechanism mounted on a hopper containing the charge powder. For ruby growth, the powder Verneuil used was a mixture of ammonium and chrome alums. The chromium oxide concentration in the boules was $\sim$2.5%. The oxygen content in the ambient gas phase was critical for achieving the appropriate oxidation state in the crystal. A flame rich in hydrogen and carbon was necessary to prevent introduction of gas bubbles in the molten ruby melts.

In 1909, Verneuil worked with L. Heller & Son of New York and Paris on developing his process for making blue sapphire. Instead of chromium additions, the sapphire was doped with a mixture of iron and titanium oxides, two impurities found in natural minerals. He suggested that the titanium in the crystal gave the deep blue color by converting the ferrous ions created by the flame back to ferric ions. Another pioneer of the flame fusion growth method was Leon Merker (1917–2007). He also worked with the Heller Co. starting during the early days of World War II, after he escaped fascism in Europe and came to the U.S. to study at the University of Michigan. Based on a friend's recommendation, he met Mr Heller from France. After some fruitful discussions, Heller assigned Merker the task of setting up the Verneuil Process for ruby and sapphire in New Jersey. The venture was successful and the General Synthetics Corporation was formed

in 1941 to provide ruby to the military and for the gem industry [64]. Merker also worked on barium, calcium and strontium titanates; the latter two with greater success.

Since Verneuil dedicated much of his career to the successful development of a commercial process for growing sizable crystals with controllable properties, he might be considered the father of the commercial crystal growth industry.

1.3.1.2 The Twentieth Century

Even while Verneuil was improving on his method, other researchers at the turn of the century were beginning to produce crystals in the laboratory to study both their solidification behavior and physical properties.

One of the earliest was Gustav Tammann (1861–1938). He was born in Russia of Baltic and German parents and spent most of his life in Germany. Among other notable achievements, he established the first Institute of Inorganic Chemistry in Germany at Göttingen University in 1903. Tammann's interests led him to study the solidification of metal alloys and their nucleation behavior. He made important contributions to the fields of heterogeneous equilibria, crystallization and metallurgy. One of his important contributions to crystal growth involved the solidification of metal alloys in long narrow tubes tapered to a point to both confine nucleation and supercooling to a small volume and thereby promoting the propagation of a single crystal along the tube [65]. He was probably one of the first to understand the relationship between grain selection and growth rate anisotropy and the concept of confining the melt to control the number of grains that form. His method would be classified today as the gradient freeze method. He also grew crystals of a number of organic compounds and studied their crystallization behavior.

Within the same time period, Obreimov and Schubnikov from Saint Petersburg, Russia, [66] published a paper describing the growth of metal crystals using a modification of Tammann's method, i.e., in a long glass tube with an imposed temperature gradient along its length. They also briefly discuss the easy to operate Czochralski process (to be discussed below) but rejected it in favor of the Tammann's method because the free-standing Czochralski crystals were not of uniform shape and some of the low-melting metals could deform during growth without being supported. In their experiments, they used a vertical cylindrical tube tapered at the bottom like Tammann. Nucleation was achieved by cooling the tapered tip with cold air and then, after crystallization in this region was accomplished, slowly cooled the furnace to propagate the crystal up the length of the tube. Both these methods distinguish themselves from the Bridgman and Stockbarger methods (also to be discussed later) in that growth is not achieved by moving either the ampoule or furnace to solidify the melt.

1.3.1.3 The Czochralski Crystal Pulling Method

1.3.1.3.1 The Invention

Following Verneuil's pioneering work, a number of other researchers began to grow metal and alkali halide crystals for property studies. In 1918, Jan Czochralski, a well-known young Polish metallurgist (head of AEG's metals laboratory in Berlin), published a

paper [67] that would describe a technique that quickly became one of the most powerful methods for growing crystals in use today. In a story related by Tomaszewski [68], Czochralski, while working late at night in his laboratory, discovered by accident the crystal pulling method for which he because famous. His studies concerned the crystallization rate of metals and while working on his experimental notes, instead of dipping his pen tip into the ink well, dipped it instead into a crucible of molten tin on his desk. When he pulled it out he found a long filament of solidified tin on the end. He subsequently found it to be a single crystal. He then realized the value of studying crystallization rates using such a device. His early apparatus (see Figure 1.6(A)) consisted of a clock-motor-driven lifting mechanism. Replacing the pen tip, a short tapered glass rod with a hook on the end was held on a silk thread connected to the pull mechanism. The rod could be raised or lowered in a continuous fashion. By dipping this rod into the surface of the melt, he was able to solidify metal onto it and pull out crystals of tin, lead and zinc in a continuous and controlled fashion. Czochralski later modified the glass rod, incorporating a capillary at the bottom to draw up the molten metal. This had the effect of restricting nucleation to the limited volume of melt in the capillary. With this apparatus he produced 1 mm thick single crystal wires at maximum crystallization velocities of up to 140 mm/min and in lengths up to 19 cm. Czochralski's life and research accomplishments can be found in Tomaszewski's monograph [68].

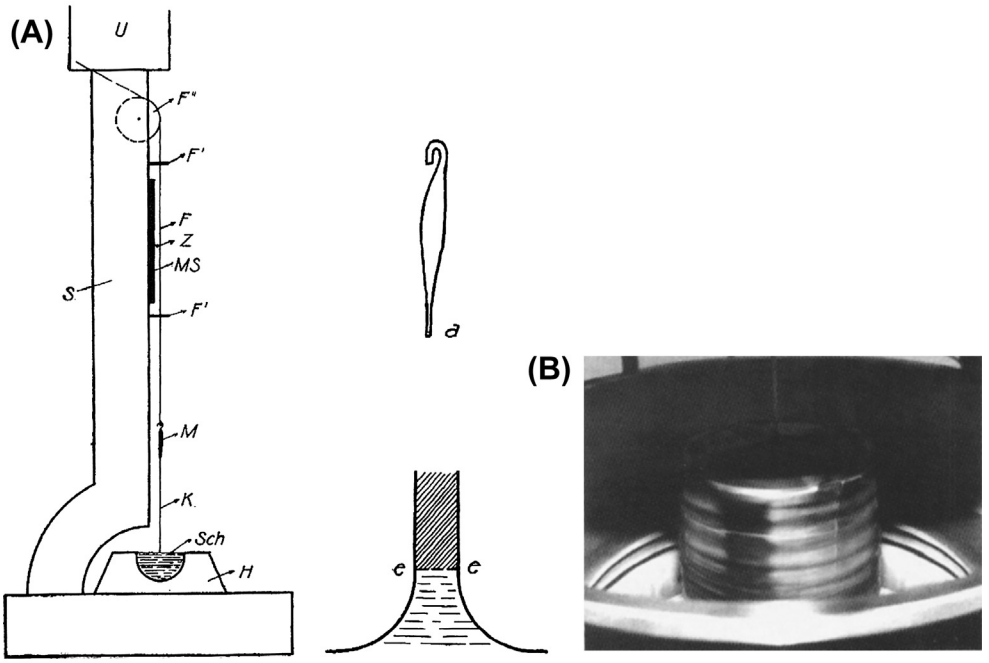


FIGURE 1.6 (A) Czochralski's original experimental setup [67], and (B) a photograph of a five-inch diameter commercial single crystal silicon boule growing by Czochrakski's method. *(From the front cover of the AACG Newsletter 13 (1983)—photo courtesy of the Siltec Corp).*

Czochralski's pulling method was almost immediately put to use by von Wartenberg [69] to grow zinc single crystal wires onto oriented seed crystals. Somewhat later, Von Gomperz [70] pulled single crystal fibers of metals through a hole in a mica plate floating on the surface of the melt. He also used a capillary for seeding. His work was the forerunner of the edge-defined film-fed growth (EFG) and the laser heated pedestal fiber growth methods. In 1928 E.P.T. Tyndall [71] wrote a paper on the *Factors Governing the Growth of Zinc Crystals by the Czochralski-Gomperz Method*. In 1937, Henry Walther of Bell Laboratories published the first paper on the use of Czochralski's method for the growth of nonmetals [72]. He rejected other methods such as the Kyropoulos method because he was intent on growing long, uniform, cylindrical bars of NaCl single crystals for property measurements. He therefore was attracted to the pulling methods that Czochralski and von Gomperz used to grow low-melting metals. He used a quartz crucible to hold the melt and dipped a platinum rod or closed tube into its surface, sometimes with an oriented seed attached. In the beginning, he used Kyropoulos's method of pulling the tube up slightly after the first melt solidified on the rod to reduce melt contact with the crystalline solid that formed. He placed an air-cooled coil containing small holes above the melt to cool the growing crystal and pulled up at rate of 5 cm/hr while rotating the crystal at 10 rpm. He was the first to apply rotation to the Czochralski method and produced the first bulk crystals of a high melting point compound by this method. Walther successfully produced NaCl boules 2 cm in diameter and 30 cm long. It is rather amazing that this paper, although published in a prominent journal and referenced twice a few years after its publication, was only found very recently (by Reinhard Uecker [73]). Strangely, even many Bell Laboratories researchers from that period to the present time seem to have been unaware of Walther's work, and it was not mentioned in the rather extensive review of engineering and science research in the Bell system during the period 1925–1980 [30]. In 1940 Evans [74] used Walther's method to grow single crystals of NaCl, KCl and KBr.

1.3.1.3.2 Semiconductors

Bardeen and Brattain discovered the transistor in 1947 using large-grained Ge samples produced by unidirectional solidification [30]. Shortly afterward, it was demonstrated that single crystals were better, and this led to a dramatic expansion of the crystal growth field in general, and the Czochralski method in particular. This versatile technique has been applied to a wide variety of materials of commercial importance—particularly semiconductors and optical materials.

According to [30] (p. 422), "A single crystal growth technique, first used by J. Czochralski in 1917, was adapted and improved in 1950 by G.K. Teal and J. B. Little for the growth of single crystals of germanium" [75]. They dipped an oriented Ge seed crystal into the melt surface and, while rotating, pulled modest sized crystals (by today's standards) of 2.5 cm diameter and 10 cm in length. The minority carrier lifetimes were significantly better than in polycrystalline materials, and therefore the semiconductor researchers shifted their efforts to producing bigger and better crystals with control of

the dopant concentrations and uniformity. In addition, the single crystal technique allowed for the creation, during growth, of n-p-n junctions by perturbing the growth conditions [76]. While this junction technique was eventually superseded, it was an important step in transistor technology. The importance of homogeneity in semiconductor devices led many researchers to study the thermodynamic and kinetic aspects of impurity and dopant incorporation. One such seminal study, as mentioned earlier, was that of Burton, Prim and Slichter [28,29]. Shortly after the germanium research activities began, the focus shifted to silicon whose properties were deemed to be superior. In 1952, Teal and Buehler [77] reported on the Czochralski growth of silicon crystals—a much higher melting compound (1414 °C compared with 938 °C for Ge) and more difficult to grow due to its reactivity. Over the years, crystal sizes have constantly increased (see Figure 1.6(B) above), and today commercial systems are available to grow Si boules 12 in in diameter and 6 ft long from which substrates can be cut for the preparation of integrated circuits. In situ recharging to grow longer crystals and controlling melt flows using magnetic fields were added over the years to boost production rates and quality. One of the most important factors in producing high quality crystals was not only to control impurities and other point defects but their complex interactions with each other, as well as with dislocations. Removal of one defect can lead to the redistribution of other defects to lower the overall energy of the system.

It was recognized early on that purity of the starting material was critical to semiconductor performance. In 1951, William Pfann [78] invented the zone refining method for ultrapurifying Ge. This very important method, in wide use today, has been successfully adapted to the purification of all classes of materials. Shortly afterward, Theuerer [30] invented the crucible-free float zone process to grow O_2-free silicon. Oxygen incorporation during Czochralski growth was due to the use of SiO_2 crucibles. Theuerer's method is still in commercial use today to produce O_2-free Si for special device applications. It also has been used with other materials for which melt-crucible interactions are problematic. With the use of optical heating systems (such as lasers or xenon lamps), its simplicity makes it very useful for growing crystals of numerous materials for physical property studies. Pfann [79] also invented the zone leveling crystal growth method that is a combination of horizontal Bridgman growth coupled with zone refining. In this case a seed and a dopant are placed at one end of a horizontal tube and by moving a molten zone along the tube, the dopant could be uniformly distributed along the boule. This method was used early on to produce transistors and diodes.

It became apparent that dislocations were affecting the electrical properties of Si single crystals and the need for zero-dislocation material arose. In 1959, William Dash [80] developed a method for doing this during Czochralski growth. Since dislocations propagate mainly from the seed, he used high-quality seeds together with careful control of the initial growth conditions. He was able to produce dislocation-free crystals by "necking" down the growing boule to a very small diameter before widening it back out to the desired size.

Important advances in both purification and control of crystalline perfection has led to the improved performance of Si devices.

When compound semiconductor materials such as GaAs, InP and their alloys became important to the electro-optic field, special Czochralski techniques had to be developed because they have high vapor pressures at elevated temperatures. In 1962, Metz et al. [81] were the first to report the use of molten B_2O_3 as a melt encapsulant for the Czochralski growth of PbTe. Both Pb and Te are volatile at the compounds melting temperature, and they successfully sought to cap the melt to prevent losses. In 1965, Mullins et al. [82] demonstrated that molten B_2O_3 was also a useful encapsulant for the growth of GaAs and InAs. They were able to adapt a commercial low-pressure Czochralski system for the growth of these compounds. This method is now known as the liquid encapsulation Czochralski (LEC) method. A few years later, Mullins et al. [83] extended their work to include the growth of GaP and InP in high-pressure furnace systems. One problem with group V elements was that when the crystal emerges from the encapsulating layer, it starts to lose P or As from its surface. In 1983, Azuma [84] came up with an innovative approach to inhibit these losses. He used a pressure balancing system to control the partial pressure of P in the InP growth chamber. In the upper chamber was extra P_4, maintained at a pressure such as to prevent evaporative losses at the crystal surface.

1.3.1.3.3 Oxide Growth

Following the success of the Czochralski method for growing elemental Si and Ge single crystal boules at Bell Laboratories, they and other laboratories started to use this method extensively for growing bulk single crystals of oxide compounds for laser, nonlinear optical, scintillator and numerous other applications.

The laser was predicted by Arthur L. Schawlow and Charles H. Townes in 1958 [30, 358, 359] but not actually demonstrated until the work of Maiman in 1960 [85] with a single crystal ruby rod prepared at the Union Carbide company. In the same year, Nassau and Van Uitert [86] were the first to use the Czochralski's method to grow a high-quality oxide crystal. They prepared laser crystals of Nd:$CaWO_4$. During the following decade, the Czochralski method was vigorously pursued in many research and industrial laboratories around the World. A wide variety of important optical materials were grown, including $LiNbO_3$ [87,88], $LiTaO_3$ [89], $Bi_{12}Ge(or\ Si)O_{20}$ and $Sr_xBa_{1-x}Nb_2O_6$ [90], YAG ($Y_3Al_5O_{12}$) [91], Nd:YAG [92], Sapphire [93,94], and $Gd_3Ga_5O_{12}$ (GGG) [95]. Many of these materials are still commercially important. A concise history of oxide crystal growth by the Czochralski method was given by C. D. Brandle [96].

Many improvements to the method were made over the succeeding decades. Compositional variations along the length and diameter were of major importance and stimulated the construction and analysis of related phase diagrams. It was found, for example, that the stoichiometric composition was not always the congruent composition [97] and to get uniformity one needed to shift the composition to the off-stoichiometric congruent composition to achieve homogeneity. Another problem often

encountered was that the shape of the phase field boundary of the compound might be curved instead of straight leading to precipitation of a second phase.

During the 1960s, very little was known about how the growth interface shape could influence crystal quality. This was very important in Si growth where zero-dislocation crystals rely on a particular interface shape. Cockayne et al. [98] were the first to show that interface shape could be modified and controlled by crystal rotation. Nominally, crystals growers find that a slightly convex interface toward the melt is most desirable. Another factor of major importance in melt and solution growth is fluid convection. It affects mass and heat transport and therefore interface shape, boundary layer and growth rate instabilities, etc. In Czochralski growth natural convection and crystal rotation can interact to modify both the interface shape and the composition in the melt near the growth interface (boundary layer). Various researchers have achieved significant improvements in crystal quality by controlling these parameters.

The application of computer modeling to help solve crystal growth problems was begun in the 1980s by Robert Brown and his group at MIT. One example is a paper written by Derby and Brown [99] on the dynamics of Czochralski growth. One of the major tasks of computer simulations is to model the flow regimes in a system in which the thermal configurations are adjustable. In recent years, facilitated by the dramatic increase in computing power, almost all types of crystal growth processes (Bridgman, float zone, etc.) have been modeled. Simulations performed have been very successful in helping design and guide refinements to laboratory and commercial crystal growth process.

During Czochralski growth, the melt level in the crucible drops as the crystal grows. This changes a number of factors including the thermal gradients and convection patterns. Often the temperature has to be changed during growth or some other parameters modified. Whiffin and Brice [100] have shown that melt height can affect thermal oscillations in the melt. These thermal fluctuations can lead to growth rate variations and crystalline imperfections such as striations. A striation is a compositional variation parallel to the growth interface, usually caused by poor temperature control and/or melt oscillations. In the 1960s most growth was carried out manually, i.e., the temperature was changed or the crucible position altered by analog temperature and motor controllers. With commercialization came the need for automated diameter control systems. These were based on either crucible or crystal weighing or by controlling the meniscus position optically.

1.3.1.4 Bridgman–Stockbarger/Gradient Freeze Methods

Little did Percy Bridgman (1882–1961) or Donald Stockbarger (1895–1952) know at the time of their respective discoveries that their names would become historically intertwined in describing one of the most popular techniques for growing crystals. Their versatile method(s) made possible the growth of many different types of materials including metals and their alloys, semiconductors, and both inorganic and organic compounds. It was also a method that allowed the preparation of some of the largest manmade crystals ever produced.

Percy Williams Bridgman was a Noble Prize-winning American physicist working at Harvard University, only a few miles away from MIT where Donald C. Stockbarger worked as an Associate Professor of Physics. His prize (1946) was for his work in the field of high pressure physics. The crystal growth method he developed and published in 1925 [101] departed from the work of Tammann [65] and Obreimov and L. Schubnikov [66] in that the vertical tube containing the melt was not stationary during growth. Growth was initiated in a capillary tube at the bottom end of a larger cylindrical ampoule and propagated upward along the tube by lowering it down through a single zone vertical tube furnace and out the bottom. The capillary was used for seed selection and was further enhanced by reducing the capillary diameter at the juncture between the capillary and the larger bore container. His first experiments were done using bismuth melts. Not long afterward, various other researchers used his method or variants of his method to grow other metal crystals, such as copper, and zinc.

In the late 1920s Stockbarger started his work on the growth of large, high-optical-quality crystals of LiF and later CaF. At first, he tried Bridgman's method [101] for CaF but it required more careful atmosphere control to prevent hydrolysis, better starting material purity and temperature stability to produce useful crystals. This led Stockbarger to modify Bridgman's method [102,103]. He used a so-called vertical "elevator furnace" that had two graphite heaters separated by a Mo baffle through which a covered crucible containing the melt could be passed from the upper higher temperature region into a lower temperature section by a motorized translation device. The use of a two-zone furnace led to better control of the thermal gradient at the growth interface. The crucible, support rod and pedestal were graphite. The V-shaped crucible bottom rested in the pedestal. There was no capillary region below the tapered region for seed selectivity, and the included angle was much larger than those used in the Bridgman and other earlier methods. So it is speculated that seed selection was controlled by the locally steep gradient at the tip created by the thermally conducting graphite support rod coupled with the baffle, thereby limiting the volume of supercooled melt that can form.

The Bridgman–Stockbarger method (shown in Figure 1.7 below) has been widely used to grow crystals of varying sizes from its development in the 1920s until the present day. It has also been used extensively in a horizontal configuration. While initially used for metals and then shortly afterward for inorganic optical materials, it has since been used to grow hundreds of other compounds including semiconductors (GaAs, CdTe, HgCdTe, and chalcopyrite compounds such as $CdGeAs_2$, $ZnGeP_2$), organic materials, oxides such as $Pb(Mg_{1/3}Nb_{2/3})O_3$–$PbTiO_3$ (PMNT) and halides such as Tl:CsI and Tl:NaI, and Eu:SrI_2 etc. Process improvements include the use of the accelerated crucible rotation technique to improve melt homogeneity and interface boundary conditions [104], vibroconvective mixing [105], baffles in the melt near the interface, growth under high pressure, etc.

The gradient freeze (GF) method of Tammann [65] differs from the Bridgman–Stockbarger approach in that there are no moving parts. Neither the ampoule nor furnace

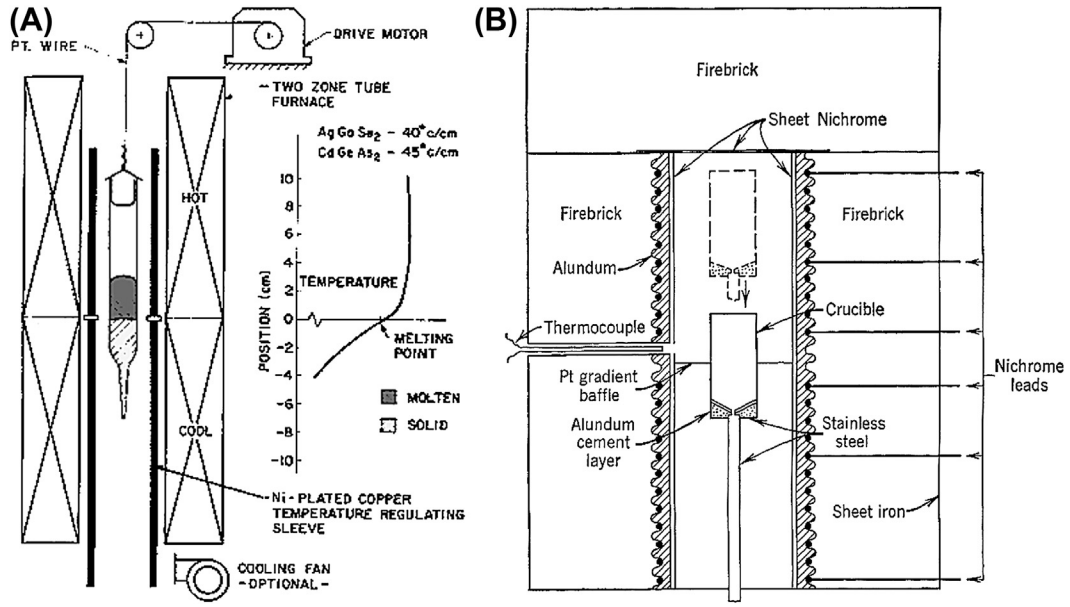


FIGURE 1.7 (A) Drawing of a Bridgman apparatus showing a tapered crucible being lowered through a stationary furnace having a steep gradient at the growth interface. In Bridgman's experiments the crucible is lowered out of the furnace. (B) A schematic diagram of Stockbarger's growth apparatus. Note the platinum baffle that separates the two furnace zones for gradient control and the crucible pedestal.

is translated with respect to the fixed thermal gradient in the furnace. Instead, a temperature gradient is maintained across the melt-containing crucible such that when the temperature of the furnace is lowered, the cool end of the melt solidifies first, and the rest of the melt solidifies layer by layer as the freezing point moves across the melt. This method is simple to implement and was used for the growth of a number of materials. Its big disadvantage was that as the furnace temperature decreased, so did the gradient across the remaining melt. Under constant cooling conditions this change in gradient could lead to changes in growth velocity and thereby variable crystal properties along its length due to component segregation and perhaps interface breakdown. One method used to solve this problem was by continuously changing the furnace-cooling rate to maintain constant freezing rate in the melt. In 1986, Gault et al. [106] successfully applied the vertical gradient freeze method (VGF) to the growth of large diameter GaP, InP, and GaAs crystals.

Attempts to grow some important III–V compounds by the vertical Bridgman and gradient freeze methods were complicated by the fact that these compounds expand on cooling and can aggressively stick to the walls of many crucible materials [107]. These methods both exist in horizontal versions that are applicable to certain important commercial crystals. While many different types of crystals have been grown by the horizontal Bridgman and gradient freeze techniques, their sizes are limited compared their vertical counterparts, and the boules have noncircular cross-sections.

1.3.1.5 Nacken–Kyropoulos Methods

During the early decades of the twentieth century, many new developments in crystal growth technology came out of Germany. During the 1920s in particular, a burst of activity in the field led to numerous growth techniques being developed, many of which are being used today in either their original or modified form.

One of the most important crystal growth pioneers of this period was the German mineralogist Richard Nacken (1884–1971). A few years before Czochralski's discovery, he reported on a process for growing crystals from the surface of a melt using a cooled copper rod with a rounded end and a seed attached [108]. Nacken's apparatus is illustrated in Figure 1.8(B) below. The general idea was to locally supercool the melt adjacent to the rod and initiate growth under controlled conditions. After growth started, the furnace temperature could be lowered to keep the seed growing. No pulling was involved. As the crystal grows, the melt level drops due to the higher density of the crystal. The method was later used by J. M. Adams and W. Lewis [109] to grow very large ice crystals. Nacken also developed a viable hydrothermal process for growing quartz crystals. His unpublished work was found in secret WW II German reports. E. Buehler and A.C. Walker at Bell Laboratories [110] based their successful hydrothermal quartz growth technology on Nacken's process.

About 10 years later, Spyro Kyropoulos (1887–1967), a student of Tammann and professor of Applied Physics at the Gottingen University (later he taught at the California

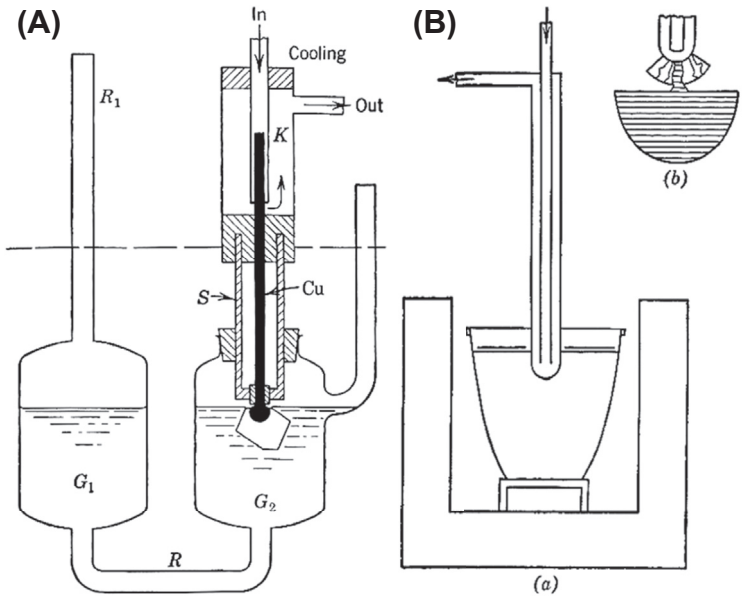


FIGURE 1.8 (A) A drawing of Nacken's apparatus illustrating the growth of a faceted crystal using a seeded cold "finger" inserted into the melt's surface. (B) A schematic diagram of Kyropoulous's experimental setup where, unlike Nacken, a cold rod is place in the melt surface without a seed. The inset shows how seed selection can be accomplished using a rounded seed rod.

Institute of Technology) took up Nacken's melt growth method to grow crack-free alkali halide crystals for precision optics. The advantage of Nacken's method is that the crystal is grown within the melt rather than being confined to a container that can induce strain in the crystal during cooling. Instead of a using a seed, Kyropoulos [111], using an air-cooled platinum tube, nucleated a few crystallites on the end of his tube and then lifted it up slightly so that the melt stayed in contact with only one grain. Kyropoulous's apparatus is shown in Figure 1.8(B) above. This seeding method had to be carefully controlled so that the tube did not break free from the melt surface. After the seeding stage, the furnace is slowly cooled to allow the crystallite with the fastest growing direction of heat flow to grow to cm-size crystals before being pulled out of the melt. The thermal gradients in the melt are generally quite small. As mentioned before, the melt level drops in systems where the density of the solid is greater than the melt density. Kyropoulos used this process for growing many alkali halide crystals [112]. The method is attractive because of its general simplicity, reliability and low operating costs. Two other advantages of the method are (1) the ability to see what was going on and to make adjustments to enhance the crystal quality and (2) its use of lower thermal gradients than in the Czochralski method. On the other hand, the lower gradients lead to faceting at the interface and thus chemical inhomogeneities in the crystals. Several years later, Korth [113] took up this method but used a seed attached to the cooled rod as did Nacken many years earlier. He grew crystals as large as 6×8 cm. A few years later, Katherine Chamberlain in the United States used this method to grow very large KBr crystals up to five inches in diameter and weighing up to seven pounds [114]. Typically Kyropoulos's method does not involve continuous pulling or rotation as in the Czochralski method. Growth rates are of the order of mm's/hr with cooling rates below 1 °C/hr. The crystal diameters usually are up to 90% of the crucible diameter. Bliss [115] gave a detailed review of Kyropoulos's life and method.

The Kyropoulos method has been in commercial use for over 75 years. From its inception until the present time, the method has been used to grow large alkali halide crystals for windows, prisms and scintillators (e.g., Tl:NaI and Tl:CsI) compounds. Due to the development of GaN-based light emitting diodes, there has sprung up a very large industry around the growth of large sapphire crystals for use as substrates. The Kyropoulos method is one of the most widely used methods today for this applications. Up to 12 inch-long crystals have been produced. It is also used in the commercial production of Ti:Sapphire laser crystals. A variety of other materials have been grown by the Kyropoulos method in laboratory settings, including organic materials, semiconductors such as Si, ZnSe, and InP and other types of laser crystals. For the growth of InP crystals [116], liquid encapsulation together with magnetic fields has been applied to the Kyropoulos configuration to improve crystalline perfection. This has been called the MLEK method. Over the years, furnaces have become much more complex. Heat shields are now used to control thermal profiles. Numerical analysis has helped to define the optimal conditions for growth of specific materials through proper baffling and positioning of the crucible in the heater. Other improved capabilities include the ability to weigh the crystal or crucible during growth to control the rate of mass increase with time.

1.3.1.6 Stöber/Heat Exchanger Methods

In 1925, F. Stöber [117] reported on the growth of large crystals of anisotropic materials such as sodium nitrate, zinc and bismuth by removing the latent heat of crystallization from the bottom of a stationary bowl-shaped crucible containing a melt (growth from the bottom upward). A heater plate was placed above the surface of the melt and a water-cooled plate at the bottom creating an axial temperature gradient. The radial heat flow, present in most other growth systems, was minimal. Stöber's method, along with one of Tammann's techniques, were perhaps the first gradient freeze methods (i.e., moving the gradient along the melt rather than by moving the crucible or furnace). Stöber also found that crystal singularity was enhanced when the thermal conductivity in one crystallographic orientation was significantly greater than in other directions. One attractive feature of Stöber's method was that you could produce very large crystals in near-net shape, i.e., in the exact shape of the container. In addition the method is very simple to implement since there are no moving parts, and lower thermal gradients employed help reduce stresses in the final boule. He grew crystals of $NaNO_3$ up to 10 pounds by his method. He also grew ice crystals by inverting the cooler and immersing it into the liquid surface.

In 1970, Frederick Schmid and Dennis Viechnicki [118] from the Army Research Laboratory at the Watertown Arsenal, reported on a new method to produce large-diameter sapphire crystals from the melt. This work was stimulated by the need by the military for very large transparent armor plates. They called their new technique the heat exchanger method (HEM). The method is similar to the Stöber process in that the crystal grows upward from the bottom filling the crucible and taking its shape. Neither the furnace, crucible nor crystal moves during growth. A He-cooled cold finger (similar to the water-cooled bottom plate used by Stöber) extracts heat from the crucible bottom in a controlled manner and independent of the heat input. A furnace that surrounds the crucible replaced the upper pancake heater. In addition, the technique allows for a small-diameter seed to be centered over the cold spot created by the He heat extraction tube. Heat can be removed from the crucible bottom by increasing the He flow rate. Since then (1975), a He recirculation system was developed along with the technology to grow large-diameter, flat-bottom crystals free of light scatter from the small seed centered over the heat exchanger. High-quality sapphire crystals have been grown commercially up to 44 cm in diameter and weighing 160 kg [119]. In situ annealing in shallow gradients is used to relieve stresses. The crystals produced are competitive with commercial Kyropoulos sapphire. The method has also being used to produce single crystal ingots of spinel, ruby, Ti:sapphire, Nd:$Y_3Al_5O_{12}$ and silicon. Commercial growth systems are available.

1.3.1.7 Kapitza's Method

Pyotr Kaptiza (1894–1984) was an important Russian physicist who spent many years in England before returning to Stalinist Russia. He was awarded a Nobel Prize for his work in low-temperature physics. He used the Bridgman method to prepare metal single

crystals for his experiments, but when it came to preparing Bi rods of a specific orientation, he found that its expansion on cooling was sufficient to cause enough strain to prevent the seeded orientation from propagating down the rod. To solve this problem, he developed a method in which the crystal was not completely constrained [120]. He placed a bismuth rod on a copper plate and covered it with loose-fitting glass plates to reduce drafts and then melted the rod. At one end, an oriented seed was attached and by unidirectionally solidifying the rod by cooling he was able to produce an oriented Bi single crystal at rates up to 5 cm per hour. This method was taken up and modified by others, and today it might be considered the forerunner of the horizontal Bridgman method which is widely used today for growing materials like GaAs for LED's [121] and $ZnGeP_2$ and $CdGeAs_2$ [122] for nonlinear optical applications.

1.3.1.8 Zone Melting

Zone melting methods have played an important part in (1) purification of materials for crystal growth and other materials processing (zone refining), (2) for producing crystals with uniform composition (zone leveling) and (3) for growing crystals without crucibles (the float zone method). There are a number of variants to all these methods including different methods of heating, horizontal and vertical arrangements, traveling mechanisms, etc. The first reported use of zone melting was by Kapitza in 1928 [120]. In his experiments, he passed a short resistance heater along a Bi filled tube to produce a single crystal. In 1937, Andrade and Roscoe [123] used zone melting (also a traveling heater) to grow lead and cadmium single crystals having low strain. By far the most important and extensive work on zone melting for purification (zone refining) and zone leveling was that of William Pfann [78]. In 1952, Pfann conceived of the zone refining method when asked by the transistor researchers at Bell Labatories (where he had worked since the late 1930s) to develop a method to produce higher purity Si and Ge for use in growing crystals with better electronic properties. Toward this end, it was used with great success in the early days of semiconductor processing. This very versatile method was found applicable to numerous other materials, including inorganic and organic compounds, metals and semiconductors. Many papers on its adaptation to different materials and operational improvements have been published since then, and it has become an important tool for both research laboratories and industry. Pfann's book on zone melting [124] is the seminal publication on the method, comprehensively covering both the theory and practice involved.

The method consists of moving a molten zone through a bar of material. Two solid interfaces are created. As the zone moves (by moving the heater or furnace), material from one interface dissolves in the zone and is recrystallized at the other interface. Purification occurs for those impurities whose solubility in the liquid is different than that in the solid (segregation coefficients, K, greater or less than one). Materials with $K < 1$ will be rejected at the growth interface and will build up in the last to freeze region. Those with $K > 1$ will tend to congregate at the start of the ingot. It is also a very useful method for growing crystals from materials that melt incongruently.

Various types of heaters have been used with the zone melting method, including resistance, RF, electron beam, plasmas, lasers and xenon lamps. Refining can be done with one zone moved through the ingot many times (passes) or by moving multiple zones simultaneously. The ends of a sample can be removed and the sample zone refined again as many times as needed. More passes are needed when K for a particularly detrimental impurity is close to 1. Volatile species can complicate the process but techniques to control or minimize melt losses have been developed.

Zone leveling is a way to produce material with uniform composition along its length. It has many of the features of zone refining. Once the zone has reached a steady-state composition, the species of interest will transfer from one interface to the other. The composition will be the same on both the melting interface and the solidifying interface. In both zone methods, crystal growth can be carried out during the purification or leveling procedures.

The floating zone technique is a very important variant of the zone melting method. It allows for crucible-free growth and eliminates possible contamination from the crucible material and also stresses due to differential expansion between the crystal and container. The method was invented and patented by H. Theuerer [125], a close colleague of Pfann, to grow ultra pure Si. For better uniformity, the rods can be rotated during growth. As mentioned before, the method is used commercially today for growing low or oxygen-free silicon. In recent years, automated commercial optical lamp heated float zone systems have become available and have permitted researchers from a variety of disciplines to grow crystals of a wide variety of materials for physical property studies [126].

The pedestal growth method is essentially a floating zone process. However, because the pull rate of the crystal (smaller diameter) is different from the push rate of the source rod (larger diameter) it has some characteristics of the Czochralski method. Parenthetically, one might classify the Verneuil method as a pedestal growth method since the crystal is grown on a pedestal and the molten zone is fed by molten powder rather than a solid rod. The first use of a pedestal growth technique was in 1958 by F. Horn [127] at the General Electric Corp. His method was a hybrid technique between the Czochralski and float zone methods. The charge in the crucible was only melted near the top surface and the crystal, of smaller diameter, pulled from this melt. As the crystal grew he changed the heater position to melt some more of the solid below. He grew boules of Sb-doped Ge having a more uniform composition than achievable by the Czochralski technique where the entire charge was melted. Dash [128], and Poplawski and Thomas [129] used this method to grow dislocation-free crystals of Si and Ge.

Two techniques that have been found particularly useful for producing small-diameter crystals for property studies are the *laser-heated* pedestal fiber growth (LHPG) [130] and micro pull-down (μ-PD) [131,132] methods. The LHPG method is a zone melting method in which, rather than a zone traversing a bar of material of uniform diameter, a fiber is grown from a source rod of larger diameter. The pedestal configuration was first used by Horn [127] and Poplawsky [129]. The source rod forms a pedestal

whose upper surface is melted with a small spatially fixed laser beam. A seed crystal is lowered into the melt and is withdrawn at a rate faster than the rate at which the source rod is pushed upward to maintain constant melt volume. It is a crucible-less method, minimizing contamination; the growth rates are much faster than bulk growth (mm/min) because the temperature gradients at the interface are very large (>1000 °C/cm). These high-growth velocities can lead to greater dopant incorporation and to the growth of metastable phases. The method is useful for incongruently as well as congruently melting compounds, although it is limited to systems where the vapor pressures are modest and dissociation is not a problem. The small diameter fibers were often found to have better crystalline perfection than bulk materials. The LHPG method has been used to grow fibers of a vast array of materials including oxides, halides, borides, carbides, and metals. Haggerty [133] was the first to use the pedestal method with laser heating. He grew LaB_6 single crystal fibers. A few years later, it was used to grow single crystal Nd:YAG fiber lasers [134]. One of the big advances in LHPG growth was the replacement of individual laser beams (two or four) with reflaxicon optics [135] giving a circular beam and a much more uniformly heated molten zone. An interesting discovery came about during the growth of $LiNbO_3$ fibers. Lithium niobate is a ferroelectric material whose as-grown bulk crystals contain numerous parallel and antiparallel domains. To be useful in nonlinear and other device applications, these have to be aligned after growth in an electric field at elevated temperatures. During the growth of $LiNbO_3$ fibers using a two-beam laser system, however, the small diameters and steep gradients led to single domain fibers when grown along the c-axis and a bi-domain fiber of opposing 180° domains when grown along the a-axis [136]. The axial gradients were responsible for the single domain c-axis fibers and the radial gradient a-axis fibers. These observations later led to a method to produce fibers with periodically poled domain structures by periodically shuttering one of the laser beams during growth [137]. Later, other periodically poled structures for quasiphase matching applications were produced by other methods.

In 1980, Mimura et al. [131] published a paper on the growth of KRS-5 fiber crystals using an inverted pulling system (modified floating zone technique). They had a crucible filled with melt on the top with a feed rod continuously feeding the melt as the fiber grew. At the bottom of the crucible was a long heated capillary tube with a shaper at the end. The growth interface was below the shaper. The growing crystal was pulled downward. This method was taken up by D. Yoon et al. [132] and called the micro-pull down method. It has been used successfully for many materials, and growth systems are available commercially. This method was reviewed by V. Chani [138].

1.3.1.9 Shaped Growth

Shaped growth generally means a method for producing a crystal with a predetermined cross-sectional configuration. The quest to develop such methods is associated with a need to reduce product cost and/or improve crystal quality. Notable savings can be achieved in device fabrication, such as cutting and polishing, reducing the loss of expensive material and reducing mechanical damage. In addition, the method usually

allows significantly higher than normal growth velocities, thereby increasing production rates. In a very real sense, shaped growth is a hybrid method that borrows from other more established methods. For example, some are simple modifications of the crystal pulling or directional freezing methods.

One might consider the Bridgman–Stockbarger method to be one of the first shaped growth processes because the crystal retains the shape of the ampoule used. It is not quite so obvious since almost all Bridgman crystals are grown from cylindrical ampoules. One of several recent exceptions being a paper by Feigelson and Route [139] on the growth of square cross-section crystals of $AgGaSe_2$ in vacuum-formed quartz crucibles. Using oriented seeds, they grew crystals not only aligned along the c-axis, but also so that the flat crystal side faces would to be normal to the [110] planes in which light propagates during type 1 phase-matched nonlinear interactions. A comprehensive review of the various shaped crystal growth methods is given in Ref. [140] and elsewhere.

Perhaps the earliest attempt at shaped crystal growth is attributed to the 1921 work of von Gomperz [70] at the Institute of Fiber Chemistry in Berlin-Dahlem. He worked in Michael Polyani's group. Polyani [141], reminiscing some 40 years later, recalled "Some metallurgists, interested in my work on the hardening of single crystals, told me of a method invented by Czochralski for producing metal crystals in the form of wires. It consisted in pulling out a thread from a pool of molten metal, so that the thread continued to solidify at the rate at which you were pulling it out. Erwin von Gomperz, who was doing his thesis with me, was put to growing single crystals of tin and zinc in this way. Unfortunately, the metal tended to come out in lumps, and the project was saved only by the intervention of Hermann Mark who covered the liquid metal by a sheet of mica with a hole in the middle, through which the thread came out as a smooth cylindrical wire. But for this ingenious intervention, our subsequent investigations of the plastic flow of metals might not have come about" [142].

In 1938, Stepanov at the Ioffe Institute in St. Petersberg began his extensive studies on shaping crystals during growth using wetted and nonwetted dies [143,144]. These dies have one or more capillaries or slots to transport melt from the crucible to the growth interface. The shape and height of the melt column is dependent on capillary properties such as surface tension, density, melt viscosity, impurities and wetting angle. Over many years, Stepanov's group produced a wide variety of shaped crystals including single and multibore tubes, rectangular bars, sheets, discs, etc. Shaped crystals of a number of different types of materials were grown, including oxides, metals and compound semiconductors.

Shaped growth in the United States began in the late 1960s with the preparation of sapphire filaments (later sapphire tubes for Na-vapor lights) by the edge-defined film-fed growth method (EFG) and single crystal superalloy turbine blades by directional solidification in complex molds [145,146]. The EFG process is in effect one of Stepanov's techniques, but it specifically focused on the advantages of wetted dies. It was discovered independently by LaBelle [147] who made a significant observation during his early attempts to pull sapphire fibers from a die placed in the melt surface. He noticed that the

melt wet the die and instead of being the diameter of the capillary within the die was the shape of the outer rim of the die. This was recognized by Mlavsky [148] as being of significant benefit and thus began extensive work on the EFG method. A sketch of the EFG method is shown in Figure 1.9. In a relatively short time after its discovery, it became a commercially viable technology. This was due to the early recognition that concurrent with experimental work, theoretical studies were needed to thoroughly understand the mechanisms involved in the process and ways to maintain better shape stability and rapid growth rates [149,150]. In 1980, an entire volume of the *Journal of Crystal Growth* was devoted to the subject of shaped growth methods. The EFG method has been successfully used in the commercial production of sapphire single fibers, tubes and ribbons and other materials. Today, the EFG method has gained an increasingly greater market share in the production of sapphire ribbons for GaN substrates used in lighting applications. Automated commercial EFG equipment can now be readily obtained, permitting companies to produce such wafers themselves.

The Stepanov and EFG methods are a meniscus-controlled process like Czochralski growth, and like it, growth is driven by crystal pulling. However, instead of pulling directly off the melt surface, the crystal is pulled from a suitable die face located above the melt surface. It can either float on the surface, like in von Gomperz's early experiments [70], held in a fixed position with respect to the crucible or moved during growth. The die position leaves the growth interface some distance above the hot melt surface and the thermal gradients are much steeper permitting enhanced growth velocities. The die material is chosen on the basis of its wettability with the melt and its reactivity. For sapphire, Mo dies have been used. In commercial systems the die can be moved during growth to maintain constant conditions and be equipped with an automatic monitoring system based on crystal weighing.

During the 1970s, the oil energy crisis led to serious efforts to produce silicon solar cells at a much lower cost than using cut wafers from Czochralski boules. This led to a robust effort to produce Si sheet at high growth rates. One of the most promising methods at the time was the EFG process. Extensive efforts went into adapting it to produce low-cost Si solar cells, but with limited success. Other innovative Si shaped crystal growth methods were studied during this time period, including the dendritic web process [151] in which a silicon dendrite is used a seed. It grows out laterally

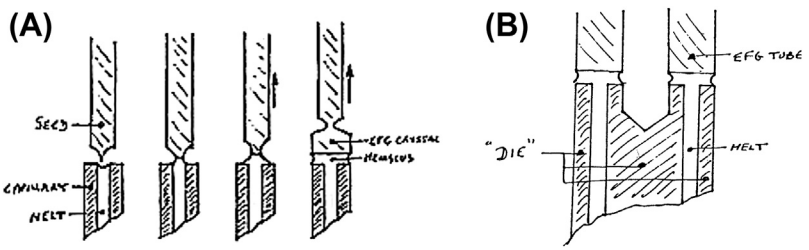


FIGURE 1.9 Illustration of the EFG shaped growth process. (A) the sequential steps involved in seeding and growing a crystal from a cylindrical die. (B) a die used to grow a hollow tubular crystal [148].

forming a thin sheet of silicon bounded by two new dendrites that define the width of the sheet. Some other methods include the ribbon against drop (RAD) method [152], where silicon is deposited as aligned grains on a carbon substrate, and the silicon on ceramic process [153], which is similar to the RAD process except that the substrate is a ceramic material. Two other methods of note are the ribbon to ribbon (RTR) [154] and horizontal ribbon (HRG) [155] growth methods. The former is a laser heated float-zone technique using a poly ribbon as the source and the latter involves pulling a ribbon (cooled from the top) horizontally from a free melt surface. The growth rates achievable in these processes are in the 5–10 cm/min range except for the HRG method where controlled cooling of the upper surface permitted growth rates of 10–40 cm/min. These methods are reviewed in Ref. [140].

The unidirectional casting method used for making single crystal jet engine turbine blades, as mentioned above, has had an important influence on aircraft performance. These blades, made from nickel-based superalloys, were found to have superior creep resistance if they have aligned grains [140] or better yet be one single crystal [156]. The method is like a Bridgman technique with the mold having the shape of the blade and extending down below is a zig-zag-shaped capillary tube mounted on a hollow pedestal that sits on a chill plate (see Figure 1.10). Growth is upward, initiated first from the melt in the pedestal by cooling the chill plate. This produces elongated grains along the mold's vertical axis, one of which will be in line with the capillary to provide seed selection. If by chance more than one grain makes it into the capillary from the pedestal, the crooked capillary will aid in seed selection.

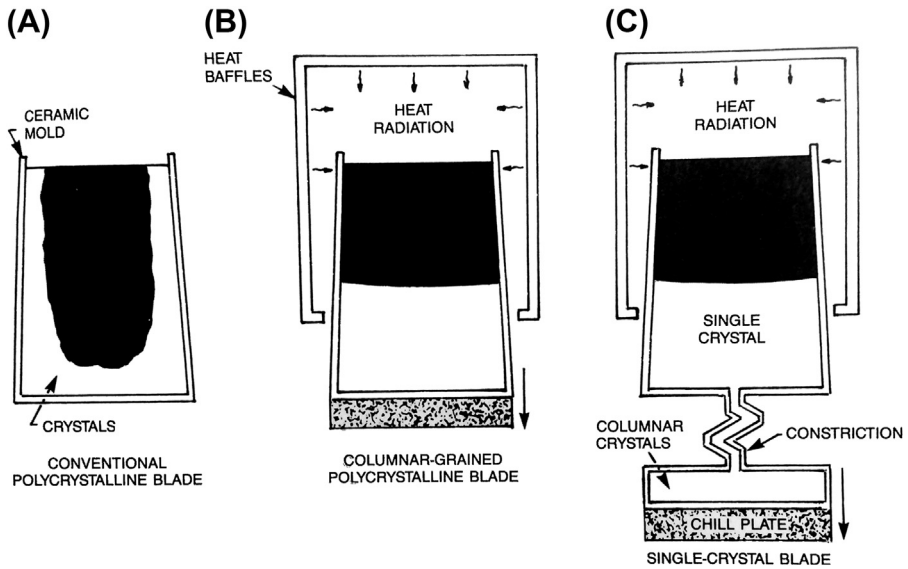


FIGURE 1.10 Schematic diagrams showing various methods for making jet engine turbine blades. (A) original casting method (polycrystalline), (B) single crystal growth by unidirectional solidification and (C) unidirectional growth with grain selection [146].

1.3.1.10 Skull Melting

The skull melting method is a quasi-crucible-less crystal growth method that was developed in the late 1960s in Russia, especially for growing large, high-purity oxide crystals for laser and gemstone production [157]. Since the 1970s it has become an important method for manufacturing cubic zirconia gemstones in a variety of colors (J.F. Wenckus et al. at the Arthur D. Little Company [158]). The method is adaptable to Czochralski and Bridgman growth methods using seed crystals and also in centrifugal casting.

In skull melting, a relatively large powder charge is contained within a water-cooled cylindrical Cu crucible surrounded by an RF heating coil. The RF field generates a magnetic field that in turn generates eddy currents due to ohmic losses in the material within the crucible. The process works for materials whose electrical conductivity increases with temperature, even through the melting stage. Due to the cold crucible wall, a skin (skull) of unmelted materials surrounds the melt and keeps it from coming into direct contact with the crucible, thus preventing contamination. Temperatures of over 3000 °C can be achieved and a wide selection of gas atmospheres are possible. Unless coupled with seeding, it is impossible to grow one single crystal due to the seeding effect of grains in the skull. For the gem industry, large slabs of crystals are retrieved from large-grained poly masses. The nature of the heat flow encourages the grains to grow along the vertical direction.

Cubic zirconia (ZrO_2) is used in the gem industry as a substitute for diamond since its optical properties, hardness, and fracture toughness are similar. Pure zirconia, however, undergoes a number of destructive phase transformations upon cooling. In order to grow single crystals from the melt, therefore, its composition has to be modified (stabilized) by doping to allow the high temperature cubic phase to persist to room temperature. The most common stabilizer is yttrium (YCZ), but CaO and MgO are also used. Concentrations of dopants vary from 10 to 40 mol%. These stabilizers work by creating many vacancies on the oxygen sublattice that prevent the cubic phase from transforming to phases of lower symmetry. Partially stabilized zirconia (PSZ) can be produced by reducing the dopant concentrations to less than 6 mol%. In these materials, part of the material transforms into the tetragonal phase creating a composite structure with excellent mechanical properties, making them attractive for applications such as drilling, threading, medical instruments such as scalpels, etc. [157]. The wide range of colors possible in zirconia gemstones are created using rare earth or transition metal element dopants.

Recently a Ukrainian group [159] developed a technique for growing large Tl:NaI scintillator crystals by a method similar to skull melting. Since the melting temperatures are quite low (~661 °C), they did not need RF heating and reconfigured the system from horizontal heating to vertical resistance heating. Basically, they hold a heater plate about 1 cm over the charge, which is held in a rectangular aluminum tray sitting on the bottom of water-cooled vacuum chamber. Since the heater is smaller in area than the container, only the center part of the charge is melted leaving a skull 5–10 mm thick surrounding

the melt and keeping it out of contact with the aluminum. Growth is achieved by slow cooling from the bottom up by lowering the temperature of the upper heater. It is not clear what the grain structure of the resulting slab is like, since seeding is not used and the plates produced are quite large. However, the scintillator properties are in line with those of single crystals produced by other methods and in a more cost-effective manner.

1.3.2 Solution Growth

1.3.2.1 Introduction

Solution growth methods involve dissolving material in a liquid (or gel) medium and then recrystallizing it under controlled conditions to produce a crystal of a desired size, shape and perfection for a specific application. The control of crystal shape and size can extend from very large crystals for optical applications down to fine powders for pharmaceutical, agricultural, or specialty chemical uses. The solvent media may be a low-temperature solvent like water or a high-temperature flux like PbO. Pressure-enhanced solution growth (the hybrid hydrothermal growth method) has also been a commercial success for the preparation of large crystals of quartz. The most common solvent used is water, and an impressive number of inorganic salts have been converted into single crystals using this technique, some weighing over 50 pounds. Other solvents include organic liquids (for the growth of organic crystals) and liquid ammonia.

Growing crystals from water solutions was undoubtedly the earliest effort by early man to replicate what he observed in nature. Natural salt ponds drying up in the hot summer months and then redissolving during winter rains assuredly piqued his interest and led him to experiment. Since salt became such an important commercial product, it is not surprising that this material and method became one of the first industrial crystal growth activities. Sugar was another material of early commercial interest. Through trial and error, a rudimentary understanding of saturation and supersaturation began to develop and, along with techniques such as seeding and solution homogenization (via stirring), better control of nucleation, crystal size and purity was achieved. Later, the role of additives to enhance growth behavior and tailor crystal properties was incorporated into the growth procedures. The eventual use of solubility diagrams greatly aided crystal growers in choosing appropriate growth conditions, i.e., temperature and composition regions. In these early days, as today, control of purity and size were of great commercial significance. Several basic methodologies are employed in solution growth: (1) controlled evaporation, (2) temperature programing, (3) mass transport in a concentration gradient at constant T and (4) changing the composition of the solution (salting out method).

1.3.2.2 Aqueous Solution Growth

Like all other crystal growth methods, a variety of modifications have been made over the years to facilitate the growth of a specific type of material, and to achieve an appropriate dimension and degree of crystalline perfection (purity, homogeneity, strain,

etc.) required for the application intended. Materials vary so widely in their thermodynamic and kinetic properties that even the growth of related materials of similar composition and structure can require changes in growth procedure or even in overall methodology.

In 1901, G. Wulff [160] published his famous theorem on the influence of surface energy on equilibrium shape of a crystal (morphology). It established a relationship between the crystal habit and the structure of crystals. It was derived from studies on the growth rates on different faces of crystals grown in water solutions. He used a rotating cylindrical crystallizer, in which a seed crystal was placed along the axis of the cylinder at its center-point. This allowed the crystal to grow out of contact with the vessel walls and be exposed to nutrient equally on all faces. Other early pioneers advancing the development of water solution methods during the first half of the twentieth century included Kruger and Finke [161], and Valeton [162]. Kruger and Finke were the first to investigate growth under constant temperature and supersaturation conditions. Their apparatus, shown in Figure 1.11, had in common two vertical chambers connected by upper and lower tubes through which solution passed in a specific way. One chamber contained source material and the other a stirring paddle (growth chamber) to move nutrient and depleted solution from one to the other. After equilibration of the growth chamber in a slightly undersaturated state, a seed crystal was added. Valeton's apparatus had a more precise way to control temperature in each bath. These methods utilized mass transported from the source chamber to the growth chamber to control the growth process. Crystals of potash alum and potassium sulfate were grown in these early experiments. In 1916, Nacken [163] developed a similar but more sophisticated apparatus using a vertical configuration for solute transport. Some 30 years later, Walker and Kohman [164] at Bell Laboratories developed a large-scale commercial crystallizer similar to these earlier methods known as the constant temperature process. Together with Holden's contributions on seed mounting [165], this apparatus was capable of growing four large EDP (ethylene diamine tartrate) piezoelectric crystals at a time. These crystals were used to replace natural quartz in telephone circuits. Crystals weighing up to 40 pounds could be

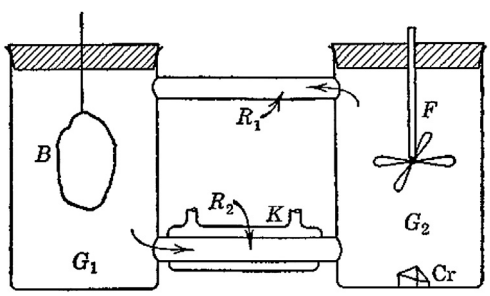


FIGURE 1.11 The aqueous solution crystallizer used by Kruger and Finke [161]. The nutrient was contained in a porous bag in heated beaker G_1. The stirrer in beaker G_2 recirculates saturated from G_1 through a water-cooled tube K, where is becomes supersaturated, into beaker G_2 where the crystals grow. Large potassium sulfate crystals 2 cm in size were produced in this reactor.

grown by this method. The method involved several large chambers with solution in one chamber saturated at one temperature being continuously fed the into the crystallizer chamber being held at a slightly lower temperature, thereby providing the supersaturation conditions necessary for growth.

In 1919, R.W. Moore [166] working at the General Electric Company needed large Rochelle salt crystals for electrical property studies. Since suitable material was not available commercially, he was forced to develop his own method. At first he started with the method of Kruger and Finke [161]. This proved unsuccessful. After thoroughly studying the available literature he came up with a new and simple approach based on the temperature lowering method. First, a saturated solution was formed 10–15 °C above room temperature, the solution decanted to separate it from the excess salt and then filtered. After heating the saturated solution 7–8° above its saturation point, it was poured into a vessel containing small seeds suspended on silk threads or metal wires, covered with a glass plate and then placed in water bath at 0.5 °C above the saturation temperature. From that point onward, the temperature was lowered to cause the seeds to grow. No means of stirring the solution was provided in these early experiments.

Moore's temperature lowering method was eventually modified to supply some fluid movement and distribute nutrient more uniformly to all the faces. The so-called rocking tank method was applied in 1947 by Walker [167] to the growth of large ammonium dihydrogen phosphate (ADP) crystals needed for submarine detection. The tanks were large rectangular trays that were gently rocked to replace depleted solution at the growth interfaces with fresh supersaturated material. Like EDP, crystals as large as $6 \times 6 \times 20$ in were produced by this method. Many crystals could be grown in each tray and for production (Western Electric Co.) rooms were filled with many rocking trays. Evaporation was inhibited and the room temperature had to be carefully controlled. One famous name associated with water solution growth was Alan Holden [165]. Aside from his research at Bell Laboratories, he also wrote a very popular book on the subject [168]. Numerous amateurs have used it to initiate crystal growth experiments. In 1949, he introduced the "rotary crystallizer" originally to grow EDP and then later for ADP crystals. It consisted of a large, one-foot-diameter cylindrical vessel holding the solution. It was heated from the bottom by two concentric heaters, an inner one to keep the bottom center somewhat under saturated. In this way, errant crystallites that have fallen to the bottom would dissolve. The outer heater controlled the overall solution temperature. The important feature introduced by Holden was the rotating seed holder (called a "spider"). The seed crystals were mounted on spokes emanating from the rotation shaft. Several sets of spokes holding the seeds were used along the vertical axis. The seeds were rotated first in one direction for a selected period of time and then in the other direction. The system was sealed and some water condensed on the upper lid forming droplets. When large enough, the droplets fell to the solution surface keeping it under saturated and thus preventing the nucleation of spurious grains.

Most of the growth methods mentioned above have inherently slow growth rates (0.5–1 mm/day) due to low solution supersaturation. The higher supersaturation

needed to yield increased growth rates were hard to achieve and control. In 1983, Loiacono et al. [169], using a three-stage crystallizer (modeled after Walker and Kohman's growth system from 1948 [164]), and under conditions of constant temperature and supersaturation were able to achieve 5 mm/day growth rates for large potassium dihydrogen phosphate (KDP) crystals. KDP is an important nonlinear optical material for modulating lasers. It has been grown commercially for many decades using aqueous methods similar to the ones described above. The need for even larger crystals of KDP for electro-optic switch and frequency converter plates in inertial confinement fusion research led to a big advance in solution growth methodology. It started in 1982 in Rashkovich's group at Moscow University. There they developed a rapid growth process for KDP from water solutions [170]. Over the next several decades, that work was taken up by Zaitseva and colleagues at the Lawrence Livermore National Laboratories in California [171]. They demonstrated that the standard Holden [165] crystallizer with temperature reduction could be used to grow large high-optical-quality KDP and deuterated KDP (DKDP) crystals up to 50 cm on a side at rates 10–100 times faster than older methods and without spontaneous nucleation and macroscopic defects. From their research on growing large crystals under fast growth conditions, they were able to develop a more thorough understanding of the mechanisms involved in solution growth. In addition to the influence of temperature, supersaturation and dislocations on growth rates, they realized the importance of impurities, mass transport (via high solution velocities) and in having a highly stable growth system. It was found imperative to control secondary nucleation. The most important feature of their rapid growth process was using highly supersaturated solutions (70–76 °C) coupled with elaborate techniques for preventing spontaneous nucleation. Toward that end a continuous filtration system and a seed protector were important modifications to the growth system.

1.3.2.3 Growth of Biological Macromolecules

Determining the crystal structure of complex biological molecules such as DNA, proteins, enzymes, etc., is important to both our understanding of animal and plant biology and functionality and our ability to develop pharmaceutical products to combat various illnesses that afflict these species. During the past century, protein crystallographers have slowly worked out the structures of a myriad of important species using X-ray diffraction methods. To accomplish this, researchers needed small, high-quality single crystals of controlled composition.

The first recorded protein crystallization experiments were done by German scientist F.L. Hunfeld in 1840. He prepared acicular crystals of earthworm hemoglobin by pressing blood between glass slides and allowing it to slowly evaporate. Since then, many techniques have been developed to prepare such crystals and the quality and size of the crystals produced were essential to the success of the structural detail obtained. The working out of the structures of myoglobin (1950) and hemoglobin (1955) using heavy metals covalently bonded to the protein led to the Nobel Prize for their researchers.

Another Nobel Prize winner who grew crystals in order to study their crystal structures was crystallographer Dorothy Crowfoot Hodgkin (1910–1994) [172]. As a child, Hodgkin was influenced by books that described how to grow crystals and on the interaction of X-rays with crystals. She won her prize in 1962 for her part in unraveling the crystal structure of the Vitamin B-12 molecule [173]. She also, together with Kathleen Lonsdale, grew crystals of penicillin and potassium and rubidium benzyl-penicillin. One of her major accomplishments was deciphering the structure of insulin.

While the crystal growth of biological macromolecules was primarily the domain of protein crystallographers and biologists during most of the century, in the 1980s the effort expanded to include experts more familiar with the theoretical and practical aspects of crystal growth (albeit small molecule materials). This came about in two ways. First, NASA had received requests from the crystallography community to fund protein crystal growth experiments in the low gravity environment of outer space. It was believed that the quality and size might be enhanced under these conditions. At the same time, NASA had been funding a variety of small molecule crystal growth experiments in low gravity with some promising results. NASA decided to try to engage some of the small molecule crystal growth community in the protein growth field. Several such programs were funded, one of which was my group at Stanford University. I immediately realized that we did not know enough about biological species to carry out his program successfully, and so when I found out that protein crystal growers did not have their own forum to discuss growth problems of mutual interest, I decided to bring them to Stanford basically to teach us about the field. Together with Alex McPherson [174], and with support from the American Association of Crystal Growth and NASA, we organized the first international conference on protein crystal growth at Stanford University in 1985. It not only brought together protein crystal growers for the first time, it also included well-known scientists and engineers from the small molecule crystal community. A total of 140 attendees were present. It was a somewhat contentious meeting at first, but as it proceeded, both sides, who spoke quite a different scientific language, came to understand more clearly the relevant issues, i.e., the physics behind the growth process and the influence of various processing parameters on the size and quality of the crystals produced. This international conference series has been held regularly every since.

Crystals of proteins and other biological species can be grown by a number of techniques including dialysis, sequential extraction, interface diffusion, vapor diffusion (plates, hanging or sitting drops), via pH and temperature changes, evaporation and in thermal (concentration) gradients. McPherson's original book, *Preparation and Analysis of Protein Crystals* [175], gives a comprehensive review of growth methods (see also Ref. [176]). Purification of starting materials and the composition of the growth solutions, like in most small molecule systems, are critically important to produce suitable crystals for X-ray structural analysis. Most of the growths are done in small batches. To establish the correct crystallizing conditions, a matrix approach is often used. Here, small samples with a systematically varied concentration of protein, salting agent,

solvent, etc., are placed within many cells and crystallized under the same conditions. Regions in the matrix that contain crystals are then regrown on a more refined compositional scale to enhance the results. In effect, this is the combinatorial chemistry approach and one that lends itself to automation using robotics to meter out the desired quantities into each cell. This latter approach was pioneered by Ward et al. [177] during the mid-1980s and is in wide use today.

One final comment worthy of mentioning again (see Scientific Study section) is that the large size of the growth units in biological macromolecules gave crystal growth scientists a unique opportunity to dynamically study the morphology and kinetics of step and ledge movement (including step bunching) during growth using the relatively new atomic force microscopy technique. The first such in situ studies were carried out in 1995 by Land et al. [178] and Malkin et al. [179].

1.3.2.4 Growth from Gels

In 1896, the German chemist Raphael E. Liesegang slowly put a drop of silver nitrate-water solution onto a thin gel layer containing potassium dichromate, and in doing so discovered the precipitation ring phenomena named after him [180]. This initial discovery stimulated a strong interest in understanding how the process worked. A gel is a semisolid containing small pores of angstrom dimensions in which a variety of salts can be dissolved. Early efforts on growing crystals in gels include the work of Hatschek [181] in 1911 and Dreaper [182] in 1913. The former grew small crystals of gypsum by letting sulfate ions diffuse in a gelatin containing a dilute solution of calcium chloride and the latter lead chloride crystals in a test-tube-shaped vessel. The idea for growing crystals in a gel media was stimulated by the research work of Fisher and Simons [183] in 1926. They were intrigued by some earlier work with gold and copper crystals produced by the reduction of their metal salts in a silica gel and the coincident occurrence of gold in quartz veins. From their early experiments, they predicted that this method would be "far-reaching" and this enthusiasm caught the attention of later researchers and became an area of vigorous research, particularly from the early 1960s onward. The work by Heinz Henisch's group at Pennsylvania State University stimulated researchers around the world and was summarized in his book *Crystals in Gels* [184].

The gel growth method has been used to prepare an impressively wide range of inorganic and organic crystals, including proteins. Gels provide a medium where mass transport is by the slow diffusion of suitable ions to a region where they can react during crystallization. It is a convection-free method and the crystals, when nucleated under carefully controlled conditions, are suspended from one another. These factors, plus the near room-temperature growth conditions purportedly result in higher crystal quality. Crystal dimensions can vary from micron to centimeter sizes depending on the system under study, but typically they only reach mm sizes. Like other solution growth crystals, they exhibit growth rate anisotropy and faceting. Typical gels used are silica hydrogel (sodium metasilicate), agar (derived from seaweed) and gelatin; however, many other gel compositions have been used as well. Crystals can be grown within gels by a number of

techniques, including (1) chemical reaction, (2) complex dilution, (3) chemical reduction and (4) solubility reduction. An extensive list of crystals grown is given in Ref. [185]. During the last decade or so, interest in gel growth has diminished along with the number of publications.

1.3.2.5 Nonaqueous Solution Growth

Organic crystals are useful for a number of applications including semiconductors and scintillator devices. Organic materials, like other substances, vary widely with respect to their thermodynamic and physical properties. Therefore it is not surprising that a suitable crystal growth method will depend on the specific properties of the material in question. They can be grown by a variety of common crystal growth techniques including vapor, melt and solution methods. Organic materials that melt without dissociation are prime candidates for melt growth methods. Others have been grown in solution or by vapor phase techniques. Solution growth methods usually involve organic solvents such as ethyl alcohol, acetone, hexane, and carbon tetrachloride. The techniques used are similar to water solvent methods and include solvent evaporation, slow cooling or heating, vapor diffusion and liquid–liquid diffusion and are nicely summarized in Ref. [186]. One of the recent examples of solution growth using organic solvents is the work at the Lawrence Livermore Laboratories on the growth of large, high-quality crystals of trans-stilbene ($C_{14}H_{12}$) for fast neutron detectors [187]. The solvents used were toluene or anisole, the latter preferred due to its lower evaporation rate. Melt growth techniques did not yield large, high-quality crystals. Building crystallization systems to withstand the organic solvents and by using the temperature reduction method together with rotation, very high-quality crystals in dimensions up to four inches have been produced.

1.3.2.6 High Temperature Solution (Flux) Growth

1.3.2.6.1 Bulk Crystals

As with other solution growth methods, the high temperature flux growth method also relies on the careful control of the supersaturation and melt composition. Like the low temperature processes, there are three general methods for controlling supersaturation: (1) slow cooling, (2) evaporation and (3) transport in a concentration gradient. In its early incarnations, the method was unseeded and crystals grew on the surface of the melt where supersaturation is usually greatest (due to volatility) or on the crucible walls where heterogeneous nucleation is favored. Later, the use of seeds or cooled probes helps facilitate growth. Generally solvents are classified as common ion or noncommon ion fluxes. An example of the former is the growth of (Ba,Sr)TiO_3 from excess TiO_2 melt [188], and the latter, the growth of $Ye_3Fe_5O_{12}$ from BaO–B_2O_3 based fluxes [189]. The BaO–B_2O_3 flux, while used early on, was not nearly as successful as PbO–B_2O_3 or PbO–PbF_2–B_2O_3 fluxes that form ionic solutions. The 1975 book *Crystal Growth from High Temperature Solutions* by Elwell and Scheel [190] still remains the most thorough, encyclopedic treatment on the history, theory and methodology of flux growth.

Crystal growth from high-temperature solutions probably began during the nineteenth century. One of the earliest documented studies (1823) was by Friedrich Wöhler (1800–1882), a famous German chemist who used a flux-reaction technique to grow sodium tungsten bronze (Na_xWO_3, where x is ≤ 1) crystals by passing H_2 gas over a Na_2WO_4 flux [191]. He is also famous for his synthesis of urea and the codiscovery of Be, Si and silicon nitride. By mid-century, early experiments were of a similar nature and a variety of materials were produced including CdS (using a $CdCl_2$ solvent and H_2S reactant), alkaline earth and transition metal oxides, silicates and sapphire [190]. Later in the century, more traditional flux methods were used to grow a variety of binary, ternary and higher order compounds using chemically compatible solvents. Some examples include Doelter [192], who in 1886 grew Ag_2S by dissolving and recrystallizing it from AgCl or $SbCl_3$ melts. AlB_{12} crystals were grown from B_2O_3 melts by Wöhler and Deville in 1857 [193], and ruby was grown by Fremy and Feil in 1877 [194] using PbO as the flux (which from the 1950s onward became a popular flux for growing oxide crystals for optical applications). The results varied from laboratory to laboratory, and generally the crystal sizes were on the small side. Over 100 flux-grown compounds and their solvent phases are listed in Ref. [190]. In this time period this was probably the most important method used for preparing crystals of non water-soluble materials.

Solvent requirements include (1) a suitable melting temperature and solubility in the temperature range of interest, (2) a temperature coefficient of solubility, (3) low volatilization (an issue with PbO and halide fluxes), (4) compatibility with available crucible materials (reactivity and wettability) and (5) a relatively low viscosity. In the 1950s, the flux growth method again became an important adjunct to other developing crystal growth methods such as the Czochralski and Bridgman methods. It was particularly important for compounds that melt incongruently, have high vapor pressures at their melting temperatures, are refractory with excessive melting temperatures or have destructive phase transformations, etc.

After World War II, the flux growth method gained interest commensurate with device researchers' interests in finding new and better materials for optical, electronic and magnetic applications. The use of this method was very extensive during the 1950s through about 1980 and took place in many laboratories around the world. As a result, it is only possible to cite a few illustrative examples in this brief historical review to give a sense of what growth activities were like during this period. In 1964, Edward Giess [195] successfully prepared $Cr:Al_2O_3$ (ruby) using a PbF_2 flux. Lead fluoride is quite volatile, and so it was often combined with PbO with some added B_2O_3 to stabilize it. At about the same time, Stanley Austerman [196] used the flux growth method to grow BeO crystals. BeO substrates were of interest because of its very high thermal conductivity and therefore its ability to remove heat from electronic devices. This would permit higher power operation. Austerman grew BeO from a Li_2MoO_4-based flux. Later, Newkirk and Smith [197] grew BeO from PbO-based fluxes. In Russia, V.A. Timofeeva's group was very active during this time period, studying the flux growth of many different oxide compounds including Al_2O_3, Cr_2O_3, Fe_2O_3, rare earth oxides and the garnets $Y_3Fe_5O_{12}$, (YIG)

and $Y_3Al_5O_{12}$ (YAG). In the latter case, she investigated how growth defects formed as a function of the flux growth conditions and solvent composition [198].

Single crystal garnets ($A_3B_5O_{12}$) were of great interest for laser, microwave and ultrasonic applications. In 1958, Nielsen and Dearborn at Bell Laboratories were the first to report on the flux growth of $Y_3Fe_5O_{12}$ (YIG) [199]. Major improvements in size and quality were achieved by adding impurities such as CaO to a PbF_2–PbO–B_2O_3 flux. Van Uitert et al. [200] worked on the growth of large, optical-quality $Y_3Al_5O_{12}$ (YAG) crystals in very large platinum crucibles. Remeika, whose extensive crystal growth activities also included the flux growth of YIG, found that pure YIG crystals contained small amounts of Fe^{4+}, which resulted in reduced optical quality [201]. By adding small additions of tetravalent ions such as Si, Sn and Ge to the flux, the problem was eliminated. Some examples of flux grown crystals are shown in Figure 1.12.

Flux growth methods vary from simply slow cooling a melt without seeding (self-nucleation) to more complicated seeded growth techniques. The bottom cooling method helps control nucleation in unseeded melts. A small supersaturated region is created in the melt to limit the volume in which nucleation can take place. This can also be used to prevent a seed from dissolving before it starts to grow (like in the heat exchanger method). In 1955 and 1956, Reisman and Holtzberg were the first to prepare single crystals of potassium niobate $KNbO_3$ (KN) and potassium tantalate $KTaO_3$ (KT). These compounds were of interest for ferroelectric and piezoelectric applications [203]. A K_2O flux was found suitable for this purpose. KN in particular became a very important material for efficient direct diode doubling and other NLO frequency conversion processes, such as generating blue light from a Ti:Sapphire laser via critical phase matching. Over the years, large-scale

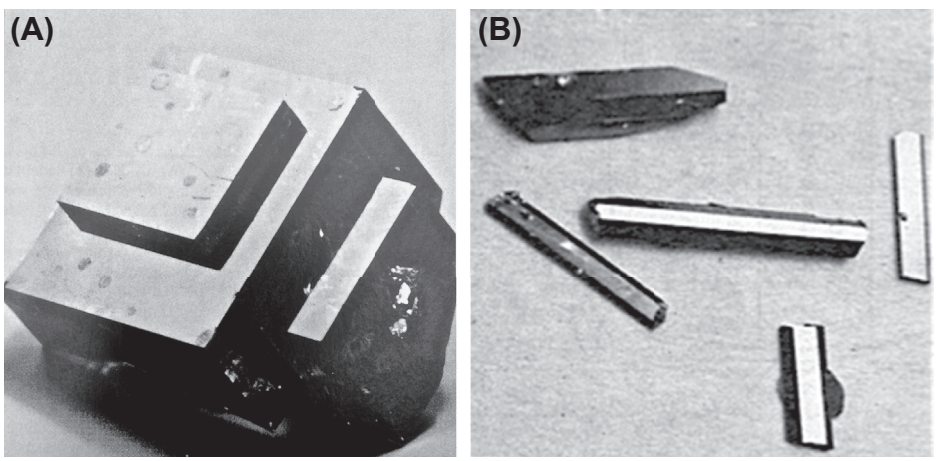


FIGURE 1.12 (A) A photograph of a large ($3.5 \times 3 \times 2.5$ cm), flux-grown crystal of $GdAlO_3$. It contained large inclusion-free regions. A PbO–PbF_2–B_2O_3 flux containing some other minor additives was used. The large size and quality is attributed to ACRT stirring method [202]. (B) Photograph of some highly facetted acicular crystals of $CdGeAs_2$ grown from a Bi flux (author). Some are solid, others contain a core of solidified Bi solvent. Strong growth rate anisotropy is evident.

crystal growth methods capable of producing 100 gm crystals were developed [204,205]. While the K_2O flux was still used, top seeding was added later to control nucleation and growth conditions. This seeding technique, now known as the top-seeded solution growth method (TSSG), became a very important flux growth method. The first description of this method was in a 1958 paper by Miller [206]. He used a seed crystal mounted on a rod that rotated in a reciprocating fashion. Without pulling, he grew KN crystals up to 15 gm from a charge containing K_2CO_3 and Nb_2O_5. He also used an electrical circuit between the crucible and seed to determine the exact moment when the seed touched the melt. The method was used later by Linares for YIG growth [207] and at MIT by Belruss et al. [208] for the growth of $SrTiO_3$ and $BaTiO_3$ crystals from melts containing excess TiO_2 and GeO_2. More recently, the method has been used for growing beta barium borate crystals (β-BaB_2O_4, BBO) from melts containing some Na_2O [209] or other solvents to lower the $\eta\varepsilon$ viscosity and to permit growth at temperatures below the α–β phase transformation temperature. BBO and other similar compounds like LiB_3O_5 (LBO) and $CsLiB_6O_{10}$ (CLBO) are very useful for optical applications in the visible and ultraviolet regions.

As seen with aqueous solution growth, stirring during growth is very beneficial to the enhancement of crystal quality and growth rates. In addition to top seeding with rotation that provides some fluid flow, other methods have been devised. One useful approach is the accelerated crucible rotation technique (ACRT). It was first used by Nelson and Remeika [210] in 1964 for pregrowth stirring. Scheel and Schultz-Dubois first demonstrated its usefulness during growth in 1971 [104]. The method is very helpful with growth from volatile melts that need to be grown in sealed crucibles. No moving parts, like stirring rods, need to be placed in the melt. The method relies on acceleration and deceleration of the crucible, thereby decoupling the fluids movement from the crucible's trajectory in a periodic fashion. Two major flow mechanisms, spiral shearing distortion and Ekman-layer flow, are operative during acceleration and deceleration. It has been found to limit nucleation and to help produce large, inclusion-free crystals. It was first applied to the growth of $GdAlO_3$ crystals from a PbO–PbF_2–B_2O_3 flux and yielded the largest such crystal to date [202].

One other method worth mentioning was that devised by Tolksdorf [211] at the Phillips Central Laboratories in Hamburg. In the past, one of the problems with growing YIG crystals from the volatile PbO–PbF_2 flux was that it redissolves below 950 °C. To prevent this, Nielsen [212] poured off the flux at 1040 °C outside the furnace. Unfortunately, the crystals cracked due to thermal shock. An improvement on this method was by Grodkiewicz, Dearborn and Van Uitert [213]. They punctured the bottom of their large platinum crucibles draining off the melt. This was expensive, as the flux material could not be reused. In Tolksdorf's method he used a sealed crucible that could be rotated on it axis. It was half filled with a PbO/PbF_2/YIG melt and, after the crystals were grown by slow cooling, the crucible was spun 180° separating the crystals from the melt. In a similar way, he could mount a seed on the empty side, and when the melt was saturated, could rotate the seed into the melt, slow cool to initiate growth onto the seed, and when done, rotate the crucible back to its original position to remove the flux from

the crystal. Toldsdorf's method and apparatus was later used to grow $KTiOPO_4$ (KTP) crystals (an important nonlinear optical material) using the flux method.

1.3.2.6.2 Thin Film Liquid Phase Epitaxy

In addition to bulk crystal growth, high-temperature solutions have also been used to grow thin films of semiconductors, oxides (magnetic and electro-optic) and various other compounds. The method is known as liquid phase epitaxy (LPE). The LPE method involves the crystallization of a single crystal or crystallographically oriented layer on a substrate in contact with a liquid phase. The substrate (usually a single crystal wafer) may be either of the same base composition (homoepitaxy) or a different composition (heteroepitaxy). The field of epitaxial growth extends well beyond the LPE method to include numerous vapor phase depositions techniques to be discussed later. While these other methods are very important, the LPE has certain advantages including (1) greater crystalline perfection due to the near equilibrium growth conditions and use of a near perfect substrate template, (2) better stoichiometry control, (3) higher growth rates due to higher solutes concentrations and (4) lower cost compared to other methods. A comprehensive review of LPE field is given in Ref. [214].

Epitaxial deposits have been found in various natural mineral formations. One example is rutile crystals growing on hematite facets [215]. In 1836, Moritz Frankenheim (1801–1869), a German physicist, was the first of many researchers to observe epitaxial growth in the laboratory when he produced oriented crystals of sodium nitrate on a cleaved surface of calcite crystals [216]. In 1906 Baker did some experiments on the orientation of crystals growing from droplets crystallizing on cleaved surfaces [217]. In 1928, Royer followed this line of research [218] and using X-ray diffraction analysis, was the first to describe the requirements for lattice matching between the film and substrate to achieve epitaxy (an orientation relationship between the layer and the substrate).

In the 1960s, attention turned to thin film semiconductor devices. At the RCA Laboratories, H. Nelson [219] was the first to develop an effective LPE method for growing epilayers of GaAs on GaAs substrates (homoepitaxy). He used a horizontal graphite boat system with a GaAs + Sn melt at one end. When the substrate, located at the other end was at 640 °C, the boat was tilted and the melt flowed over the substrate. After cooling for a period of time the melt was poured back off, leaving the substrate covered by a single crystal layer of GaAs. Other LPE techniques involved dipping substrates (either vertically or horizontally) into an appropriate solution and then, after deposition, withdrawing them back out. Substrate rotation has also been used to achieve better uniformity. By using these various methods p-n junctions could be produced using doped layers. These epilayers were used in a number of important device applications including GaAs [220] and $Al_xGa_{(1-x)}As$ [221] lasers and the extensive commercial production of light emitting diodes (LED's). The LPE method has also been used to prepare thin film structures from silicon, germanium and their solid solutions, II–VI compounds such as ZnSe, CdTe, $Hg_{1-x}Cd_xTe$ (MCT), SiC, the III–V nitrides (AlN and GaN) and many other alloy compositions [214]. As with any flux growth method, the solvents have to be tailored to the specific film

composition to be grown and the substrate and its orientation carefully chosen. The requirement for high-quality substrates for both LPE and vapor phase epitaxy have kept the bulk crystal growth industry very active.

In addition to semiconductor research, various applications for magnetic garnet crystals (e.g., $Y_3Fe_5O_{12}$ and $Ca_{2x}Bi_{3-2x}Fe_{5-x}V_xO_{12}$) were developing during this same time period. LPE became an important method for the preparation of various types of microwave, magneto-optic and bubble memory thin film devices. One of the first attempts to use the LPE method for YIG film deposition on garnet substrates was in 1965 by Linares et al. [222]. In 1968 Linares [223] grew high quality YIG films on GGG using the tilting boat method and a lead borate flux. Magnetic bubble memory thin film devices became an important research activity during the 1970s [224]. These thin film structures were once expected to replace Si-based memory chips and various groups extensively studied both their preparation and properties. Magnetic garnet single crystal LPE films were typically grown from PbO-B_2O_3 fluxes. The substrates used were nonmagnetic $Gd_3Ga_5O_{12}$ (GGG). It is a very good lattice match with YIG. Various techniques such as substrate dipping and rotation during growth were studied. Withdrawal after growth could be problematic due to cracking and film pealing. The process gave high growth rates and crystalline perfection, film thickness uniformity, and compositional homogeneity. Over time, ever more complex film compositions evolved to enhance their properties [225,226]. The technology reached its zenith with the growth of bismuth-doped rare earth iron garnet thick films up to one-half mm thick from Bi_2O_3–PbO–B_2O_3 fluxes onto large lattice parameter-matched Ca–Mg–Zr substituted GGG substrates.

In 1986, Bednorz and Müller discovered Hi–Tc superconductivity while working at the IBM research laboratories near Zurich [227]. The ceramic material they produced was an oxygen-deficient Ba-doped Lanthanum cuprate ($La_{2-x}Ba_xCuO_4$), a perovskite-like compound that exhibited zero electrical resistance at 35 K, twice the highest transition temperature achieved to date. This set off a whirlwind of research activities to find other cuprates with even higher Tc's including $YBa_2Cu_3O_{7x}$ (92 K) followed by $Bi_2Sr_2Ca_{n-1}Cu_nO_{4+2n+x}$ ($n = 1$, 2 and 3) with a Tc between 85 and 110 K. Thallium- and mercury-based cuprates had even high Tc's, the latter a record at 134 K. At this point a great effort was made to grow single crystals of these compounds for physical property studies and to enhance their properties. Due to the complex nature of the phase equilibria in these systems, crystal growth was very complicated, making difficult the preparation of large, high-quality single crystals. All manner of bulk and film deposition methods were tried with varying success. In Scheel et al. [228], the LPE method was used to prepare very flat, high quality epitaxial layers of $YBa_2Cu_3O_{7-x}$ on $NdGaO_3$ substrates. Step heights were between 1.2 and 7.2 nm and did not exhibit the spiral islands found using vapor phase deposition techniques.

1.3.2.7 Hydrothermal Growth

Hydrothermal growth is a solution growth method operated at modest temperatures and elevated pressures. Byrappa and Yoshimura [229] authored an exhaustive treatise on the

history and technology behind the hydrothermal growth process. The subject has also been discussed to a lesser extent in numerous other books and journal articles, for example [230]. The process involves the controlled growth of crystals in an autoclave (see Figure 1.13(A)) onto seeds immersed in a water solution containing the nutrient and usually a mineralizing agent. The driving force for growth is the solubility difference generated by a temperature gradient. The method has several advantages. Growth takes place below the material's melting temperature and often below a destructive phase transformation (e.g., α-quartz, the low-temperature polymorph of SiO_2). Since the growth process takes place in a sealed system, the atmosphere can be modified to suit the material being grown (i.e., maintaining an oxidizing or reducing environment). In addition, the method generally produces less stress on the crystal and can lead to an increased crystalline perfection. Another attractive feature of hydrothermal growth is that the growth rates are relatively fast compared to other solution growth methods. High-pressure vessels can be made of various materials depending on the temperature and pressures required.

Hydrothermal growth's principal use has been for the commercial growth of large, highly perfect (dislocation-free) α-quartz crystals for piezoelectric applications. A rack of crystals produced from one large-scale commercial autoclave is shown in Figure 1.13(B). Piezoelectric materials such as quartz generate an electrical polarization when subjected

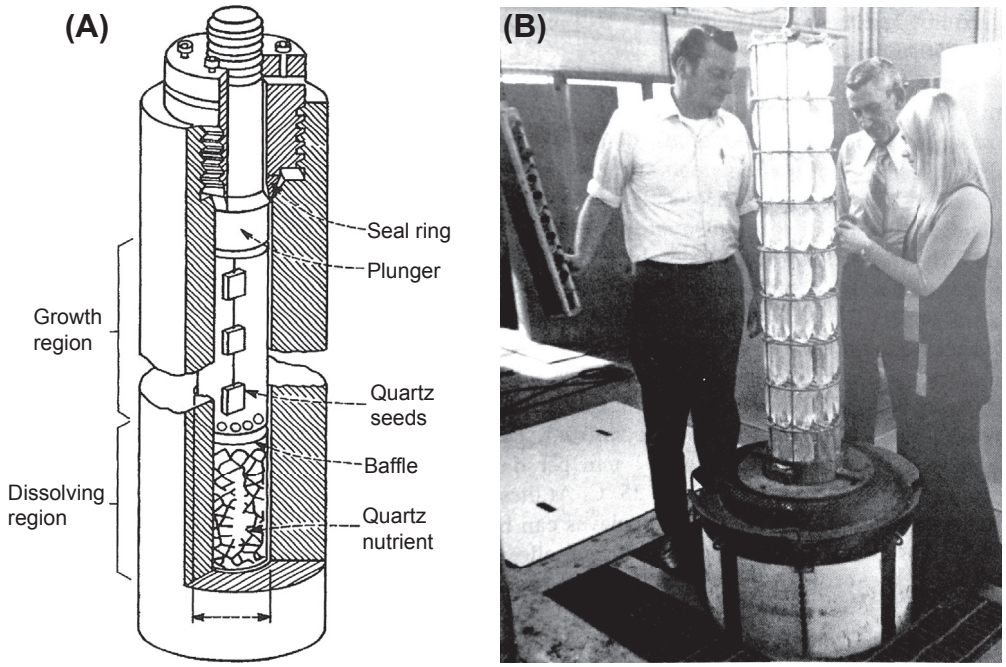


FIGURE 1.13 (A) A schematic diagram of a hydrothermal quartz autoclave. (B) The commercial harvesting of quartz crystals at AT&T's factory in Massachusetts [30].

to a mechanical stress. The hydrothermal method has also been used to successfully grow a variety of crystals of many different classes of compounds from simple binary compounds, such as ZnO, ZnS and GaN, to more complex compounds, such as the phosphates ($AlPO_4$, and $KTiOPO_4$), calcite, hydroxyapatite, zeolites, silicates, metal borates, vanadates, tungstates, and rare earth garnets (e.g., YIG and YAG).

Geologists trying to understand how crystals grew in nature (in an aqueous media under high pressures and temperatures) were the first to carry out hydrothermal phase equilibria studies. Experiments have been traced back to the twelfth century. Their main interest was in phase relationships rather than the growth of large crystals. One of the first published papers on hydrothermal crystallization was by Karl Emil von Schafhäutl in 1845. He prepared microcrystals of quartz. A short time later, in 1848, Robert Bunsen prepared some of the alkaline earth carbonates. The first attempt to grow large crystals hydrothermally was by Henri De Sénarmont in 1851 [231]. He introduced the use of seed plates. It was one of his many studies on the hydrothermal crystallization of minerals. Fifty years later, Giorgi Spezia (1905) published one of the seminal papers on the seeded growth of α-quartz [232]. These early research efforts in Europe eventually formed the basis of the commercial quartz crystal industry.

The modern synthetic quartz crystal growth industry arose during World War II. Supplies of natural Brazilian quartz were not getting to the U.S. due to German submarine attacks on allied shipping. Ironically, one of the applications for quartz crystals was for submarine detection. Piezoelectric materials are needed in single crystal form to take advantage of their anisotropic properties. In oscillators for example, the frequency depends on crystal orientation and devices require precisely oriented parts. Quartz is also used for watches and clocks for precision time management and in signal processing applications.

Major developments in the hydrothermal growth technology were centered on the growth of quartz crystals at the Bell Telephone Laboratories and the Western Electric Company during the 1950s. Walker and Buehler [233] developed a hydrothermal growth method capable of producing very large crystals. They used a welded steel autoclave that was capable of temperatures of 450 $^{\circ}$C and pressures up to 3000 atm. Improvements in autoclave designs were based on some early high pressure studies by Bridgman. Over the following years, improvements were made by the Western Electric Company that led to its successful commercialization. It was in large part due to the efforts of Laudise and Sullivan [234]. Systematic kinetic studies by Laudise [235] led to significant improvements in crystal growth rates. An effort to improve the resonance of quartz oscillators was undertaken by Bell Laboratory scientists and other researchers. Lithium and nitrite ions added to the growth solutions led to improved mechanical Q values [30].

Important hydrothermal growth parameters include (1) operating temperature, (2) temperature gradient, (3) pressure, (4) percent fill, (5) impurity and mineralizer concentration, and (6) seed orientation and surface area. Using quartz as an example, the autoclave is placed in a two-zone furnace with the upper section, containing the oriented seed plates, being cooler than the bottom in which the nutrient, SiO_2, is placed.

A perforated metal disc serves as a baffle between the two zones. The vessel is then filled with solvent to a desired level. The nutrient dissolves in the hot region of the furnace and the upper becomes supersaturated and deposits crystal layers on the seed plates. Convective currents generated by the temperature gradients help move saturated solutions to the seed chamber. For commercial quartz production (see Figure 1.13(B) above), a typical set of conditions might include a hot zone temperature of 400 °C, a seed zone temperature of 360 °C, a fill factor of 80% and a solution containing 1.0 M NaOH, a baffle opening of 5% and a pressure of 21 kpsi (144 MPa) [230].

An important more recent application of hydrothermal growth has been in the preparation of ZnO crystals. ZnO is a transparent, wide bandgap semiconductor (n- or p-type when doped) with a range of useful properties. It is piezoelectric, ferroelectric, exhibits room temperature ferromagnetism, has a large magneto-optic response, etc. It is useful for chemical sensors, catalysis and optoelectronic applications. When doped, its conductivity can vary from insulating to metallic. Bulk crystals, thin films and various nanostructures can be produced using this method. Nanostructures, in the form of platelets, rods, columns, and complex bilayer (column-to-rod), have recently been prepared using hydrothermal methods [236–239].

1.3.2.8 Electrochemical Crystal Growth

Another useful solution growth technology is electrodeposition. It can be carried out in both aqueous and molten salts solutions, and both bulk and thin film single crystals can be grown in this way. The process involves introducing an anode and cathode into either type of solution (of appropriate composition) and applying a suitable voltage across the cell. The driving force for crystallization is the passage of current between the electrodes. The electrode can be a single crystal substrate (wafer or seed), a wire or a more complex structure. Electrodeposition has recently become an attractive method for use in preparing nano-, bio- and micro-structures. It can be used to make functional materials with the aid of three-dimensional masks and scaled up from the deposition of a few atoms to thick deposits.

The first use of electrolysis in chemical processing is attributed to the famous English chemist Sir Humphry Davy (1778–1829). Davy, who was responsible for the discovery of several alkali and alkaline earth metals, separated K from KOH in 1807, the first metal isolated by electrolysis. It has been used since then for the synthesis of a variety of materials. The Hall process, developed in the 1930s for separating Al metal from bauxite (dissolved in a molten salt), is one of its most important industrial applications of electrodeposition. It has also been used to produce many refractory compounds such as borides, phosphides, silicides and carbides. The application of this technology to crystal growth had a late start, surprising since Kunnmann [240] observed that "materials electrochemically precipitated from fused melts can almost always be obtained in the form of reasonably large crystal when sufficiently low current densities are employed."

The potential advantages of electrocrystallization for crystal growth include (1) growth can be accurately controlled solely by electrochemical parameters (current

density and electric potential), (2) the process is isothermal (thermal gradients and temperature ramps are unnecessary, (3) its insensitivity to temperature fluctuations, (4) low growth temperatures minimize thermal decomposition and stresses, as well as vapor losses, (5) purification can be achieved electrochemically and (6) growth features can be studied quantitatively by varying electrochemical parameters [241]. On the other hand, the material to be grown and substrate have to be electrically conducting, and the growth rates are typically slow due to the generally low solute concentrations in the solution. An interesting hybrid method, developed by DeMattei et al. [242], combined molten salt electrodeposition with the Czochralski pulling technique. They demonstrated the method by growing long [110] oriented crystals of sodium tungsten bronze from [110] oriented seed crystals.

Silicon was first electrodeposited in 1854 by Claire-Deville [243]. He used a $NaAlCl_4$–Si molten salt solution. This work was followed a decade later by Ullik [244] who used a K_2SiF_6–KF flux. Cohen and Huggins [245], using a similar flux, were the first to produce coherent epitaxial layers of Si on Si substrates. Metal substrates yielded polycrystalline films. Other semiconductors electrodeposited from molten salt fluxes include the III–V compounds GaP, GaAs and InP. A review of molten salt electrochemical crystal growth was given by Feigelson in 1980 [246].

In the 1990s, an active research area developed around the growth of heteroepitaxial thin films of chalcogenide semiconductors using the low temperature aqueous solution electrodeposition method [247]. Large-scale solar cells were made from electrodeposited polycrystalline CdTe films [248]. Epitaxial films of CdTe can be electroplated from solutions containing cadmium sulfate and TeO_2 onto an InP substrate [249]. In addition, epitaxial films of PbS [250], CdS [251], ZnSe [252] and other related compounds have been electrodeposited. Schlesinger et al. [253] presented a comprehensive review on the subject of semiconductor electrodeposition.

As mentioned above, ZnO is an important and versatile material of great interest to the research and industrial communities. The electrodeposition of ZnO was first demonstrated by Izaki et al. [254] and later Peulon et al. [255]. The growth of oriented rods and flat, disc-shaped crystals were described in Refs. [256–258]. Xu et al. [259] electrodeposited well-defined nano- and micro-structures onto to indium-doped tin oxide substrates using low molecular weight salts in the solutions to control crystal shape. They produced hexagonally shaped tapered ZnO rods and platelets and rhombohedral rods by using amine and other inorganic ions in their solutions.

1.3.3 Vapor Growth

1.3.3.1 Introduction

Vapor phase crystal growth methods have been used extensively for the preparation of both bulk crystals and single crystal thin films. The latter, called vapor-phase epitaxy, are usually deposited on single crystal substrates and have found their greatest utility in the preparation of films and patterned nanostructures for electronic and electro-optic devices.

Important film deposition techniques include OMVPE, MBE, sputtering, etc. In most of these techniques the crystalline lattice of the film needs to be tailored to the substrate upon which it is deposited. Their properties depend not only on their composition, but also on lattice matching between the film and substrate and the crystalline defects that might arise from any misorientation. The substrates are typically thin, crystallographically oriented wafers cut from bulk crystals usually grown by melt growth techniques. The process can be homoepitaxial (growth on a substrate of the same material) or hetero-epitaxial (growth on a different substrate material). Some common examples being Si integrated circuits, GaAs LED devices and more recently GaN on sapphire for lighting applications. Artificial epitaxy, to be discussed later, involves the creation of a geometric pattern (containing some orientational relationship with the film lattice) on a substrate by etching or deposition. The base substrate can be an amorphous material like glass.

Vapor growth technology does not have as long a history as other crystal growth methods. Most research and development work began mainly from 1960 onward. However, it has been traced back a bit further to the German chemist Robert Bunsen (1811–1899) [260]. In 1852, Bunsen observed that Fe_2O_3 crystals formed together with HCl in volcanic gases through a reaction between ferric chloride and water vapor, i.e., a chemical vapor deposition (CVD) process [261]. Not long afterward, in 1861, French chemist Henri Claire-Deville (1818–1881) became the first person to put a CVD process to use preparing artificial oxide minerals of magnesium, titanium and tin [262].

The first commercial CVD process was inaugurated in 1880 for the fabrication of filaments for the new incandescent lamp industry [263]. In 1914, F.C. Brown, studied the crystal habits of Se crystals deposited by sublimation of its vapor in a closed tube under either vacuum or atmospheric pressure [264]. He held the Se at 270 °C for up to a week and the crystals formed along the tube where the temperature was lower. The largest crystals always formed at the higher condensation temperatures (~210 °C). During the 1920s, Fritz Koref and immediately afterward Anton Eduard Van Arkel used WCl_4 to deposit W on single crystal tungsten wires. Koref [265] used a hydrogen reduction method to dissociate WCl_4 gas near the wire, which was heated to between 110 and 1000 °C. This led to a W deposit containing oriented grains. In Van Arkel's process [266], H_2 was not needed as the process was operated at much higher temperatures (1600–1700 °C). In 1921, research began on growing metal crystals by sublimation. Gross [267] and Gross and Volmer [268] grew leaflet crystals of Zn and Cd by directing vapors onto cool glass plates. This work led Volmer to his adsorption-layer theory discussed earlier. In 1932 Straumanis [269] grew Mg, Zn, and Cd crystals by a similar technique. The metals were held at temperatures somewhat below their melting points.

Three basic techniques have been used to grow crystals from the vapor phase: (1) direct sublimation or evaporation of material followed by condensation, (2) chemical transport reaction and (3) chemical vapor deposition. Chemical thermodynamics (shifts in vapor–solid equilibrium) and mass transport are some basic differentiating features between these methods. The process relies on mass transport of species from the source through the gas phase to its incorporation onto the crystal surface. The sublimation and

chemical vapor deposition methods have been operated in either an open system, using an inert carrier gas, or in a closed growth chamber containing vacuum or higher gas pressures. The chemical vapor deposition method, most often used in thin film epitaxy, involves the decomposition of molecular species (precursors). The chemical transport reaction method, a reversible process, converts nonvolatile species into volatile ones prior to crystallization in closed systems.

Crystals can usually be grown by vapor growth techniques at lower temperatures than from melts of the same composition. Vapor phase methods are especially useful when a compound is difficult to grow because of a high vapor pressure, dissociation prior to melting, etc., or where a thin film is required. While these methods are used more extensively to grow epitaxial thin films, bulk crystals of a wide variety of elements and compounds (inorganic and organic) have also been prepared in useful sizes. Seeds are often used but many studies have involved heterogeneous nucleation on the walls of an ampoule. Vapor grown bulk crystals have been particularly useful for the preparation of small crystals for physical property studies, and in a few cases larger crystals, such as SiC and CdS, have found commercial markets. Crystals prepared by vapor techniques include halides, chalcogenides, oxides, pnictides and organic compounds. Growth rates vary for different materials systems and process details, but generally tend to be slower than melt growth methods. Comprehensive reviews of vapor growth theory and methods have been given by Kaldis [270], Faktor and Garrett [271], and Wilke [272].

1.3.3.2 Bulk Growth

1.3.3.2.1 Physical Vapor Transport

A volatile compound that congruently sublimes (or evaporates from the liquid state) can form crystals when it condenses in a cooler region of a furnace. In its simplest form, a closed glass ampoule containing the source at one end is placed in a temperature gradient. The source sublimes at a selected temperature and condenses at the cooler end either as self-nucleated crystallites or on a seed crystal. When no seed is used, many nuclei usually form and some may outgrow the others due their temperature of deposition or their orientation with respect to the heat flow in the system. In 1954, Pizzarello [273] made an important modification to the method that helped improve crystal size and quality. It involved translating the ampoule in the furnace gradient and has some similarities to zone melting with the source and crystal separated by the gas phase rather than the melt. The amount of vaporization at the source end is balanced by amount deposited on the crystal [274]. This “zonal sublimation method” has been used to grow doped crystals of Cd and Zn chalcogenides (see Ref. [275]). Both vertical and horizontal methods have evolved. By controlling the nutrient flux toward the growth interface, seeded growth is possible. This was demonstrated both by Fochs in 1960 with CdS [276] and by Prior in 1961 with PbSe [277].

Some refractory materials such as SiC and ZnO require high temperatures to achieve useful vapor pressures. SiC and ZnO are important wide bandgap, high-temperature/high-voltage semiconductors. Large crystals are sought-after for the

fabrication of substrates. Since melt growth would require pressures of 100,000 atm and 3200 °C, other growth methods were sought. Various novel growth chambers and heating methods were developed for the vapor growth of these crystals. One of the earliest was by Frisch [278] in 1935. He grew ZnO needles by sublimation of a ZnO pellet heated to very high temperatures. In 1955, Jan Anthony Lely [279] developed a sublimation process for growing SiC single crystals that forms the basis of all commercial processes today. In his method, silicon carbide was placed in a graphite crucible and heated to 2500 °C in an argon atmosphere. Large hexagonal platelets of 6H–SiC formed. The crystals were of different sizes up to $2 \times 2\ \text{cm}^2$ and were of very high quality. In 1978, Tairov and Tsverkov [280] modified the process to include seeding. They placed the source at the bottom of the reactor and the seed at the top. Growth rates of 0.5–1 mm/h were achieved. Further modifications have been made since then, and now crystals greater than 50 cm in diameter can be produced. The formation of defects in these crystals, in particular micro-pipes has been a challenging problem [281]. Much effort has gone into their reduction or elimination.

Large crystals of organic compounds such as urea have also been grown by sublimation techniques [282]. However, the formation of gaseous byproducts such as ammonia during growth were problematic. To remove these unwanted species, which either slowed down or stopped the growth process, a vacuum pumped effusion hole was incorporated into the ampoule. Large cm-size high-optical-quality boules of urea were grown on [001] seeds at rates of 2.5 mm/day (by comparison growth from methanol solutions was 0.3 mm/day). The reactor used for the PVT growth of large urea crystals is shown in Figure 1.14.

1.3.3.2.2 Chemical Transport Reaction

When a material is nonvolatile under convenient processing conditions, it can often be chemically converted into a volatile species. Crystal growth can then proceed in a reversible process. This method is called the chemical transport reaction method (CRT). For example, ZnSe crystals can be grown in a sealed ampoule in the presence of a small amount of I_2 vapor (the transport medium). The ZnSe charge at the hotter end will react to form ZnI_2 (v) + Se (v). These gaseous species will then be transported to the cooler end reforming ZnSe on the growing crystal and thereby releasing I_2. The freed iodine can then react with more ZnSe source material and return to the growth zone. In this case, for deposition at the cold end, the required conditions are that the enthalpy (ΔH) and entropy (ΔS) are <0. For materials where ΔH and ΔS are >0 deposition takes place in the hot zone.

The chemical transport method is based upon the pioneering work of Van Arkel and de Boer [283]. In 1925, they prepared the refractory metals titanium, hafnium and thorium using this iodine transport technique with deposition taking place on a heated wire. By 1963, the field had expanded to encompass many different materials using a variety of transporting agents. Harold Schafer's book entitled *Chemical Transport*

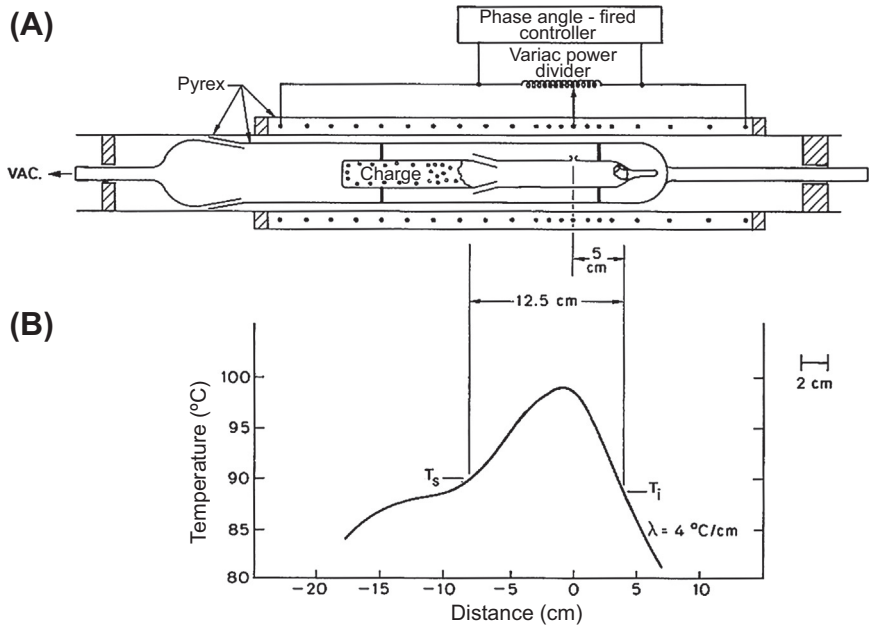


FIGURE 1.14 An example of a physical vapor transport bulk crystal growth apparatus. This growth system was used for growing urea crystals [282].

Reactions [284] has been an invaluable reference source for workers in the field since 1964.

Some important process requirements include (1) a chemical reaction that produces only one stable phase, (2) a free energy close to zero to facilitate reversibility, (3) a non-zero ΔH^o and (4) the ability to control nucleation and the growth kinetics via crystallization zone temperature, temperature gradients, etc. Dopants have been added by incorporating volatile species of the desired element(s) into the growth ampoule. The choice of transporting agent is based on the thermodynamic propensity for the volatile species to form and dissociate in a useful temperature range. Sometimes additional species such as CO are added to the ampoule to facilitate the desired reaction. Sagal showed in 1966 that the growth of Y_2O_3 crystals solely by halogen transport was not favored due to relatively high values of ΔG^o (near 60 kcal/mol for the Cl_2 gas and higher for Br_2 and I_2) [285]. However, by adding CO to Br the ΔG^o value could be shifted closer to zero. The reaction therefore would be:

$$Y_2O_3(s) + 3CO(g) + 3Br_2(g) \rightleftharpoons 2YBr_3(g) + 3CO_2(g) \tag{1.5}$$

During the 1950s and 1960s, considerable research work was in progress using the CTR method. Metals such as iron, cobalt, copper and nickel crystals were produced, as well as classical semiconductors such as silicon, gallium arsenide and gallium phosphide. In addition, various oxide crystals such as alumina, beryllia and silica were grown.

1.3.3.2.3 Chemical Vapor Deposition

The same basic chemical transport process can be use in open systems; however, the process is not reversible. This method is typically known as chemical vapor deposition (CVD). Its most important application by far is the growth of epitaxial thin films to be discussed later. Bulk crystals or thin oriented films are grown by reacting and/or decomposing one or more volatile precursors in the vapor state and depositing them onto the crystal or substrate. These sources can be in many forms; gases, liquids, solutions and aerosols. Any unwanted reaction byproducts can exit the system in the gas stream. Very pure crystals can be produced by this method depending on the type of precursor used. Perhaps the earliest example of this technique was reported by Lorenz in 1891 [286]. He reacted Cd vapor with H_2S gas to form fairly large crystals of CdS. In 1947, Frerichs [287] modified the technique by using a slow stream of H_2 gas to drive the H_2S over Cd metal that was heated to 800–1000 °C. His open tube system produced crystals up to 2 cm^2. An extensive discussion of bulk crystal growth from the vapor phase is given by Schönherr [288]. He provides many useful and practical details including the various methods used, ampoule designs and furnace systems, etc. Bulk growths can be grown in vertical or horizontal configurations or any angle inbetween. Translating the growth chamber or ampoule in a temperature gradient is an often-used procedure. The reactors can be operated at pressures ranging from atmospheric to ultra-high vacuum. Materials produced by the CVD method include refractory metals (such as tungsten), semiconductors (such as silicon and III-V compounds), oxides (such as SiO_2), silicon carbide, nitride and oxynitride, and various carbon structures, including diamond as discussed later. Since the late 1990s, it has found use in the preparation of nanocrystals, one important example being carbon nanotubes and fibers. The nanotubes can be produced by a number of methods including the catalytic decomposition method [289], a CVD technique using metal catalysts together with hydrocarbon precursors. Depending on the details of the process, i.e., the metal catalysts used, etc., aligned single- or multi-walled nanotubes can be produced.

1.3.3.3 Vapor Phase Epitaxy

1.3.3.3.1 Organometallic Vapor Phase Epitaxy

Organometallic vapor phase epitaxy (OMVPE aka MOCVD) is a subset of the more general Chemical Vapor Deposition (CVD) method. It uses at least one organometallic precursor (OM) but may also be combined with other types of volatile species to produce films of many different II-VI and III-V semiconductor compounds and their solid solutions. Like other methods, there are lots of variations in technique. One of the earliest recorded descriptions of the OMVPE process was in Scott et al.'s little known 1957 United Kingdom patent [290]. In it, he describes the deposition of InSb in a cold wall reactor by the pyrolysis of a Group III alkyl (i.e., triethylindium) and a Group V hydride (i.e., stibine-SbH_3). The second, in a 1965 U.S. patent, described the pyrolysis of a Group III alkyl (i.e., triethylindium or trimethylgallium) and a Group V reactant such as arsine to produce a III–V semiconductor [291]. However, the first published work in the

scientific literature was in 1969 in a paper by Harold Manasevit and W. Simpson [292]. They grew single crystal Ga-group V compounds on insulating, GaAs, GaP or Ge single crystal substrates. Either trimethylgallium or triethylgalllium in the presence of arsine, phosphine and arsine-phosphine or arsine-stibine gas phases was used in these experiments. In 2004, Manasevit, now considered one of the founders of OMVPE technology, published his recollections on how the OMVPE field developed [293]. A schematic drawing of his apparatus is shown in Figure 1.15. Along with his colleagues at the Autonetics Division of North American Rockwell, Manasevit published numerous other papers on this topic. In 1975, Seki et al. [294] produced the first important device quality (i.e., very high mobility) GaAs layers. This advancement was due to the enhanced purity and crystalline perfection of the films. Other major technological advancements followed soon afterward.

One of the important virtues of the OMVPE method is that it is can be used to grow epitaxial semiconductor alloy films. In 1977, Dupuis and Dapkus [295] grew low oxygen and carbon films of AlGaAs by the OMVPE method. This material had excellent minority carrier lifetimes making them useful for light-emitting diode devices. In 1978, Gerald Stringfellow, from the Hewlet-Packard Laboratories, both proposed [296] and demonstrated [297] that with OMVPE one could grow very bright LEDs from AlInP and AlGaInP epitaxial films.

In the 1960s, Isamu Akasaki's group at Nagoya University started working on GaN-based LED's devices. In 1989, his work culminated in the invention of a bright gallium nitride p-n junction by using the low temperature OMVPE method with an AlN buffer layer on sapphire [298]. A major step was in creating p-type GaN using magnesium as the dopant and n-type with silicon. In 1994, Nakamura et al. [299] grew the first very bright InGaN/AlGaN double-heterostructure blue-light-emitting diodes also on

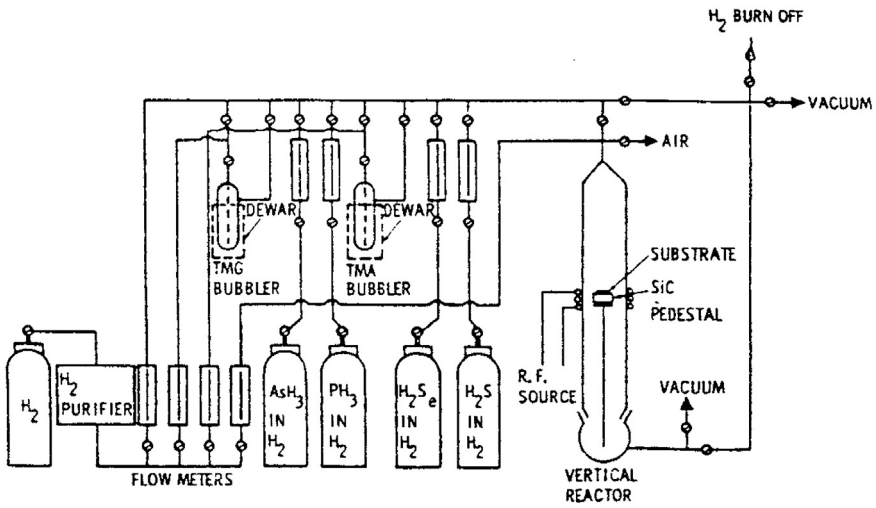


FIGURE 1.15 Schematic drawing of Manasevit's MOCVD deposition system [293].

sapphire substrates using a two-flow OMVPE method. The work of Nakamura's group at the Nichea Chemical Company, as summarized in Ref. [300], was a pivotal step in the development of the now multibillion dollar industry centered on the fabrication of highly efficient GaN-based/sapphire epitaxial films for optoelectronic devices. OMVPE development since then has been continuous with emphasis not only on improving the quality and properties of the epilayers, but also the quality and cost of the sapphire substrates. In an example of a more recent work (2002), Liu et al. [301] grew GaN single crystal epitaxial layers on sapphire in a three-step process using the low-pressure OMVPE method together with an AlN buffer layer and via atomic layer epitaxy (to be discussed later).

Stringfellow reviewed the development and status of the OMVPE method several times (e.g., Refs [297,302]). In the latter, he commented "One reason that OMVPE is so widely used today is that it is the most versatile technique for the growth of materials and structures for a wide range of devices."

1.3.3.3.2 Molecular Beam Epitaxy (MBE)

Molecular beam epitaxy (MBE) is a process in which a thin single crystal layer is deposited on a single crystal substrate using atomic or molecular beams generated in Knudsen cells contained in an ultra-high vacuum chamber. The source beams can be created in a number of ways, including (1) melting and evaporation of solids or liquids contained in crucibles (2) solid sublimation from a crucible, (3) ion beam bombardment, and (4) cracking various chemical species, etc. Its greatest use is for making multilayer semiconductor device structures. Details of the MBE method, as well as other bulk and thin film growth techniques used to prepare compound semiconductors, are given in Ref. [303].

One of the earliest published studies on the use of the MBE method for single crystal film growth was that of Joyce and Bradley [304]. In the mid-1960s they grew homoepitaxial layers of Si from SiH_4. The growth rates were very low comparative to other Si film methods and therefore not competitive in a market that needed 10 um-thick films. A few years later, J. Davey and T. Pankey [305] from the Naval Research Laboratory, and J. Arthur [306] and A. Cho et al. [307] from Bell Laboratories expanded the MBE method for the deposition of GaAs. Arthur focused on surface kinetic studies, Davey and Pankey grew large-area GaAs epitaxial films on GaAs and Ge substrates using the three temperature technique, while Cho focused on device applications. The MBE technique is a powerful method both for film deposition and in situ analysis. It has yielded, in addition to device structures, a wealth of data on the surface atomistic phenomena such as surface reconstruction. It has also been applied to other semiconductor material systems such as the nitrides and has facilitated the construction of novel structures such as periodically poled GaAs for IR nonlinear applications and quantum dots. Today, it is a very important research tool and is used extensively in commercial optoelectronic device processing. A historical review of the MBE method was given by Joyce and Joyce in Ref. [308].

1.3.3.3.3 Sputtering

In 1852, Sir William Robert Grove (1811–1896), a noted Welsh judge, physical scientist and inventor of the fuel cell, was the first to discover the deposition process now known as sputtering [309]. He was able to deposit material from the tip of a wire in a chamber at a pressure of 0.5 Torr onto a polished silver surface when the latter was the positive electrode in an electrical circuit. An interesting factoid is that the first commercial application of the sputtering method may be attributable to Thomas Edison who early in the development of his wax phonograph cylinders using a sputtering methodology for plating them.

The sputtering method, as used to prepare thin films, became popular and of commercial importance from the mid-1960s onward. It has the advantage of not requiring high temperatures to deposit materials, even very refractory ones. The films have compositions similar to the target material and large areas can be deposited. While this physical vapor deposition method is more often used to deposit polycrystalline and amorphous films, single crystal films have been produced by carefully controlling the processing parameters. The method is used for fabricating integrated circuits, antireflection coatings, solar cells and optical waveguides, etc. Typical materials sputtered include metals, semiconductors, oxides, and nitrides, etc.

The simplest sputtering process involves just a temperature-controlled cathode and anode, a source of energetic particles, i.e., ions or atoms, and a vacuum chamber. A DC potential of several thousand volts is usually maintained across the electrodes. Radio frequency sputtering, where the sign of the electrodes is varied at a high rate, has also been found beneficial. The material to be deposited is ejected from the target (at the cathode) by bombarding it with ions or atoms, and the ejected material is transported in the plasma formed to the substrate (at the anode). In addition to the ions released from the target, electrons are also produced and they play an important role in maintaining the plasma. However at the same time they can cause excessive heating of the substrate. The transport mechanism within the gap between the two electrodes is complicated and depends to a great extent on the background gas pressure. For efficient ejection, the sputtering gas should have a similar atomic weight as the target elements. Nonreactive gases such as argon, krypton and neon are often used to eject atoms from the target, but reactive sputtering, using oxygen or nitrogen gas, has been employed to deposit oxide and nitride films (e.g., ZnO and TaN). In the latter process a chemical reaction takes place between the gas and the sputtered ions near the cathode before being transported to the substrate. Higher substrate temperatures encourage the deposition of single crystal films. Williams has given an extensive overview of the sputtering field and sputtered ion emission [310].

Conventional sputtering has some disadvantages including low deposition rates, low ionization efficiencies and substrate heating. One major improvement to this technology was the introduction of magnetron sputtering [311]. In this process a magnetic field is incorporated into the sputtering apparatus with the magnetic field positioned parallel to the target and confining the secondary electron movement close to the target. This

maximizes the probability of electron–atom interactions, increasing ionization efficiency. The result is higher sputtering and deposition rates. It also permits the use of lower operating pressures and voltages. The magnetron was originally conceived by P.M. Penning in 1936 [312]. In 1980, Naoe et al. [313] were the first to use it in a sputtering application. Over the years magnetron sputtering configurations have been modified, and these developments have led to improved film quality and device performance [311].

1.3.3.3.4 Atomic Layer Deposition

In 1977, Dr Tuomo Suntola from Helsinki University in Finland patented a novel technique to prepare highly oriented compound thin films [314]. The method was called atomic layer epitaxy (ALE). More recently, the nomenclature atomic layer deposition (ALD) has been favored. The ALD technique provides precise control of the film thickness and composition and with the proper substrate composition, orientation and temperature, can produce single crystal thin films. It involves the periodic (alternating) pulse deposition of a film's components in a vacuum chamber. In between pulses there is an equilibration period during which the excess components can desorb from the surface and exit the growth chamber. This leaves just one atomic layer on the substrate surface. The next atomic species is then deposited, and a controlled chemical reaction at the surface between these two layers creates the desired film composition or composite structure. By way of example, to produce an epitaxial ZnS film by this technique, a single Zn atomic layer is first deposited on the substrate surface. This layer is then exposed to S_2(g) or H_2S(g), either of which react with the Zn layer to form the compound ZnS. Following equilibration, another Zn layer is deposited and then reacted again with the sulfur-containing gas. The thickness is determined by the number of cycles employed.

Historically, the idea for a sequential layering film deposition process was first mentioned in the 1952 thesis of Professor V. B. Aleskovskii as molecular layering. And years later, (during the 1970s) his group in Russia worked on the developing this concept further [315]. The efficacy and implementation of the method into a commercially viable process derived from the work of Suntola's group during the years prior to their patent application. For a definitive review of all aspects of the methodology, see Suntola [316].

The ALD technique can produce atomically flat films with almost perfect stoichiometry and surface conformity through the self-limiting reaction mechanism. It can be used with many of the chemical vapor deposition methods described above that normally deposit the requisite phases simultaneously. It can be used to produce layered films with abrupt interfaces (e.g., TiO_2/Al_2O_3 films [317]), and M. Ritala and M. Leskela [318] reviewed the method's features and its potential role in nanotechnology. One of the drawbacks of the ALD method is the slow deposition rates. This has been somewhat overcome by increasing the substrate areas during deposition. The use of bias sputtering has given the best stoichiometry to date. Besides oxides and chalcogenides, as mentioned above, the method has also been used for the preparation of various semiconductors, nitrides, and metal films.

1.3.3.4 The VLS Method

Another interesting and useful growth method is the vapor-liquid-solid (VLS) technique. This hybrid method combines chemical vapor deposition with solution growth. Deposition takes place at localized positions on a single crystal substrate to yield a nanostructure, particularly whiskers, rods and nanowires. It starts with a single crystal substrate patterned with an array of small dots made from a solid metal solvent phase (the "catalyst"). The patterning can be done using lithography or by converting a solvent film deposited on the substrate surface to droplets. The growth procedure is simple. When the substrate is heated, the solvent phase melts. The liquid phase rapidly supersaturates by adsorption of nutrient species from the gas phase. Growth subsequently takes place at the substrate–liquid interface and not on the bare substrate surface. The solvent region rises up, as a mass is deposited below it, thus propagating the growth feature. The molten zone remains on the fiber tip during growth. As an example, silicon nanowires have been produced from a Au–Si alloy droplet and with a gas phase containing $SiCl_4$ and H_2 (see Figure 1.16 below). The VLS method has been used in conjunction with CVD, MBE, laser ablation and carbothermic reduction.

The VLS method was first described in 1964 in the pioneering work of Bell Laboratories scientists R. Wagner and W. Ellis [319] (see Figure 1.16). The VLS

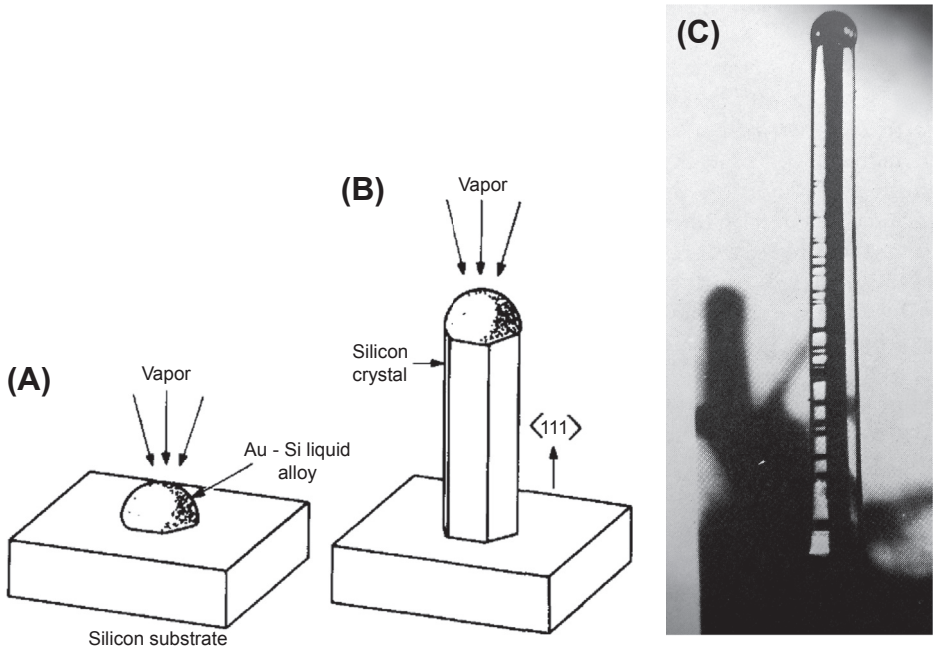


FIGURE 1.16 Original schematic diagram of the VLS process for Si whisker growth on a silicon substrate, (A) Au–Si alloy catalyst droplet on substrate surface before growth. (B) A growing whisker. A photograph of an actual Si whisker (0.5 mm diameter) grown on a {111} Si substrate is shown in (C). It has 12 side facets alternating between the {211} and {110} [318].

mechanism they proposed explained the growth of silicon whiskers on silicon substrates in the absence of the axial screw dislocation growth model described by Frank [22]. Some advantages to the VLS growth process include a lower reaction energy than with regular vapor growth techniques, the structures only grow where the solvent is located, and anisotropic-shaped columns can be produced. Important process parameters include (1) the wettability of the liquid droplet and its reactivity with the substrate, (2) the substrate orientation and surface roughness, (3) the processing temperature, etc. It is also very important and obvious that the solvent phase has to have a low equilibrium vapor pressure.

Some of the popular semiconductor materials grown by VLS include Si, Ge, GaAs, GaN, SiC and ZnO. Gold is most often used as the solvent phase, but other materials such as Ni, Pt have been successfully used. The substrates may be of the same or similar composition as the growth pillars (homoepitaxy) or on different material (heteroepitaxy). One example of the latter is the growth of densely aligned GaN wires grown on sapphire, $LiAlO_2$ or MgO substrates [320]. Schmidt et al. discussed various aspects of the growth of silicon nanowires and their electrical properties, including use of the VLS method [321].

The VLS method has been easily adapted to the growth of nanostructures. This has become a very important active area of research and should lead to exciting new commercial applications in the foreseeable future. One of the leading groups in this area is Lars Samuelson's group at Lund University. They have reviewed the fundamental mechanisms involved in the VLS processing of nanowires and the prognosis for further development of this technology [322]. Many new and unique structures have been created using the VLS process. One example being a decade old study that showed that certain material systems can phase separate into cored nanofibers. In a one-step VLS process, Choi et al. [323] grew GaN cored nanowires with a thick AlGaN skin. Other techniques have also been used to coat the nanowires. An excellent review of the VLS method was given by Choi [324].

1.3.3.5 Artificial Epitaxy (Graphoepitaxy)

Up to now, we have discussed the epitaxial growth of single crystal thin films on substrates made from materials of related composition and/or structure. The usefulness of high-quality single crystal or highly oriented films is well-known and crystalline perfection of the film depends on various attributes of the substrate and its quality. However, it would be highly desirable to be able to grow oriented films on inexpensive amorphous substrates or on substrate layers important to device function. In the early 1970s, Prof. N.N. Sheftal from the Russian Institute of Crystallography first described the concept of growing films on an artificial lattice. The paper, which described the technique as "artificial epitaxy," was translated into English three years later [325]. In 1982, the technique was renamed "graphoepitaxy" [326], and this more catchy term has gained favor in much of the subsequent literature.

Graphoepitaxy involves inscribing a micro-relief pattern onto a flat amorphous substrate surface. The surface patterns consist of only four symmetries, two-, three- four- and

six-fold (i.e., arrays of stripes, triangles, squares and hexagons). The walls of the relief pattern simulate the kinks and ledges associated with a crystal growth surface as described by the Kossel-Stranski model. Rather than atoms or molecules, however, the growth units that attach to these relatively large steps are microcrystallites of nanometer or micron sizes. The pattern chosen is determined by the crystal structure of the material to be deposited. For example, the three-fold relief would be used for the growth of the diamond lattices of Si and Ge. The reliefs can be achieved by a number of techniques including photolithography and etching, etc.

Graphoepitaxy can be accomplished using a variety of gas phase, melt or solution growth techniques, including the VLS method mentioned above. Like other methods described in this chapter many different materials have been prepared by graphoepitaxy, including very large-molecule biological materials [327]. By way of illustration, Figure 1.17 shows two crystals of catalase (an enzyme) that were deposited from an aqueous solution onto a silicon substrate that contained a striated micro-relief [327]. The pattern used here had a 5 μm periodicity and a groove depth of 1–2 μm. The crystals are clearly aligned with the micro-relief. Deposited on an unpatterned substrate, the crystals would have no orientational relationship with one another.

Graphoepitaxy has also been used to prepare oriented single crystal nanowires of semiconductor oxides such as ZnO, SnO_2, In_2O_3. The method has been exhaustively covered in a book by Givargizov [328]. Since this book was written, the technique has become very popular, particularly in the semiconductor field. For example, so-called nano-graphoepitaxy has been used to prepare semiconductors for three-dimensional integration devices [329].

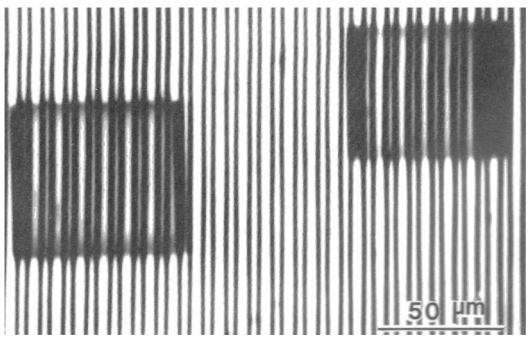


FIGURE 1.17 Two catalase crystals grown from solution onto an oriented Si substrate having an etched micro-relief pattern. The crystals, which are aligned with the pattern, grew by artificial or graphoepitaxy [327].

1.3.4 Synthetic Diamond Crystals

Diamond, while highly sought after as a gemstone, also has a unique combination of properties that make it very useful in industrial applications. Diamond not only has the highest known hardness, it also has a very high thermal conductivity and electron mobility, low thermal expansion coefficient and excellent optical transmission over a

broad spectral range. In addition to its major commercial market, i.e., cutting tools and polishing powders, diamond has a myriad of other uses. Diamond-based devices include high-power electronic devices, high-frequency field effect transistors, LED's, ultraviolet and high-energy particle detectors, substrates and optical windows. The two principal methods used to grow synthetic diamonds are (1) crystallization of bulk crystals from solution at high pressures and temperatures (HPHT) and (2) deposition at low pressures and relatively low temperatures using the chemical vapor deposition (CVD) method. Both were developed during the early 1950s within a few years of each other and are still in use today to manufacture synthetic diamond products.

The earliest known reference to diamond can be found in the Old Testament [330]. It was not until near the end of eighteenth century that it was realized that diamond, while transparent and colorless, was made up solely of carbon atoms like graphite. The discovery came from the French chemist Antoine Lavoiser (1743–1794), who shortly before his death, decomposed a diamond by heating it in oxygen and found CO_2 as the only byproduct. Thereafter, a number of credible researchers tried to synthesize diamond, one of the first being Scottish chemist James Hannay (1855–1931). His attempts in 1879 [331], later questioned, were followed in 1895 by the French Noble Prize-winning chemist Henri Moissan (1852–1907). He tried to synthesize diamond in the laboratory [332] starting with charcoal and iron heated to temperatures as high as 3500 $^{\circ}$C using an electric arc furnace. The heated mixture was then quenched in water to hopefully create the high pressures under which diamond formed in nature. Other researchers who tried to duplicate these studies either failed or had their various claims discredited. Sir Charles Algernon Parsons (1882–1922), the inventor of the steam turbine, spent considerable time and energy over many years trying to duplicate the work of Hannay and Moissan. He also tried to develop his own method to produce diamond. In 1928, as reported by Desch [333], Parsons concluded that synthetic diamonds had not been produced. Kathleen Lonsdale used X-ray diffraction methods to study some of Hannay's "diamonds" held at the British Museum. She concluded in a 1962 paper [334] that they were natural diamonds and doubted that "… neither Hannay, Moisson or Parsons ever, in fact, made diamonds by their respective methods." Percy Bridgman, who as mentioned before won a Nobel Prize for his high pressure work, spent the better part of 50 years (from 1905 to 1955) trying to synthesize diamond. His efforts were apparently unsuccessful as well [330]. In addition to the researchers mentioned above, the nineteenth century was littered with numerous unsuccessful attempts to synthesize diamonds by various means. One particularly engaging and well-researched book on the history and growth of diamond crystals is *The Diamond Makers* by Robert Hazen [335].

In 1941, the General Electric Research Laboratories, in conjunction with the Norton and Carborundum companies, set about to develop a process to synthesize industrial diamonds. The effort was suspended during WW II but started up again in 1951. While GE put together a large staff charged with designing a furnace that could go to both high pressures and high temperatures, it was not until H. Tracy Hall, came

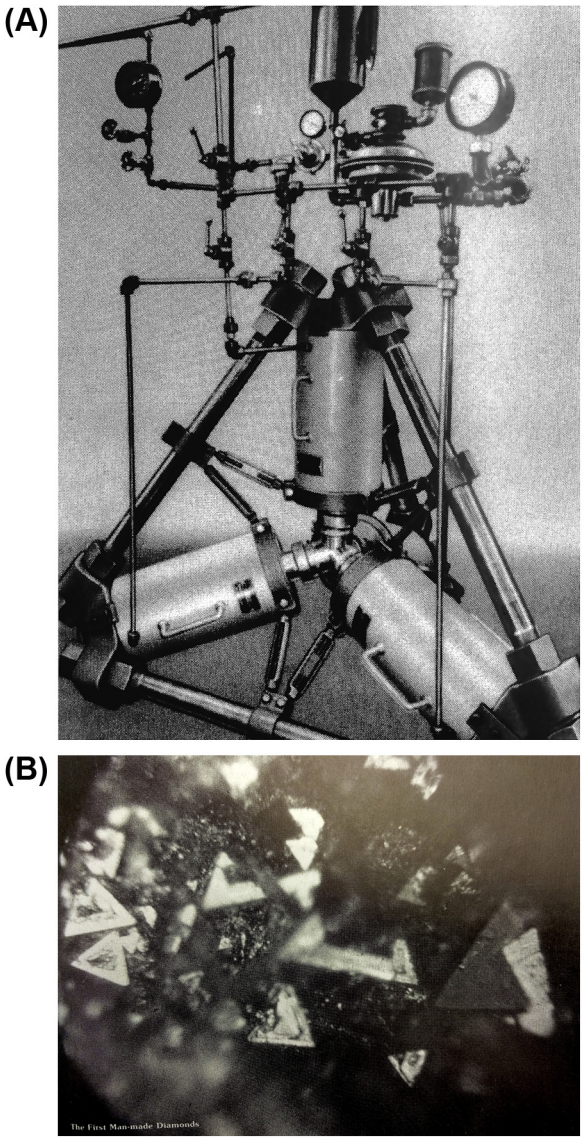


FIGURE 1.18 Early diamond crystal growth (A) Schematic of the belt high-pressure, high-temperature apparatus built at the General Electric Corp., and (B) the first synthetic diamonds produced using this apparatus [330].

up with the "Belt" press (see Figure 1.18(A)) that a breakthrough was imminent. This device exceeded the original specifications of 35,000 atm and 1000 °C to achieve 250,000 atm and 1800 °C [336]. The growth chamber consisted of a graphite tube surrounded by a pyrophyllite container. Inside were placed Ni, Fe, or Co to act as a solvent-catalyst in which the graphite dissolved. The bottom was in contact with a Ta disc. However, even with this capability, diamond was not readily produced. It was

not until the end of 1954, after much experimentation and frustration, that the first small, micron-sized faceted diamond crystals were produced (see Figure 1.18(B)). The addition of FeS to the container did the trick. These were not gem-quality stones but appropriate for abrasive applications. Hall's personal reminiscences, given in Ref. [330], describe how the process was invented and the subsequent events that led to his other important invention—the tetrahedral anvil press shown in Figure 1.19. Oddly, just before the GE success, the Swedish company ASEA, also managed to produce small diamonds in a top-secret project that only went public in 1980 [337].

The first successful preparation of gem-quality diamonds by HPHT was in 1970, again accomplished by GE. The process was similar to that described above with the addition of thin diamond seeds. The first crystals were 5 mm (1 carat) in size and took a week to grow. Longer growth times were required to produce larger crystals. These early crystals were yellow or brown in color due to nitrogen contamination and contained inclusions. By adding nitrogen getters such as Al and Ti, clear colorless crystals could be produced. On the other hand, other dopants have been used to modify the color of synthetic diamonds (e.g., boron gives it a blue color). A variety of colored stones have been produced.

The inherent technological difficulty in preparing diamonds or other materials at high pressures and temperatures and the high cost of equipment led many researchers,

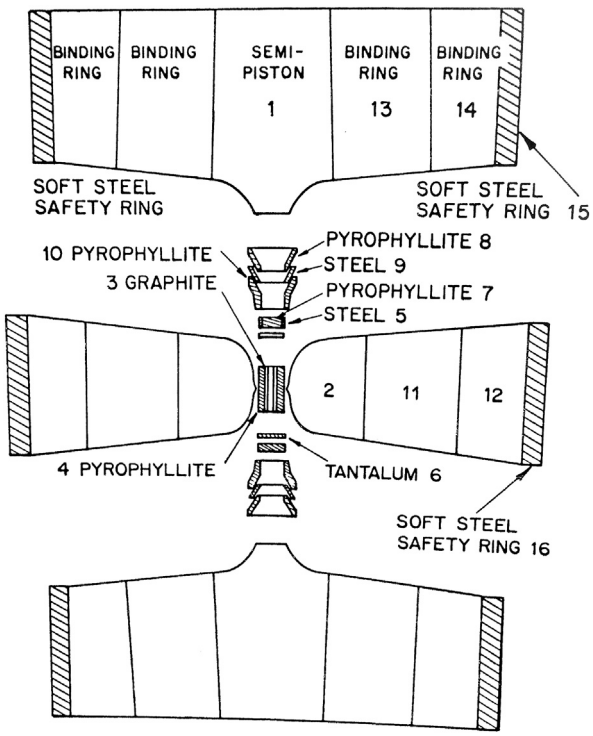


FIGURE 1.19 The original tetrahedral press for producing synthetic diamond [330].

particularly in the U.S. and the Soviet Union, to try to find a low-pressure method. Not many scientists, however, thought this would be possible. Nevertheless, in 1952–1953, William Eversole, at the Union Carbide Corporation, became the first person ever to prepare synthetic diamonds [338] by the chemical vapor deposition technique. It involved slowly depositing carbon atoms on clean diamond particle surfaces (i.e., substrates) in a vacuum at temperatures in the 800 °C range. Carbon monoxide (CO) or methane (CH_4) was used as the source of carbon. Due to the propensity for graphite formation under these conditions, the residual graphite had to be removed after a period of time. This involved removing the crystals from the deposition chamber and cleaning them in an autoclave at 1000 °C and 50 atm of H_2 gas. Numerous such cycles were required. While he accomplished this feat about two years before General Electric and just a few months before ASEA in Sweden [337], his work (described in a patent only) was not published until 1962. In 1968, John Angus [339] independently verified Eversole's work, as did Deryagin and Fedoseev in 1970 [340]. While Angus also deposited diamond on single-crystal diamonds, Deryagin and Fedoseev made epitaxial films on other substrate materials such as Si and metals. From this point onward this very versatile method was aggressively pursued and refined by a number of groups in the USSR (Russia), the U.S. and Japan for the growth of both bulk crystals and homoepitaxial films on diamond substrates and by heteroepitaxial growth on suitably oriented materials like Si. A useful review of diamond growth by the chemical vapor deposition method was written by Garcia et al. [341].

One of the initial problems with the commercialization of CVD bulk diamond was the slow growth rates. In 1969, the former USSR scientists Spitsyn and Dervagin, who had been working on this problem since 1956, were finally able to increase the growth rates reported by Eversole by an order of magnitude. The improvement over their own previous work was due to the use of methane at higher pressures (13–40 Pa), together with an increased deposition temperature (950–1050 °C). Just a year later, a significant breakthrough was made independently by J. Angus (USA) and V. Varnin (USSR) [342]. They found that the use of atomic hydrogen in the growth chamber would remove the graphite co-deposits that form along with diamond due to a large difference in etching rates. A decade later, a group of researchers at Japan's National Institute for Research in Inorganic Materials made a series of important process improvements. They developed the microwave plasma, hot filament, and RF-Plasma CVD methods, new ways to dissociate the carbon-containing gases into reactive species [343–345]. Growth rates up to several μm/hr were achieved. This in turn led to the development of a variety of other processes and process refinements by this and other groups, leading the commercial success of the CVD method for a variety of diamond products including gemstones and coatings for various types of electronic and optical devices.

Two other methods have been employed for growing small-size diamonds. One is the explosive detonation method in which a carbon-based explosive is detonated inside a metal tube containing graphite [346]. The procedure, an HPHT process, produces

nanoscale diamonds. The diamonds are prevented from reconverting to the more stable graphite form due to the quenching effect of a surrounding water-filled chamber. However, they have to be separated from the remaining carbon by dissolution in acid. The second method is the ultrasonic cavitation technique [347]. This more recent process is carried out at room temperature and atmospheric pressure. It involves the application of ultrasonic energy to a suspension of graphite particles in an organic liquid and results in micron-sized diamond crystals.

1.3.5 Solid State Recrystallization

The next to last topic to be covered in this chapter is the growth of sizable single crystals from the solid phase. This technology has a much smaller impact on the crystal growth field than the other methods described above. First, it is mainly limited to metals, and second, there are various processing difficulties associated with controlling nucleation and growth over extended length scales. The mechanisms involved are related to ceramic and powder metallurgy processing, where control of crystallite (grain) size and morphology in polycrystalline structures is a major concern. There are a myriad of important industrial applications for these polycrystalline materials (piezoelectric elements, magnets, etc.) and all aspects from theory to sample preparation are covered in various books on ceramics and powder metallurgy. The use of solid-state methods for crystal growth is covered in the book by R. Laudise [285].

At the heart of solid-state crystal growth (recrystallization) is grain growth. As mentioned before, the method is mainly used with metals such as aluminum, tungsten and iron. The material from which the crystal is grown contains grains of varying sizes and morphologies, plus grain boundaries and dislocations. Single crystals can be formed by controlling the growth of preexisting grains or by nucleating new grains with lower free energies. Wilhelm Ostwald's pioneering work in 1896–1897 explained how crystallites behave at elevated temperatures [348]. He showed that smaller particles adjacent to larger ones would decrease in diameter while the larger grains increased in size (an effect now known as Ostwald ripening). When the more energetic surface atoms on the smaller crystallites redeposit on the larger grains, the total energy of the system decreases. The driving force for grain growth can also be related to orientation differences between grains.

The principal method used to grow large metal crystals involves strain annealing techniques. A suitable polycrystalline sample, a bar, rod, plate, etc., is strained by one of a number of techniques such as rolling, drawing and extrusion. It then may be fabricated into a suitable shape to facilitate growth. The amount of strain induced is usually between 1 and 10% and the amount is critical in controlling the nucleation of strain-free grains. Growth is most often done in a temperature gradient and nucleation control is similar to other growth methods. In some respects, it is related to the Bridgman–Stockbarger method where the sample can be a rod with a tapered

end that is translated through the gradient. The gradients, however, are reversed from melt growth. The polycrystalline charge is heated until the tapered end reaches the recrystallization temperature and strain free grains are created in a localized region at the tip. Further movement propagates the strain-free grain(s) along the axis of the sample consuming the strained grains. If nucleation creates several strain-free grains, the sample can be notched somewhere along the length to permit only one grain to propagate through into the main part of the charge. Seeds can also be used. Suppressing nucleation ahead of the growth front has been a concern, and generally slow translation rates and sharp gradients minimize this problem (e.g., Ref. [348]).

The concept for strain-annealed crystal growth can be traced to Robert Anderson in 1918 [349]. This led, in 1921, to Carpenter and Elam's demonstration of the growth of large aluminum crystals by the strain-anneal method [350]. Over the ensuing years, various other metals have been grown by this method. One technique used to prevent random nucleation ahead of the growth interface during the strain-annealed growth of alpha iron crystals was the incorporation of pulsed heating [351]. Large, 25cm-long oriented single crystal rods and strips could be produced in a few hours using this method.

The solid-state recrystallization method has also been applied to semiconductor fabrication (e.g., Si and Ge). One technique of note is the solid phase epitaxial growth method. In 1968, L. Kulper at IBM, patented a process for the growth of aluminum-doped silicon films by the migration of silicon through an aluminum thin film during a heat treatment process [352]. The solid-state growth process provided doped layers with a maximum amount of aluminum in silicon. Later Mayer et al. [353], from the California Institute of Technology, patented a similar but more general process for doping with other species. In this process, a single crystal substrate is coated with a thin metal film having such properties that it will permit the migration of material through it to form an epitaxial layer without acting as an active dopant itself. Upon this film a dopant layer is deposited followed by an amorphous or polycrystalline layer of the material that will make up the doped epilayer. This sandwich structure is first heated to a temperature that permits the metal layer to dissolve some of the dopant, amorphous film and part of the substrate. After a time, the temperature is raised to allow the transport and epitaxial deposition of the doped layer onto the substrate. For the solid-state epitaxial growth of silicon, for example, the substrate and amorphous or polycrystalline layers would be silicon, the dopant layer might be phosphorous, aluminum, or boron, and the metal film palladium, vanadium, or nickel. A review of this technology is included in the book by Mayer and Lau [354].

Another application for solid-state crystal growth was in the preparation of piezoelectric single crystals, such as lead magnesium niobium-titanate (PMNT) [355]. The principal motivations were its potential cost advantage and enhanced

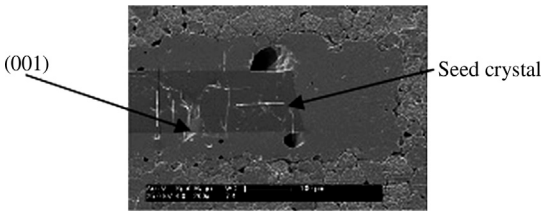


FIGURE 1.20 An example of template growth. Shown is a micrograph of a polished and thermally etched (1080 °C for 30 min) sample of $BiScO_3$–$PbTiO_3$ (BS–PT) that was heat-treated with an embedded barium titanate (BT) seed crystal (5 $mm^2 \times 100$ μm thick). A 5% excess of PbO was used in the mixture to enhance diffusion. An overgrowth of a BS–PT crystal onto the surrounding BT seed can be clearly seen [355].

manufacturing throughput over conventional crystal growth methodologies. Two techniques were explored, conventional grain growth and templated growth in which seed crystals were embedded into a powdered matrix and processed at elevated temperatures. An example of the templated growth of BiScO3-PbTiO3 is shown in Figure 1.20.

1.4 Epilogue

Crystal growth is a field that has had a major impact on modern society. The devices we have come to rely on today were made possible through the contributions of numerous scientists and engineers from a variety of disciplines. These devices are based on single crystals prepared in various ways and in forms and compositions reflecting the application intended. The foundations upon which our understanding of a crystal's structure, thermochemistry, growth mechanisms and methods is based on work from earlier centuries and dramatically expanded on all fronts (theory, growth and characterization) just after World War II. The book by Buckley in 1950 was the first comprehensive treatment (in English) of the prior art and science of crystal growth. Today, there are well over 100 books covering the topic, from surveys of the entire field to various specialized topics. This historical review of the crystal growth field is not comprehensive but was designed to highlight the major achievements. While I have tried to be as inclusive as possible, I apologize in advance if I have left out any major contributors to this field or important theories and growth methods. As a final comment, I must mention that a selection of some pioneering crystal growth papers were collected and reprinted in their original languages by D. Hurle [356].

Acknowledgments

The author is indebted to Alex Chernov, Vince Fratello, Gabe Loiacono, and Jerry Stringfellow for critiquing specific sections of this chapter, and for discussions with Al Ballman, Bill Tiller and others.

References

[1] Pliny the Elder (Gaius Plinius Secundus). Volume 9 (Books 33–35), [H. Rackham, Trans.]. Naturalis Historia circa AD 77–79. Loeb Classical Library/Harvard University Press; 1958. p. 430.

[2] Agricola G. De Re Metallica [L. Hoover and H. Hoover, Trans.]. Dover Publications; 1950.

[3] Biringuccio V. The Pirotechnia [C. Smith and M. Gnudi, Trans.]. Dover Publications; 1990. p. 477.

[4] Doremus RH, Roberts BW, Turnbull D. Proc. int. conf. on crystal growth: growth and perfection of crystals. New York: Wiley; 1958.

[5] Gessner C. De omni rerum fossilium; 1565.

[6] Caesalpinus A. De metallicis libri tres. Rome; 1596.

[7] Sunagawa I. In: Feigelson RS, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 35.

[8] Steno N. De solido intra solidum naturaliter contento; 1669.

[9] Guglielmini D. Philosophical reflections deducted from figures of salts from Doctor Domenico Guglielmini expressed in a speech recited in academy philosophical experimental affirmations. Archdeacon Marsigli; the evening of March 21, 1688. Bologna.

[10] Romé de l'Isle JB. Essay of crystallography or description of geometric figures. L'Academie Electorale des Sciences Utiles de Mayence; 1772.

[11] Bravais A. Mémoire sur les systèmes formés par des points distribués regulièrement sur un plan ou dans l'espace. J De l'Ecole Polytech 1848;19:1.

[12] Gibbs JW. The collected works of J. Willard gibbs, in two volumes. In: Longley WR, Van Name RG, editors. New Haven: Yale University Press; 1993. 1957 (1928).

[13] Jackson KA. In: Feigelson RS, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 35.

[14] Wilson HA. Phil Mag 1900;50:238.

[15] Knudsen M. Ann Phys 1909;29:179.

[16] Volmer M, Estermann I. Z Phys 1921;7:13.

[17] Kossel W, Nachr. Gesell. Wiss. Gottingen. Math-Phys. K1 1927;135.

[18] Stranski IN. Phys Chem 1928;136:259.

[19] Kossel W. In: Falkenhagen H, editor. Quantentheorie und Chemie, Leipzig; 1928. p. 46.

[20] Stranski IN, Kaishev R. Z Krist 1931;78:373.

[21] Buckley HE. Crystal growth. New York: John Wiley; 1951.

[22] Frank FC. The influence of dislocations on crystal growth. Discuss Faraday Soc 1949;5:48.

[23] Heck CM. Phys. Rev 1937;51:686.

[24] G.G. Lemmlein, Vestn. Akad. Nauk SSSR 4, 119 (1945)

[25] Donnay JDH, Harker D. Am Mineral 1937;22:44.

[26] Hartman P, Perdok W. Acta Cryst 1955;8:495.

[27] Bennema P. Growth and morphology of crystals. In: Hurle DTJ. Handbook of crystal growth, Parat 1a. North Holland (Amsterdam); 482.

[28] Burton JA, Prim RC, Slichter WP. J Chem Phys 1953;21:1987.

[29] Burton JA, Slichter WP. In: Feigelson Robert S, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 43.

[30] Millman S, editor. A history and of engineering and science in the Bell system, physical sciences (1925–1980). AT&T Bell Laboratories; 1983.

[31] Carruthers JR. In: Bardsley W, Hurle DTJ, Mullin JB, editors. Crystal growth: a tutorial approach. Amsterdam: North Holland; 1979. p. 157.

[32] Chernov AA. In: Feigelson Robert S, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. 47 and Sov. Phys. Uspekhi 4 (1961) 116.

[33] Burton WK, Cabrera N. Discuss Faraday Soc 1949;5:33.

[34] Frenkel J. Phys J USSR 1932;1:498.

[35] Becker R, Döring W. Ann Phys 1935;24:719.

[36] Burton WK, Cabrera N, Frank FC. Philos Trans R Soc London Ser A 1951;243:299.

[37] Chernov AA. J Cryst Growth 2004;264:499.

[38] Jackson KA. Interface structure. In: Doremus RH, Roberts BW, Turnbull D, editors. Growth and perfection of crystals. London: Chapman and Hill; 1958. p. 319.

[39] Jackson KA, Hunt JD. Acta Met 1965;13:1212.

[40] Woodruff DP. The solid-liquid interface. Cambridge solid state science series. London: Cambridge University Press; 1973.

[41] Rutter JW, Chalmers B. Can J Phys 1953;31:15.

[42] Tiller WA, Jackson KA, Rutter JW, Chalmers B. Acta Met 1953;I:428.

[43] Leamy HJ, Gilmer GH, Jackson KA. In: Blakely JM, editor. Surface physics on materials, vol. 1. New York: Academic Press; 1975. p. 121.

[44] Leamy HJ, Gilmer GH. J Cryst Growth 1974;24/25:499.

[45] Mullins WW, Sekerka RF. J Appl Phys 1963;34:323; 35 (1964):444.

[46] Voronkov VV. Siev Phys Solid State 1965;6:2378.

[47] Cahn JW. Crystal growth. In: Steffen Peiser H, editor. Proceedings of the international conference on crystal growth. Oxford: Pergamon Press; 1967. p. 681.

[48] Coriell SR, Sekerka RF. J Cryst Growth 1976;34:157.

[49] Chernov AA. J Cryst Growth 1974;24/25:11.

[50] Hurle DTJ. J Cryst Growth 1969;5:162.

[51] Temkin DE. Dokl Akad SSR 1960;132:1307.

[52] Bolling GF, Tiller WA. J Appl Phys 1961;32:2587.

[53] Hamilton DR, Seidensticker RG. J Appl Phys 1960;31:1165.

[54] Glicksman ME, Lupulescu A. In: Feigelson Robert S, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 97.

[55] Vogel R. Z Anorg Chem 1912;76:425.

[56] Tiller WA. Liquid metals and solidification. Cleveland (OH): ASM; 1958. p. 276.

[57] Jackson KA, Hunt JD. Trans Metall Soc AIME 1966;236:1129.

[58] Glicksman ME. Principles of solidification. New York: Springer; 2010.

[59] Wilcox WR. J Appl Phys 1964;35:636.

[60] Carruthers JR, Nassau K. J Appl Phys 1968;39:5205.

[61] Hurle DTJ, Rudolph P. In: Feigelson Robert S, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 109.

[62] Wang CA, Carlsen D, Motakef S, Wiegel M, Wargo MJ, for Witt AW. In: Feigelson Robert S, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 125.

[63] Nassau K, Nassau J. In: Feigelson Robert S, editor. 50 years of progress in crystal growth. Amsterdam: Elsevier; 2004. p. 9.

[64] Merker L. In: Feigelson Robert S, editor. 50 years of progress in crystal growth. Amsterdam: Elsevier; 2004. p. 23.

[65] Tammann G. Metallography [Dean and Swenson, New York, Trans.]. N.Y: The Chemical Catalog Co.; 1925. p. 26.

[66] Obreimov J, Schubnikov L. Z Phys 1924;25:31.

[67] Czochralski J. Z Phys Chem 1918;92:219.

[68] Tomaszewski P. Jan czochralski and his method. Wroclas-Keynia; 2003.

[69] von Wartenberg H. Verh Dtsch Phys Ges 1918;20:113.

[70] von Gomperz E. Z Phys 1921;8:184.

[71] Tyndall EPT. Phys Rev 1928;31:313.

[72] Walther H. Rev Sci Insts 1937;8:406.

[73] Uecker R. JCG 2014;401:7–24.

[74] Evans J. Phys Rev 1940;57:47.

[75] Teal GK, Little JB. Phys Rev 1950;78:647.

[76] Shockley W, Sparks M, Teal GK. Phys Rev 1951;83:139.

[77] Teal GK, Buehler E. Phys Rev 1952;87:190.

[78] Pfann WG. J Met 4, Trans AIME 194 (1952):747.

[79] Pfann WG, Olsen KM. Phys Rev 1953;89:322.

[80] Dash WC. J Appl Phys 1959;30:459.

[81] Metz EPA, Miller RC, Mazelsky R. J Appl Phys 1962;33(6):2016.

[82] Mullins JB, Straughan BW, Bickel WS. J Phys Chem Solids 1965;26:82.

[83] Mullins JB, Heritage RJ, Holiday CH, Straughan BW. J Cryst Growth 1968;3–4:261.

[84] Azuma K. Japanese Patent 60–111299, Furukawa Ind.; 1983.

[85] Maiman. Nature 1960;187:493.

[86] Nassau K, Van Uitert LG. J Appl Phys 1960;31:1508.

[87] Ballman AA. J Am Ceram Soc 1965;48:112.

[88] Nassau K. In: Feigelson Robert S, editor. 50 years of progress in crystal growth. Amsterdam: Elsevier; 2004. p. 155.

[89] Fedulov SA, Shapiro ZI, Ladyshinskii PB. Kristallografia 1965;10(22):268.

[90] Ballman AA. J Cryst Growth 1967;1:109.

[91] Bardsley W, Cockayne B. In: Peiser HS, editor. Crystal growth. Oxford: Pergamon; 1967. p. 109.

[92] Kestigian M, Holloway WW. In: Peiser HS, editor. Crystal growth. Oxford: Pergamon; 1967. p. 451.

[93] Cockayne B, Chesswas M, Gasson DB. J Mater Sci 1967;2:7.

[94] Charvat FR, Smith JC, Nestor OH. In: Peiser HS, editor. Crystal growth. Oxford: Pergamon; 1967. p. 45.

[95] Brandle CD, Valentino AJ. J Cryst Growth 1972;12:3.

[96] Brandle CD. J Cryst Growth 2004;264:593.
[97] Carruthers JR, Peterson GE, Grasso M. J Appl Phys 1971;42:1846.
[98] Cockayne B, Chesswas B, Gasson DB. J Mater Sci 1968;3:224.
[99] Derby JJ, Brown RA. J Cryst Growth 1987;83:137.
[100] Wiffen PAC, Brice JC. J Cryst Growth 1976;13:13.
[101] Bridgman PW. Proc Am Acad Arts Sci 1914;49:625.
[102] Stockbarger DC. Am Mineral 1927;12:26.
[103] Stockbarger DC. Rev Sci Inst 1936;7:133.
[104] Scheel HJ, Schultz-Dubois EO. J Cryst Growth 1971;8:304.
[105] Zawilski KT, Custodio MCC, DeMattei RC, Feigelson RS. J Cryst Growth 2005;282:236.
[106] Gault WA, Monberg EM, Clemens JE. J Cryst Growth 1986;74:491.
[107] Mullins JB. In: Feigelson Robert S, editor. 50 years of progress in crystal growth. Amsterdam: Elsevier; 2004. p. 141.
[108] Nacken R. Neues Jahrab Mineral Geol 1915;2:133.
[109] Adams JM, Lewis W. Rev Sci Instrum 1935;5:400.
[110] Buehler E, Walker AC. Growing quartz. Bell Telephone Publication; 1949.
[111] Kyropoulos S. Z Anorg Chem 1926;154:308.
[112] Kyropoulos S. Z Phys 1930;63:849.
[113] Korth K. Z fur Phys 1933;8:677.
[114] Chamberlain K. Rev Sci Instrum 1938;9:322.
[115] Bliss D. In: Feigelson Robert S, editor. 50 years of progress in crystal growth. Amsterdam: Elsevier; 2004. p. 29.
[116] Bryant GG, Bliss DF, Gabbe DR, Zach FX, Iseler GW. Photonic device engineering for dual-use applications, 2481. SPIE; June 30, 1995. 232.
[117] Stober F. Z Krist 1925;61:299.
[118] Schmid F, Viechnicki D. J Am Ceram Soc 1970;53(9):528.
[119] Kattik C, Schmid F. J Cryst Growth 2001;225:572.
[120] Kapitza P. Proc R Soc London 1928;A 119:358–442.
[121] Mullin JB. In: Miller LS, Mullin JB, editors. Electronic materials: from silicon to organics. New York: Plenum Press; 1991. p. 118.
[122] Schunemann PG, Pollak TM. MRS Bull 1998;23(7):23.
[123] da C. Andrade EN, Roscoe R. Proc Phys Soc London 1937;49:152.
[124] Pfann William G. Zone melting. Wiley series on the science. New York (NY): John Wiley & Sons; 1958.
[125] Theuerer HC. J Met 8, Trans AIME 206(1956):1316.
[126] Poplawsky RP. J Appl Phys 1962;33:1616.
[127] Horn FH. J Electrochem Soc 1958;105:393.
[128] Dash WC. J Appl Phys 1960;31:736.
[129] Poplawski RP, Thomas Jr JE. Rev Sci Inst 1960;31:1303.
[130] Feigelson RS. Mater Sci Eng 1988;B1:67.

[131] Mimura Y, Okamura Y, Komazawa Y, Ota C. Jpn J Appl Phys 1980;19:269.

[132] Yoon DH, Yonenaga I, Fukuda T, Ohnishi N. J Cryst Growth 1994;144:339.

[133] Haggerty JS, Menashi P. Prod fibers by floating zone fiber drawing technique. Final Rep. NASA-CR-120948; 1971.

[134] Burris CA, Stone J. Appl Phys Lett 1975;26:318.

[135] Fejer MM, Nightingale JL, Magid GA, Byer RL. Rev Sci Inst 1984;55:1791.

[136] Luh Y, Feigelson RS, Fejer MM. J Cryst Growth 1986;78:135.

[137] Magel GA, Fejer MM, Byer RL. Appl Phys Lett 1990;56:108.

[138] Chani VI. In: Fukuda T, Chani VI, editors. Advances in materials research-shaped crystals, vol. 8. Berlin: Springer; 2007. p. 3.

[139] Feigelson RS, Route RK. J Cryst Growth 1990;104:789.

[140] Feigelson RS. Growth of shaped crystals. In: Arend H, Hulliger J, editors. Crystal growth in science and technology. Plenum Publishing Corp; 1989. p. 275.

[141] Polanyi M. My time with X-rays and crystals. In: Ewald PP, editor. 50 years of X-ray diffraction. Utrecht: International Union of Crystallography; 1962. p. 629.

[142] Mark H, Polanyi M, Schmid E. Z Phys 1923;12:58.

[143] Stepanov AV. Zh Tech Fiz 1959;29:382.

[144] Antonov PI, Nikandrov SP. J Cryst Growth 1980;50:3.

[145] LaBelle Jr HE, Mlavsky AI. Mater Res Bull 1971;6:581.

[146] VerSnyder FL, Shank ME. Mater Sci Eng 1970;6:231.

[147] LaBelle HE. In: Feigelson Robert S, editor. 50 years of progress in crystal growth. Amsterdam: Elsevier; 2004. p. 179.

[148] Mlavsky AI. In: Feigelson Robert S, editor. 50 years of progress in crystal growth. Amsterdam: Elsevier; 2004. p. 177.

[149] Surek T, Coriell SR, Chalmers B. J Cryst Growth 1980;50:21.

[150] Tatarchenko VA, Brener EA. J Cryst Growth 1980;50:126.

[151] Seidensticker RG. J Cryst Growth 1977;39:17.

[152] Belouet C, Brissot JJ, Martres R, Ngo-Tich Phuoc. In: Proc. photovoltaic solar energy conf. Luxembourg; 1977. p. 164.

[153] Zook J, Koepke BG, Grung BL, Leipold MH. J Cryst Growth 1980;50:260.

[154] Baghdadi A, Gurtler RW. J Cryst Growth 1980;50:236.

[155] Bleil CE. J Cryst Growth 1969;5:99.

[156] Kear BH, Piearcey BJ. Trans AIME 1967;238:1209.

[157] Leomonova EE, Osiko VV. Growth of zirconia crystal by skull-melting technique. In: Scheel HJ, Fukuda T, editors. Crystal growth technology. John Wiley & Sons, Ltd; 2003. p. 461.

[158] Castonguay RA, Menashi WF, Wenckus JF, Arthur D. Little, Inc., U.S. Patent 4049384, (September 20, 1977).

[159] Kisil II, Taranyuk VI, Yaroslavkin SV. Funct Mater 2008;15:600.

[160] Wulff G. Z Krist 1901;34:449.

[161] Kruger F, Finke W. Dtsch Reich Patent. 228, 246, K1, 120, Gr2 (November 5, 1910).

[162] Valeton JJP, Sachs K. Math-physik Klasse, 67. Leipzig: Ges. Wiss; 1915. p. 1.

[163] Nacken R. Z Instrumentenk 1916;36:12.

[164] Walker AC, Kohman GT. Trans Am Inst Elec Engrs 1949;68:222.

[165] Holden AN. Discuss Faraday Soc 1949;5:312.

[166] Moore RW. J Am Chem Soc 1919;41:1060.

[167] Walker AC. Bell Lab Rec 1947;25:35.

[168] Holden AN, Singer Phylis. Crystals and crystal growing. New York: Anchor Press/MIT Press; January 1, 1982. p. 318.

[169] Loiacono GM, Zola JJ, Kostescky G. J Cryst Growth 1983;62:545.

[170] Rashkovich LN. Rapid growth of large crystals for nonlinear optics from solution. Bull Acad Sci USSR 1984;9:15.

[171] Zaitseva N, Carmen L. Rapid growth of KDP-type crystals, progress in crystal growth and characterization of materials. Pergammon 2001;1.

[172] Glusker JP, Dorothy crowfoot hodgkin (1910–1994). Protein Sci 1994;3:2465.

[173] Brink C, Hodgkin DC, Lindsey J, Pickworth J, Robertson JH, White JG. Nature 1954;174:1169.

[174] Protein crystal growth. In: Feigelson RS, editor. Proc. first int. conf. on crystal growth. Stanford (CA): Stanford University; August 1985. North Holland, Amsterdam (1986).

[175] McPherson A. Preparation and analysis of protein crystals. New York: John Wiley and Sons; 1982.

[176] McPherson A. Crystallization of biological macromolecules. NY: Cold Spring Harbor Lab. Press, Cold Spring Harbor Press; 1999.

[177] Ward KB, Perozzo MA, Zuk WM. J Cryst Growth 1988;90:325.

[178] Land TA, Malkin AJ, Kuznesov YG, McPherson A, DeYoreo JJ. Pus Rev Lett 1995;75:2774.

[179] Malkin AJ, Land TA, Kuznetsov YG, McPherson A, DeYoreo JJ. Phys Rev Lett 1995;75:2778.

[180] Liesegang RE. Nat Wochschr 1896;11:221.

[181] Hatschek E. J Soc Chem Ind 1911;30:276.

[182] Dreaper WP. J Soc Chem Ind 1913;32:678.

[183] Fisher LW, Simons FL. Am Mineral 1926;11:124 and 200.

[184] Henisch HK. Crystal growth in gels. University Park (PA): The Pennsylvania State University Press; 1970.

[185] Patel AR, Verikateswara Rao A. Bull Mat Sci 1982;4:527.

[186] Jiang H, Kloc C. MRS Bull 2013;38:28.

[187] Carman L, Zaitseva N, Martinez HP, Rupert B, Pawelczak I, Glenn A, et al. J Cryst Growth 2013; 368:56.

[188] Bethe K, Weltz F. Mat Res Bull 1971;6:209.

[189] Linares RC. J Am Ceram Soc 1962;45:307.

[190] Elwell D, Scheel HJ. Crystal growth from high temperature solutions. London: Academic Press; 1975.

[191] Wöhler F. Ann Chim Phys 1823;29(2):43.

[192] Doelter C. Z Krist 1886;11(29):40.

[193] Wöhler F, St Claire Deville H. Ann Chem Pharm 1857;101(113):347.

[194] Fremy E, Feil C. Compt Rend 1877;85:1029.

[195] Giess EA. J Am Ceram Soc 1964;47:388.

[196] Austerman SB. J Am Ceram Soc 1963;46:6.

[197] Newkirk HW, Smith DK. Am Min 1965;50:44.

[198] Timofeeva VA. J Cryst Growth 1968;3/4:496.

[199] Nielsen JW, Dearborn EF. Phys Chem Solids 1958;5:202.

[200] Van Uitert LG, Grodkiewicz WH, Dearborn EF. J Am Ceram Soc 1965;48:105.

[201] Wood DL, Remeika JP. J Appl Phys 1966;37:1232.

[202] Scheel HJ. In: Feigelson Robert S, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 173.

[203] Reisman A, Holtzberg F. J Am Chem Soc 1955;77:2115.

[204] Mizell GJ, Fay WR, Alekel III T, Rytz D, Garrett M. Visible and UV lasers, 2115. SPIE; 1994.

[205] Fluckiger U, Arend H. J Cryst Growth 1978;43:406.

[206] Miller CE. J Appl Phys 1958;29:233.

[207] Linares RC. J Appl Phys 1964;35:433.

[208] Belruss V, Kalnajs J, Linz A, Foweiler R. Mat Res Bull 1971;6:899.

[209] Chen CT, Wu B, Jiang A, You G. Sci Sin Ser B 1985;28:235.

[210] Nelson DF, Remeika JP. J Appl Phys 1964;35:522.

[211] Tolksdorf W. J Cryst Growth 1968;3/4:463.

[212] Nielsen JW. Electronics 1964;37:44.

[213] Grodkiewicz WH, Dearborn EP, Van Uitert LG. In: Steffen Peiser H, editor. Crystal growth, proceedings of the international conference on crystal growth. Oxford: Pergamon Press; 1967. p. 441.

[214] Capper P, Mauk M, editors. Liquid phase epitaxy of electronic, optical, and optoelectronic materials. Chicester (England): John Wiley; 2007.

[215] Rakovan J. Rocks Miner 2006;81:317.

[216] Frankenheim ML. Ann Phys 1836;3(7):516.

[217] Baker TV. J Chem Soc Trans 1906;89:1120.

[218] Royer L. Bull Soc Franc Mineral 1928;51:7.

[219] Nelson H. RCA Rev 1963;24:603.

[220] Rupprecht H. Symposium on GaAs, Reading (England). (1966) 57.

[221] Panish MB, Sumski S, Hayashi I. Metall Trans 1971;2:795.

[222] Linares RC, McGraw RB, Schroeder JB. J Appl Phys 1965;36:2884.

[223] C. Linares R. J Cryst Growth 1968;3/4:443.

[224] Bobeck AH, Della Torre E. Magnetic bubbles. Amsterdam: North-Holland; 1975.

[225] Hibiya T. J Cryst Growth 1983;64:400.

[226] Fratello VJ, Wolfe R. In: Francombe MH, Adam JD, editors. Magnetic film devices. Academic Press; 2000.

[227] Bednorz JG, Müller KA. Z Phys-Condensed Matter 1986;64:189.

[228] Scheel HJ, Klemenz C, Reinhart FK, Lang HP, Güntherodt HJ. Appl Phys Lett 1994;65(7):901.

[229] Byrappa K, Yoshimura M. Handbook of hydrothermal technology. Norwich (NY): Noyce Publications; 2000.

[230] Ballman AA, Laudise RA. In: Gilman JJ, editor. The art and science of growing crystals. New York: Wiley; 1963. p. 231.

[231] Senarmont H. Ann chem phys 1851;32:129.

[232] Spezia G. Acad Sci Torino Atti 1905;40:254.

[233] Walker AC, Buehler E. Indust Eng Chem 1950;42:1369.

[234] Laudise RA, Sullivan RA. Chem Eng Prog 1959;55:55.

[235] Laudise RA. J Am Chem Soc 1959;81:562.

[236] Tian ZR, Voigt JA, Liu J, Mckenzie B, Mcdermott MJ. J Am Chem Soc 2002;124:12954.

[237] Vayssieres L, Keis K, Lindquist SE, Hagfeldt AJ. Phys Chem B 2001;105:3350.

[238] E Greene L, Law M, Goldberger J, Kim F, Johnson JC, F Zhang Y, et al. Angew Chem Int Ed 2003;42:3021.

[239] Zhang J, Sun LD, Liao CS, Yan CH. Chem Commun 2002:262.

[240] Kunnmann W. In: Lefever RA, editor. Preparation and properties of solid state materials. New York: Dekker; 1971. p. 1.

[241] Bostanov V. J Cryst Growth 1977;42:194.

[242] DeMattei RC, Huggins RA, Feigelson RS. J Cryst Growth 1976;34:1.

[243] St Claire-Deville H. Compt Rend 1854;39:323.

[244] Ullik F. Ber Akad Wien 1865;52:115.

[245] Cohen U, Huggins RA. J Electrochem Soc 1976;123:381.

[246] Feigelson RS. In: Holt SL, Milstein JB, Robbins M, editors. Solid state chemistry, a contemporary overview; 1980. p. 243. Am. Chem. Soc. Washington, DC.

[247] Hodes G. In: Rubinstein I, editor. Electrodeposition of II–VI semiconductors, physical electrochemistry. New York: M. Dekker Inc.; 1995. p. 515.

[248] Turner AK, Woodcock JM, Ozsan E, Summers JG. In: Luque A, Sala A, Palz W, Dos Santos A, Helm P, editors. Proceedings of the 10th European Solar Energy Conference, held at Lisbon Portugal: Springer [Netherlands]; April, 1991 8–12, p.791.

[249] Isshiki M, Endo T, Masumoto K, Usui Y. J Electrochem Soc 1990;137:2697.

[250] Lincot D, Ortega-Borges R, Froment M. Appl Phys Lett 1994;64:569.

[251] Chaudary GN, Sardesai SN, Sathaye SD, Rao VJ. J Mat Sci 1992;27:4647.

[252] Lincot D, Furlong MJ, Froment M, Cortes R, Bernard MC. Mat Res Soc Symp Proc 1997;451:223.

[253] Schlesinger TE, Rajeshwar K, De Tacconi NR. Electrodeposition of semiconductors. In: Schlesinger M, Paunovic M, editors. Modern electroplating. New York: John Wiley & Sons; 2010.

[254] Izaki M, Omi T. Appl Phys Lett 1996;68:2439.

[255] Peulon S, Lincot DJ. J Electrohem Soc 1998;145:864.

[256] Liu R, Vertegel AA, Bohannan EW, Sorenson TA, Switzer JA. Chem Mater 2001;13:508.

[257] Vanheusden K, Warren WL, Seager CH, Tallant DR, Voigt JA, Gnade BE. J Appl Phys 1996;79:7983.

[258] Yshida T, Tochimoto M, Schlettwein D, Wöhrle D, Sugiura T, Minoura H. Chem Mater 1999;11:2657.

[259] Xu L, Guo Y, Liao Q, Zhang J, Xu D. J Phys Chem B 2005;109:13519.

[260] Allendorf M. From Bunsen to VLSI. Electrochem Soc Interface 1998. Spring.

[261] Bunsen R. J Prakt Chem 1852;56:53.

[262] St Claire-Deville H. Annals 1861;120:176.

[263] Tietjen JJ. Ann Rev Mater Sci 1973;3:317.

[264] Brown FC. Phys Revs 1914;4:85.

[265] Koref F. Z Electrochem 1922;28:511.

[266] Van Arkel AE. Physica 1923;3:76.

[267] Gross R. Z Krist 1921;56:421.

[268] Gross R, Volmer M. Z Phys 1921;5:188.

[269] Straumanis M. Z Phys Chem 1932;B19:63.

[270] Kaldis E. In: Goodman CHL, editor. Crystal growth, theory, and techniques, vol. 1. England: Plenum Press; 1974.

[271] Faktor MM, Garrett J. Growth of crystals from the vapor. London: Chapman and Hall; 1974.

[272] Wilke KT. Kristsall-zuchtung. Berlin: VEB Deutscher Verlag d. Weissenschaften; 1973.

[273] Pizzarello F. J Appl Phys 1954;25:804.

[274] Avinor M. Philips Res Rep 1959;14:211.

[275] Piper WW, Polich SJ. J Appl Phys 1961;32:1278.

[276] Fochs PD. J Appl Phys 1960;31:1733.

[277] Prior AC. J Electrochem Soc 1961;108:82.

[278] Frisch O. Ann Phys 1935;22:375.

[279] Lely JA. Ber Dtsch Keram Ges 1955;32:229.

[280] Tairov TM, Tsvetkov VF. Neorg Mater (USSR) 1977;13:448.

[281] Heindl J, Strunk PH, Heydemann VD, Pensl G. Phys Stat Sol (A) 1997;167:251.

[282] Feigelson RS, Route RK, Kao TM. J Cryst Growth 1985;72:585.

[283] Van Arkel AE, de Boer JH. Zeit anorg Allgem chem 1925;148:345.

[284] Schäfer H. Chemical transport reactions. New York: Academic Press; 1963.

[285] Laudise RA. The growth of single crystals. New Jersey: Prentice-Hall; 1970. 234.

[286] Lorenz R. Chem Ber 1891;24:1509.

[287] Frerichs R. (1996) 1701. Phys Rev 72 (1947) 594.

[288] Schönherr H. The growth of large crystals from the vapor state. In: Crystals, growth, properties and applications. Berlin: Springer-Verlag; 1980. p. 51.

[289] WZ Li, SS Xie, LX Qian, BH Chang, BS Zhou, WY Zhou, et al. Science;274.

[290] Scott TR, King G, Wilson JM. UK Patent 778,383 (1957).

[291] Miederer W, Ziegler G, Dotzer R. US Patent 3,226,270 (1965).

[292] Manasevit HM, Simpson WJ. J Electrochem Soc 1969;12:1725.

[293] Manasevit HM. In: Feigelson Robert S, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 217.

[294] Seki Y, Tanno K, Iida K, Ichiki E. J Electrochem Soc 1975;122:1108.

[295] Dupuis RD, Dapkis PD. Appl Phys Lett 1978;32:40.

[296] Stringfellow GB. Annu Rev Mater Sci 1978;8:73.

[297] Stringfellow GB. Organometallic vapor phase epitaxy: theory and practice. 2nd ed. Boston: Academic Press; 1999.

[298] Akasaki I, Amano H. Jpn J Appl Phys 2006;45:9001.

[299] Nakamura S, Mukai T, Senoh M. Appl Phys Lett 1994;64:1687.

[300] Nakamura S, Fasol G. The blue laser diode. Berlin: Springer; 1997.

[301] Liu BL, Lachab M, Jia A, Yoshikawa A, Takahashi K. J Cryst Growth 2002;234:637.

[302] Stringfellow GB. J Cryst Growth 2004;264(4):620–30.

[303] Holloway PH, McGuire GE. Handbook of compound semiconductors; growth, processing, characterization and devices. Noyes Publications; 1995.

[304] Joyce BA, Bradley RR. Phils Mag 1966;14:289.

[305] Davey JE, Pankey T. J Appl Phys 1968;39:1941.

[306] Arthur JR. J Appl Phys 1968;39:4032.

[307] Cho AY, Panish MB, Hayashi J. In: Proceedings of the symposium on GaAs and related compounds. Inst. of physics, vol. 2. Germany: Aachen; 1970. p. 118.

[308] Joyce BA, Joyce TB. In: Feigelson Robert S, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 203.

[309] Grove WR. Trans R Soc 1852;142:87.

[310] Williams P. Surf Sci 1979;90:588.

[311] Kelly PJ, Arnell RD. Vacuum 2000;56:159.

[312] Penning PM. Physica 1936;3:873.

[313] Naoe M, Yamanaka S, Hoshi Y. IEEE Trans Magn 1980;Mag-16:646.

[314] Suntola T. US Patent 4058430 (1977).

[315] Aleskovskii VB. J Appl Chem USSR 1974;47:2207.

[316] Suntola T. Mater Sci Rep 1989;4:261.

[317] Triani G, Evans PJ, Mitchell DRG, Attard DJ, Finnie KS, James M, et al. Proc SPIE 2005;5870:9.

[318] Ritala M, Leskela M. Nanotechnology 1999;10:19.

[319] Wagner RS, Ellis WC. Appl Phys Lett 1964;4:89.

[320] Kuykendall T, Pauzauskie PJ, Zhang Y, Goldberger J, Sirbuly D, Denlinger J, et al. Nat Mater 2004;3:524.

[321] Schmidt V, Wittemann JV, Senz S, Gösele U. Adv Mater 2009;21:2681.

[322] Wacaser BA, Dick KA, Johansson J, Borgstrom MT, Deppert K, Samuelson L. Adv Mater 2009;21:153.

[323] Choi HJ, Johnson JC, He R, Lee SK, Kim F, Pauzauskie P, et al. Chem B 2003;107:8721.

[324] Choi HJ. Vapor-liquid-solid growth of semiconductor nanowires. In: Yi GC, editor. Semiconductor nanostructures for optoelectronic applications, vol. 1. Berlin: Springer-Verlag; 2012.

[325] Sheftal NN. Trends in real crystal formation and some principles for single crystal growth. In: Growth of crystals, vol. 10. New York: Consultants Bureau; 1976. p. 185.

[326] Geis MW, Tsaur BY, Flanders DC. Appl Phys Lett 1982;41:526.

[327] Givargizov EI, Kliya MO, Melik-Adamyan VR, Grebenko AI, DeMattei RC, Feigelson RS. J Cryst Growth 1991;112:758.

[328] Givargizov EI. Oriented crystallization on amorphous substrates. New York: Plenum Oress; 1991.

[329] Crnogorac F, Witte DJ, Pease RFW. J Vac Sci Technol B 2008;26:2520.

[330] Tracy Hall H. In: Feigelson Robert S, editor. 50 years progress in crystal growth. Amsterdam: Elsevier; 2004. p. 193.

[331] Hannay JB. Proc R Soc London 1879;30:450.

[332] Moissan H. Comptes Rendus 1894;118:320.

[333] Desch CH. Nature 1928;121:799.

[334] Lonsdale K. Nature 1962;196:104.

[335] Hazen RM. The diamond makers. Cambridge: Cambridge University Press; 1999.

[336] Tracy Hall H. Rev Sci Instrum 1960;31:125.

[337] Barnard AS. The diamond formula: diamond synthesis: a gemological perspective. Oxford: Butterworth-Heinemann; 2000.

[338] Eversole WG. Synthesis of diamond, US Patent 3,030,188, (April 17, 1962).

[339] Angus JC. J Appl Phys 1968;39:2915.

[340] Deryagin BV, Fedoseev DV. Rus Chem Rev 1970;39:783.

[341] Gracio JJ, Fan QH, Madaleno JC. J Phys D Appl Phys 2010;43:1.

[342] Angus JC. A short history of diamond synthesis. In: Asmussen J, Reinhard DK, editors. Diamond films handbook. New York: Marcel Dekker Inc.; 2002.

[343] Matsumoto S, Sato Y, Kamo M, Setaka N. Jpn J Appl Phys 1982;21L:183.

[344] Kamo M, Sato Y, Matsumoto S, Setaka N. J Cryst Growth 1983;62:642.

[345] Matsui Y, Matsumoto S, Setaka N. J Mater Sci 1983;2:532.

[346] Decarli P, Jamieson J. Science 1961;133:1821.

[347] Galimov ÉM, Kudin AM, Skorobogatskii VN, Plotnichenko VG, Bondarev OL, Zarubin BG, et al. Dokl Phys 2004;49:150.

[348] Ostwald W. Zeit Phys Chem 1897;22:289.

[349] Anderson RJ. J Frankl Inst 1919;187:1.

[350] Carpenter HCH, Elam CF. Proc R Soc London 1921;100:329.

[351] Bailey DJ, Brewer EG. Metal Trans A 1975;6A:403.

[352] Kulper LL. U. S. Patent 3413157, November 26, 1968.

[353] Mayer JK, Nicolet MA, Lau SS. U. S. Patent 4012235, March 17, 1977.

[354] Mayer JK, Lau SS. Electronic materials science: for integrated circuits in Si and GaAs. New Jersey: Prentice Hall; 1989.

[355] Kwon S, Hackenberger WS, Rehrig PW, Lee J-B, Heo T-M, Kim D-H, Lee H-Y. In: Proceedings 2004 IEEE international ultrasonics, ferroelectrics and frequency control conference; 2004. p. 153.

[356] Hurle DTJ, editor. A perspective on crystal growth: a historical collection in celebration of 25 years of the journal of crystal growth. Elsevier Science Publishers; 1992.

[357] Ivantsov GP. Dokl Akad Nauk SSSR 1947;58:567.

[358] Schawlow AL, Townes CH. Phys Rev 1958;112:1940.

[359] Dunn CG, Nonken GC. Metal Prog 1953;71:71.

2 Phase Equilibria

Detlef Klimm

LEIBNIZ INSTITUTE FOR CRYSTAL GROWTH (IKZ), BERLIN, GERMANY

CHAPTER OUTLINE

2.1 Introduction 86
2.1.1 Basic Terms 86
2.1.1.1 Components and Concentrations 86
2.1.1.2 Phases and Phase Rule 88
2.1.1.3 System 89
2.1.2 Thermodynamic Functions and Potentials 90
2.1.2.1 Specific Heat Capacity 90
2.1.2.2 Enthalpy 91
2.1.2.3 Entropy 94
2.1.2.4 Gibbs Free Energy 96
2.1.3 Phase Transitions 98
2.1.4 Calculation of Phase Diagrams 103
2.1.4.1 Miscibility 103
2.1.4.2 Analytical Expressions 106
2.1.4.3 Minimization of the Gibbs Energy 109
2.2 Equilibria Between Condensed Phases 110
2.2.1 One Component 110
2.2.1.1 Pressure–Temperature Diagrams 110
2.2.1.2 Other Fields 111
2.2.2 Two Components 112
2.2.2.1 Total Miscibility in 2 Phases 113
2.2.2.2 Eutectics and Eutectoids 117
2.2.2.3 Segregation and Lever Rule 119
2.2.2.4 Intermediate Compounds 121
2.2.2.5 Peritectics and Peritectoids 122
2.2.2.6 Syntectics and Monotectics 124
2.2.3 Three and More Components 124
2.2.3.1 x – y *Diagrams* 124
2.2.3.2 Concentration Triangles 126

Handbook of Crystal Growth. http://dx.doi.org/10.1016/B978-0-444-56369-9.00002-2

2.2.3.3 Isopleth Sections 127
2.2.3.4 Reciprocal Salt Pairs 128
2.3 Equilibria Including Gas Phase 129
2.3.1 Volatile Species 130
2.3.2 Ellingham Type Diagrams 131
References 133

2.1 Introduction

In this chapter, phase diagrams will be introduced as basic tools for the development and understanding of crystal growth processes. The focus will be on reading diagrams that can be found in the literature, or that can be calculated with commercial or freeware computer programs such as described later in Section 2.1.4.3. If nothing else is written explicitly, equilibrium phase diagrams are presented. The crystal growth process itself, however, is a nonequilibrium process, where some nutrient phase (e.g., a melt, a solution, or a gas) is transformed to the desired solid (crystal) phase. The relevance of equilibrium phase diagrams for the description of crystal growth is given if the growth process is performed not too far from equilibrium conditions. This condition is typically fulfilled for most slow growth processes from the melt or from melt solutions such as Czochralski, Bridgman, Verneuil, top seeded solution growth (TSSG), micro-pulling-down (μ-PD), heat exchanger method (HEM), or vertical gradient freeze (VGF) technique. Also, growth processes from hot gas phases with sufficient density (not too far below ambient pressure 1 bar, such as chemical vapor transport (CVD) or physical vapor transport (PVT) can usually be described well by equilibrium phase diagrams. For further reading and for refereed collections of phase diagrams, some web sites prove to be helpful, e.g., [1–3].

2.1.1 Basic Terms

2.1.1.1 Components and Concentrations

A *component* is a substance that, under the given conditions, cannot be further divided into parts. A chemical element can always be chosen as component; but often it is more convenient to choose compounds instead, because a smaller number of components makes the description easier. It must be possible to create all compounds that may be found in equilibrium by chemical reactions between the components! If two or three compounds instead of elements are chosen to set up a phase diagram, the system is sometimes called *pseudobinary* or *pseudoternary*, respectively.

It is not always straightforward to decide whether a chemical compound can be used as component for the description of a specific system. For instance, vanadium(V) oxide V_2O_5 and molybdenum(VI) oxide MoO_3 are reacting upon heating to an intermediate phase [4]. This intermediate phase was initially described as V_2MoO_8, and binary phase diagrams similar to Figure 2.1 left, with V_2O_5 and MoO_3 as components, were published.

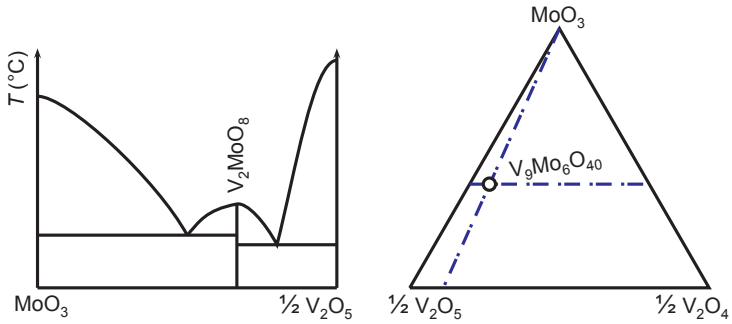


FIGURE 2.1 Left: Simplified binary phase diagram V_2O_5–MoO_3 under the incorrect assumption that the intermediate phase is V_2MoO_8 [6], entry 4467. Right: Correct phase triangle demonstrating that intermediate $V_9Mo_6O_{40}$ is ternary.

Further studies showed that $\frac{1}{9}$ of the V^{5+} ions are reduced to V^{4+} during the reaction, and the compound should rather be written as $V_9Mo_6O_{40} = 4\,V_2O_5 \cdot \frac{1}{2}V_2O_4 \cdot 6MoO_3$ [5]. Consequently, three components are required to give an appropriate description of the system, such as shown in the right panel of Figure 2.1. Of course, a description in terms of the chemical elements V–Mo–O would be valid too.

The share of the i-th component in the system is called its *concentration*. If not mentioned explicitly, concentrations x_i are given in molar fractions, or atomic fractions,

$$x_i = \frac{n_i}{\sum_{i=1}^{C} n_i} \tag{2.1}$$

where C is the number of components and n_i is the number of moles of component i. It is obvious that $\sum_i x_i = 1$, and hence in two-component systems only one concentration x is necessary. Mol-% or At-% values can be obtained by multiplying x_i with 100%.

For a specific chemical composition, the x_i depend on the choice of the components Y_i ($i = 1 \ldots C$). In dielectric systems (e.g., oxides, halides) often compounds A_mB_n (A: cation, B: anion) are chosen as components. Then it may be useful to scale them in such a way that all components bear the same number of cations (often one). The benefit of such scaling is that the resulting phase boundaries are often more symmetrical around invariant compositions. An example is given in Figure 2.1 for the system V_2O_5–MoO_3 where the right component is written as $VO_{2.5}$.

In systems with many intermediate compounds, the choice of components is ambiguous. Figure 2.2 demonstrates this for a system of the chemical elements (or basic

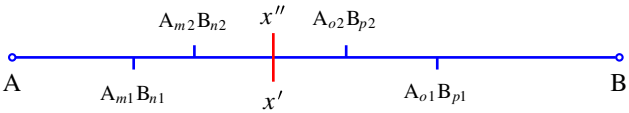


FIGURE 2.2 A specific composition $x = x' = x''$ can be expressed in terms of basic components (e.g., elements) A, B. Alternative descriptions are in terms of compounds $A_{m1}B_{n1}$, $A_{o1}B_{p1}$ (x') or $A_{m2}B_{n2}$, $A_{o2}B_{p2}$ (x''), respectively.

compounds such as simple oxides) A and B, with four intermediate compounds; the stoichiometry is expressed by indices. If a composition x is given in terms of $A_{m1}B_{n1}$ and $A_{o1}B_{p1}$ (below the concentration axis), the amounts of the basic compounds can be calculated by

$$\begin{aligned}[A] &= (1-x')\frac{m1}{m1+n1} + x'\frac{o1}{o1+p1} \\ [B] &= (1-x')\frac{n1}{m1+n1} + x'\frac{p1}{o1+p1}\end{aligned} \tag{2.2}$$

from the concentration x' and the stoichiometric indices of the components. The concentration with respect to the basic components is then easily expressed by

$$x = \frac{[B]}{[A]+[B]} \tag{2.3}$$

as a molar fraction. Rescaling of concentration data to $A_{m2}B_{n2}$, $A_{o2}B_{p2}$ (x'', above the concentration axis) is given by

$$x'' = \frac{x - \frac{n2}{m2+n2}}{\frac{p2}{o2+p2} - \frac{n2}{m2+n2}} \tag{2.4}$$

where the molar fraction x with respect to the basic components is given by Eqn (2.3). For the general case of component transformation in a multicomponent system ($C > 2$), the reader is referred to the literature, e.g., Ref. [7].

2.1.1.2 Phases and Phase Rule

A phase is a homogeneous part of a system, on scales that are large compared to atoms. This means that all physical quantities, and composition, are no function of the position $\vec{r}$, except differential fluctuations or gradients. Usually systems contain only one gas phase. Liquids (melts) of similar substances are often forming one liquid phase (molten slag, molten alloy, solution in a liquid). Dissimilar, nonmixing liquids form several phases (water and oil, molten slag and metal). Each solid compound is often forming a separate phase, with a specific crystal structure. Only if the components of a solid phase have identical crystal structure and are otherwise similar, they may intermix each other in arbitrary ratio. Such mixtures are one phase, called mixed crystal, or solid solution ($Ge_{1-x}Si_x$, $Ag_{1-x}Au_x$, $K_{1-x}Rb_xCl$; $0 \le x \le 1$). More details on phase miscibility will be given in Section 2.1.4.1.

For a given system with C components, Gibbs phase rule

$$P + F = C + 2 \tag{2.5}$$

limits the number P of coexisting phases and the number F of degrees of freedom that the system has in equilibrium. This means that in a one-component system, never more than three phases may exist at the same time. This is, e.g., demonstrated in Figure 2.3, where never more than three phase fields are touching each other in one point. Such triple points "T_i" describe conditions where $P = 3$ phases are in equilibrium.

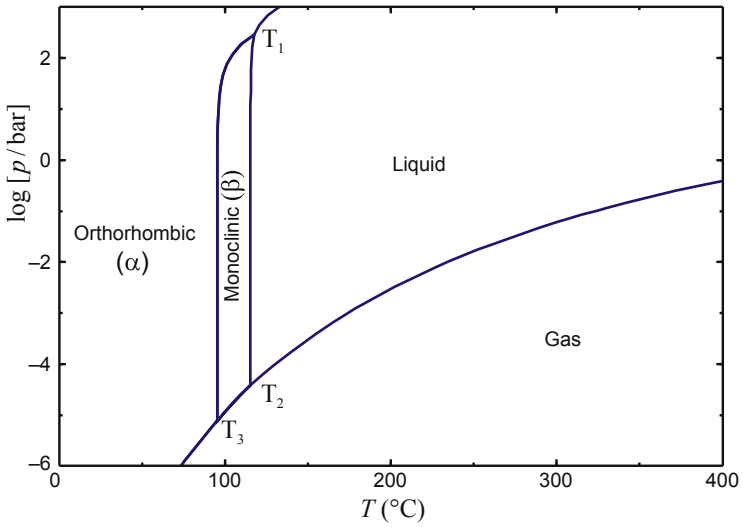


FIGURE 2.3 Phase diagram of sulfur in coordinates $T - \log[p]$. T_i are the invariant triple points where three phases are in equilibrium.

Often, experiments are performed under isobar conditions $p = 1$ bar, which consumes 1 degree of freedom. Then Eqn (2.5) simplifies to

$$P + F = C + 1 \tag{2.6}$$

which means that, e.g., in two-component systems, never more than three phases may coexist under isobar conditions. In the left panel of Figure 2.1, this is the case in the two "eutectic points" on both sides of V_2MoO_8 where the liquid phase (stable at high T) is in equilibrium with solid $V_2O_5 + V_2MoO_8$ or $V_2MoO_8 + MoO_3$, respectively. Both eutectic points are invariant, because any deviation would result in the disappearance of at least one phase. In phase fields where only one phase is stable (e.g., within the liquid phase field of Figure 2.1), one obtains from Eqn (2.6) $F = C - P + 1 = 2 - 1 + 1 = 2$. Indeed, one can move somewhat up or down (freedom of T) and left or right (freedom of x) without leaving the phase field.

2.1.1.3 System

If a specific volume V contains a limited amount of matter with mass m, and is characterized by some intensive properties, such as pressure p or temperature T, it may set up a system. Intensive properties do not change if two identical systems are united. Often, closed systems are considered, with no exchange of matter or energy with the surroundings. Then, in equilibrium T is constant for all parts of the system. However, p may be a function of the position $\vec{r}$ as a result of surface tension, e.g., in epitaxial layers. In contrast V and m are extensive properties. Their value depends on the system size.

Crystal growth is a dynamic process and matter is transported from a reservoir to a seed, where crystallization takes place as a result of supersaturation. Figure 2.4

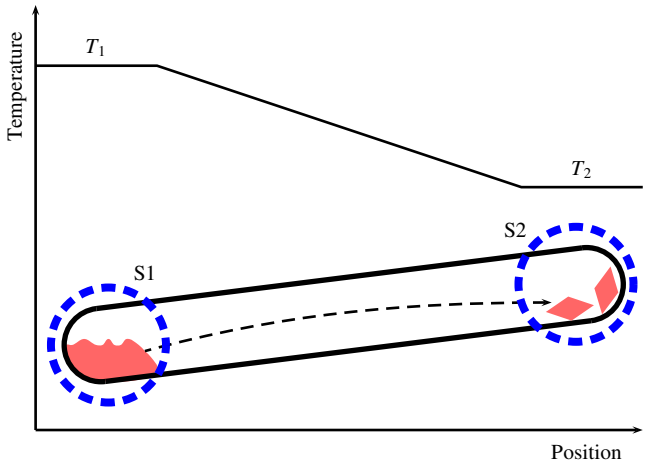


FIGURE 2.4 Crystal growth by sublimation in an ampoule. Feedstock is transported from a hot reservoir (temperature T_1) to the colder end (T_2).

demonstrates this for sublimation growth (physical vapor transport (PVT)) from the hot part of an ampoule (T_1) to the cold part (T_2). Under such conditions, it is not possible to define overall intensive conditions p, T for the whole ampoule. It may be useful, however, to consider separately system S1 where the solid feed is at p, T_1 in equilibrium with the surrounding gas, and system S2 where at p, T_2 gas is in equilibrium with the seed. Then, the growth process can be divided to three parts: (1) equilibration in S1, (2) transport of gas under cooling, (3) equilibration in S2.

2.1.2 Thermodynamic Functions and Potentials

2.1.2.1 Specific Heat Capacity

The specific heat capacity c is the basic thermodynamic quantity that can be measured. Typically, lattice vibrations (phonons) are the major contributors to the internal energy U of a substance, and U is a function of T. If an amount of heat energy Q is transferred to 1 mol of a sample, its temperature increases under isochor (volume $V =$ const.) or isobar (pressure $p =$ const.) conditions by

$$\Delta T_V = \frac{\Delta Q}{c_V} \tag{2.7}$$

$$\Delta T_p = \frac{\Delta Q}{c_p} \tag{2.8}$$

where c_V or c_p is the specific heat capacity of the material under the given conditions. Especially for technical purposes, the amount of substance is often given by the mass $m = nM$ (n, number of mol; M, molar mass). The difference

$$c_p - c_V = \frac{T}{n}\left(\frac{\partial V}{\partial T}\right)_p \left(\frac{\partial p}{\partial T}\right)_V \tag{2.9}$$

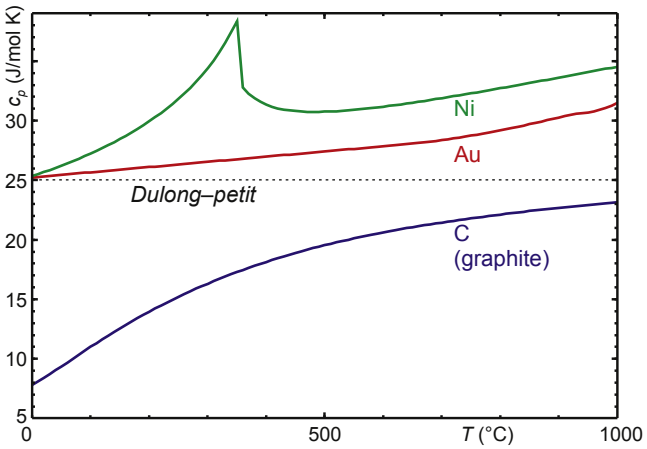


FIGURE 2.5 $c_p(T)$ functions for three chemical elements compared to the Dulong–Petit law (Eqn (2.10)).

is usually small, (but > 0) for condensed phases with small thermal expansion $\partial V/\partial T$ (solids, liquids). For ideal gases, one has $c_p - c_V = R$ (gas constant).

Dulong and Petit found that, for many solid chemical elements, the specific heat capacity approaches 25 J/(mol K) at room temperature (ref. Eqn (2.10) and Figure 2.5). For very low temperature $c_p(T)$ follows a T^3 law (see Figure 2.6) and vanishes at absolute zero (Eqn (2.11)). Neumann and Kopp found that $c_p(T)$ of a chemical compound A_mB_n is nearly the sum of the $c_p(T)$ of its composing chemical elements (Eqn (2.12)).

$$\lim_{T \to 298\text{K}} c_p \approx 25\ \text{J/mol} \cdot \text{K} \quad \text{(Dulong–Petit)} \tag{2.10}$$

$$\lim_{T \to 0} c_p = 0 \tag{2.11}$$

$$c_p^{(\text{A}_m\text{B}_n)}(T) \approx m c_p^{(\text{A})}(T) + n c_p^{(\text{B})}(T) \quad \text{(Neumann–Kopp)} \tag{2.12}$$

For most substances, the function $c_p(T)$ rises monotonously with T; the slope is steep at low T and becomes flatter near room temperature. Anomalies arise near phase transitions (see Section 2.1.3). Distant from phase transitions, the functions $c_p(T)$ are usually smoothly rising, which can be expressed, e.g., by polynoms

$$c_p = A + BT + CT^2 + DT^3 + E/T^2 \tag{2.13}$$

and data for many substances are compiled in commercial databases, on the web (e.g., http://webbook.nist.gov), or in printed reference books [9].

2.1.2.2 Enthalpy

The internal energy U that was mentioned in Section 2.1.2.1 describes the potential and vibrational energy that is contained in the material itself. In a real system, the material has a certain volume V and is exposed to a pressure p. It should be noted that for systems

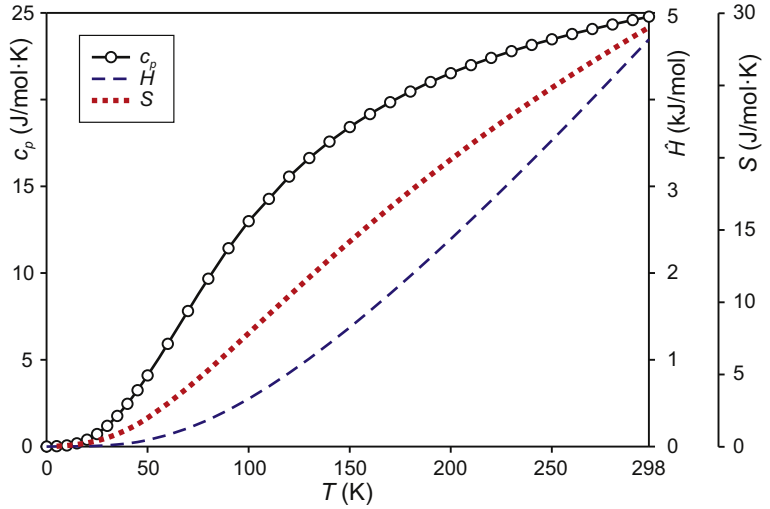


FIGURE 2.6 Measured $c_p(T)$ data for lithium below room temperature [8]. $\widehat{H} = H - H(0\,\mathrm{K})$ is the "real" enthalpy that starts from 0 at absolute zero. The entropy $S(T)$ was calculated with (Eqn (2.21)).

where surface energy can be neglected (phases sufficiently large and phase boundaries with low curvature), p is constant over the whole system, such as T. The quantity

$$H = U + pV \tag{2.14}$$

is called *enthalpy* and is the sum of the internal energy and the amount of work ("volume work") that must be performed to create the phase volume V against the system pressure p. Such as U itself, H is a thermodynamic potential (=state function) because it depends only on the current status of the system, and not on how this status was reached.

It is usually difficult, if not impossible, to measure the total amount of $\widehat{H}$ stored inside a phase, because the energy balance of all contributions to U is often unknown. Contrarily, the enthalpy change of a system from an initial state i to a final state f

$$\Delta H = H_f - H_i = Q \tag{2.15}$$

can usually be measured because it equals the heat Q added to the system provided that p = const., and no other work except volume expansion work is done by the system.

To circumvent the problem of absolute H measurements, it is useful to determine "standard conditions" and a set of basic substances where the H values are determined at these standard conditions. For this purpose, the U.S. National Bureau of Standards defined $T_0 = 25\ ^\circ\mathrm{C} = 298.15$ K and $p_0 = 1 \times 10^5$ Pa = 1 bar as "standard ambient temperature and pressure" [10]. Under these conditions $H_i = H_0 = 0$ (Eqn (2.15)) is defined for every chemical element in the phase state that is stable under these conditions (Figure 2.7). The function $H(T)$ is usually smooth, except at first-order phase transitions (cf. Section 2.1.3) where it jumps by an amount that is called *heat of fusion* (ΔH_f), *heat of vaporization* (ΔH_v), or in general *heat of transition* (ΔH_t). Some first-order

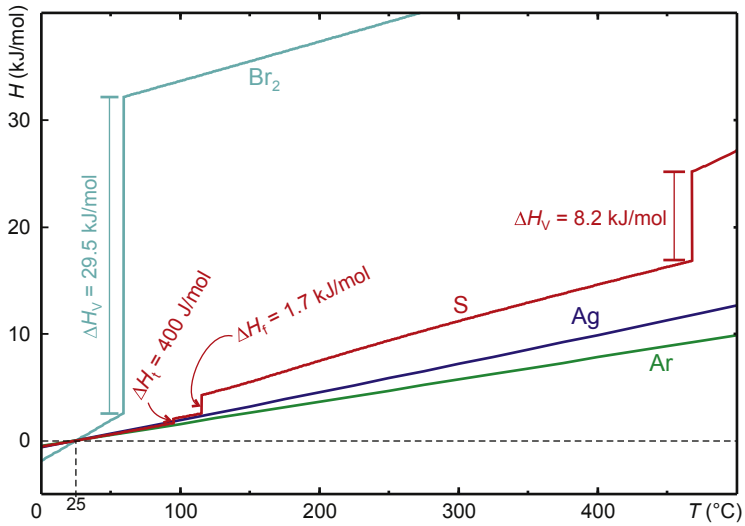


FIGURE 2.7 $H(T)$ functions for four chemical elements at 1 bar. Bromine vaporizes at 59 °C. Sulfur undergoes a monoclinic ⇄ orthorhombic phase transition at 95 °C, melts at 115 °C, and vaporizes at 469 °C. Silver and argon show no transformations in this T range.

transitions for bromine and sulfur are shown in Figure 2.7. In contrast to this, Figure 2.6 shows "absolute" $\widehat{H}(T)$ data for Li metal below room temperature.

For other temperatures, the enthalpy can be calculated by

$$H(T) = H_0 + \int_{T_0}^{T} c_p(T)\mathrm{d}T \tag{2.16}$$

from experimental $c_p(T)$ data. If one starts with $\widehat{H}(0\ \mathrm{K}) = 0$, then Eqn (2.16) represents the area under the $c_p(T)$ curve in Figure 2.6. $c_p(T)$ rises with T^3 only for low $T \lesssim 50$ K and quickly approaches the Dulong–Petit value 25 J/(mol K) (Eqn (2.10)) and is slowly changing then; consequently, $H(T)$ is an almost a linear function for high T.

As mentioned above, H is a thermodynamic potential. Hence, the enthalpy produced or consumed during a chemical reaction does not depend on the path of the reaction (Figure 2.8). Hess's Law can be used also for changes in entropy S and in Gibbs free energy G, which are state functions, such as H. With Hess's Law, the dependence of

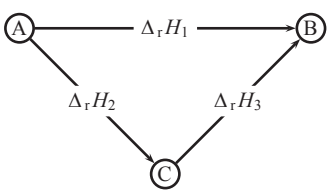


FIGURE 2.8 Hess's Law: the reaction enthalpy does not depend on the path of the reaction: $\Delta_r H_1 = \Delta_r H_2 + \Delta_r H_3$.

reaction enthalpies on T can be calculated, and also transition enthalpies under nonequilibrium conditions.

2.1.2.3 Entropy

The term *entropy* describes another thermodynamic potential that can be defined either statistically or from a thermodynamic viewpoint. The statistical interpretation is related to the Shannon entropy [11], which is used in information theory, and corresponds to the average information density in a system of symbols (or atoms). In the statistical interpretation, entropy S is a measure of "uncertainty" of a given state, that is, a measure of the number of equivalent arrangements of elements setting up the system. This is shown in Figure 2.9(A) for the case of a planar lattice where all possible sites are occupied by atoms. Another equivalent possibility for the atoms does not exist, and the probability for this state is $P = 1$. This is not so in Figure 2.9(B) where two atoms are missing. There are 48 options for selecting the first missing atom, and 47 options remain for the second missing atom, but it is indistinguishable which atom was missing first. This means $48 \times 47/2 = 1128$ equivalent microstates exist that result in the same macrostate "plane lattice" with 48 sites. If P_i is the probability of the i-th microstate, then

$$S = -k_B \sum_i P_i \ln P_i \tag{2.17}$$

is the entropy of the corresponding macrostate.

For the arrangement from Figure 2.9, this is demonstrated in Figure 2.10. The system has N atom sites that are either "faulty" (number f) or occupied (number $N - f$). The number of possible arrangements of faulty sites (=number of microstates) is

$$\frac{1}{P} = \Omega = \frac{N!}{(N-f)!f!} \tag{2.18}$$

and grows drastically with f until a maximum value that is obtained for $f = N/2$.

In closed systems in equilibrium, P is often identical for all microstates and Eqn (2.17) simplifies to

$$S = k_B \ln \Omega \tag{2.19}$$

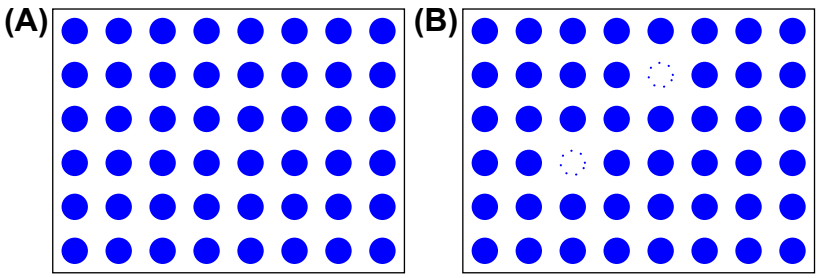


FIGURE 2.9 Statistical interpretation of entropy. (A) One option is possible to distribute $6 \times 8 = 48 = n$ atoms on n lattice sites. (B) This figure shows one of $48 \times 47/2 = 1128$ equivalent options to arrange $f = 2$ faults (e.g., unoccupied sites or vacancies) on 48 lattice sites.

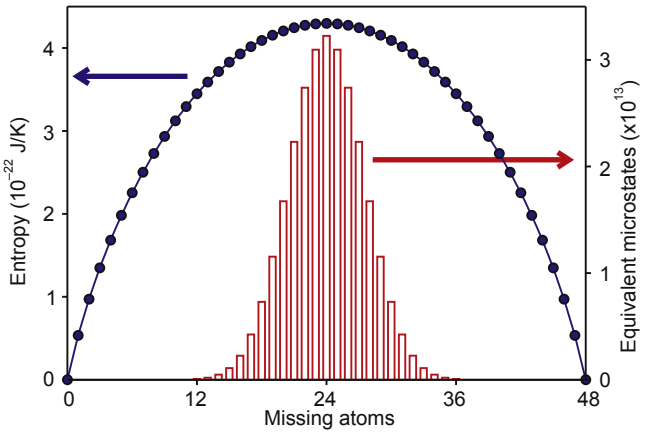


FIGURE 2.10 Derivation of entropy *S* for the ensemble that is shown in Figure 2.8 with different numbers of missing atoms. The number Ω of equivalent microstates is maximum if 50% of the sites are faulty.

where $S(f)$ is shown by the box plot in Figure 2.10. If in a real system with $f > 0$ and at $T > 0$ all faulty sites are initially ordered ($\Omega(t = 0) = 1$), this ordering will drop with time *t*. It will be shown in Section 2.1.2.4 that the growth of *S* is the driving force for increasing the "uncertainty" of the system.

The thermodynamic interpretation of *S* does not depend on the atomistic nature of matter, but can be related to it. Figure 2.11(A) shows an ideal lattice where the atoms are connected by chemical bonds. One can assume that this solid is formed because the creation of every bond reduces the internal energy of the material (compared with sole atoms) by a certain amount *E*. In Figure 2.11(B), some vacancies are introduced that increase the disorder, or the "uncertainty," of the system, and increase its entropy *S*. If for this purpose *n* bonds had to be broken, the energy $Q = n \cdot E$ was used. Assuming that this process is performed under equilibrium conditions, it is reversible, and the relation

$$\Delta S = \frac{Q_{\text{rev}}}{T} \tag{2.20}$$

describes the entropy change of the system.

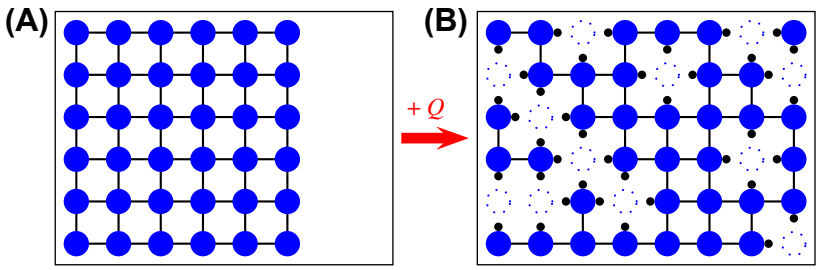


FIGURE 2.11 Thermodynamic interpretation of entropy: In (A) all atoms occupy lattice sites and form ideal bonds. In (B) faults (here vacancies) are introduced and some bonds are "dangling." For this process, a certain amount of energy *Q* was used.

In analogy to Eqn (2.16), the entropy can be derived from the specific heat capacity by

$$S(T) = S_0 + \int_{T_0}^{T} \frac{c_p(T)}{T} \mathrm{d}T \tag{2.21}$$

and one has $S_0(0\ \mathrm{K}) = 0$. Figure 2.6 shows, together with $H(T)$, the function $S(T)$ for Li metal down to absolute zero.

2.1.2.4 Gibbs Free Energy

The enthalpy H introduced in Section 2.1.2.2 is a measure for the amount of energy that is added (or extracted) from a system. This process is, however, usually not reversible: Even if lattice vibrations are considered undamped (contribution U in Eqn (2.14)) and if volume work can be restored (contribution pV), the entropy S of the system is increased upon heating. Once the disorder of a system became larger, it is in general an irreversible process.

It is useful to define the "free energy" (or Gibbs energy)

$$G = H - TS = U + pV - TS \tag{2.22}$$

which is the amount of energy that can reversibly be added or extracted from the system. Such as H and S, G is a thermodynamic potential. This means that for a given state x_i, T, p of a system, G does not depend on how this state was reached.

At elevated T (room temperature or beyond), H changes only weakly (see, e.g., Figure 2.7), and the same holds for S. From Eqn (2.22) it is obvious that $G(T)$ is a function that drops nearly linearly for most phases and systems. Figure 2.12 demonstrates this for the three aggregation states of zinc metal. The solid phase develops most stable binding forces between atoms, this way reducing H and resulting in the most negative $G^{\mathrm{sol}}[0\ ^\circ\mathrm{C}] = -11.42$ kJ/mol. The energy gain by binding is less significant in the liquid

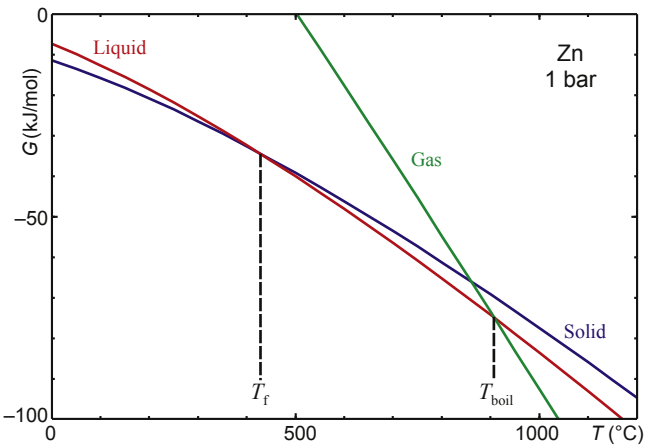


FIGURE 2.12 G as a function of T for the solid, liquid, and gaseous phases of zinc. T_f, fusion point; T_{boil}, boiling point.

($G^{liq}[0\ ^\circ C] = -7.38$ kJ/mol) or even smaller in the gas with much weaker attractive forces between atoms ($G^{gas}[0\ ^\circ C] = +86.42$ kJ/mol). On the other hand, the degree of disorder, and consequently the slope $-S$ of the G functions becomes larger from solid over liquid to gas. The result is that, in this order, these phases have the lowest G, and become stable one after the other with larger T.

For all curves in Figure 2.12, $p = 1$ bar was assumed. The influence of p on S and H is usually small for condensed phases (solids, liquids). Contrary gases show high compressibility, which leads to a high dependency of atomic interactions on pressure. For lower p, $G^{gas}(T)$ shifts to bottom left (Eqn (2.14)), leading to an intersection with $G^{liq}(T)$ at lower T—this means T_{boil} drops with p. If for sufficient small p the intersection is below T_f, the liquid phase is never stable and the substance undergoes sublimation, which is a first-order phase transition from the solid to the gas, and vice versa.

Many chemical elements can react with each other under the formation of compounds. The stability of compounds depends on the binding forces between components, hence on the lattice energy. Typically, compounds with prevalent ionic bonding consist of electropositive cations (often metals) and electronegative anions, and are named, for instance,

- pnictides (compounds with negative N, P, As ions)
- oxides (oxygen) or chalcogenides (compounds with negative S, Se, Te ions)
- halides (compounds with negative F, Cl, Br, I ions).

These anion-forming elements are gases or at least volatile already at moderate temperatures T. The partial molar Gibbs energy (= chemical potential) of an ideal gas can be written as

$$\mu_i = \mu_i^0 + RT \ln p_i \tag{2.23}$$

where μ_i^0 is the Gibbs energy at standard pressure (1 bar) and p_i is the partial pressure of this gas. As a consequence of Eqn (2.23), functions $G(\log p_i)$ are linear for $T =$ const. For different phases, the slopes $G(\log p_i)$ are different, and the resulting intersections are the limits of stability ranges for these phases. This is demonstrated in Figure 2.13 for the system Fe–O_2.

From Eqn (2.23) it is obvious that for $T =$ const. the μ_i are linear functions of the logarithmic vapor pressures p_i. Figure 2.14 shows this for equilibria between iron, oxygen, and sulfur. Only if both p_i are low, metallic iron is stable. If the oxygen pressure becomes larger, oxides FeO, Fe_3O_4, Fe_2O_3 appear in the same order as in Figure 2.13. If only the sulfur pressure becomes larger, the sulfides FeS or FeS_2 are formed. In an intermediate region where oxygen and sulfur have significant p_i, iron(II) sulfate $FeSO_4$ or iron(III) sulfate $Fe_2(SO_4)_3$ are formed. Phase diagrams of this type are valuable, e.g., for predictions, whether interfaces between a substrate and an epitaxial layer are stable. Under the given conditions, FeS_2 ("pyrite") can exist in equilibrium with $FeSO_4$ as well as $Fe_2(SO_4)_3$. This means that from the thermodynamic point of view, epitaxial growth of both sulfates on a pyrite substrate might be feasible. FeS ("pyrrhotite"), in contrast,

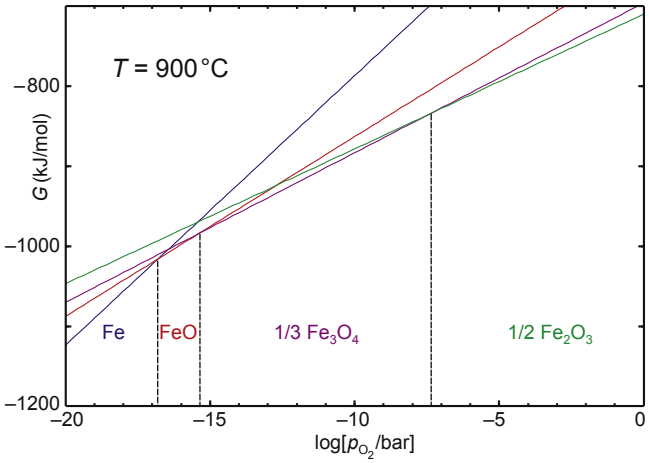


FIGURE 2.13 G as a function of the oxygen partial pressure p_{O_2} for 1 mol iron in equilibrium with $\frac{3}{2}$ mol O_2. From left to right, the solid phases Fe(bcc) → FeO → Fe_3O_4 → Fe_2O_3 have the lowest G and are stable.

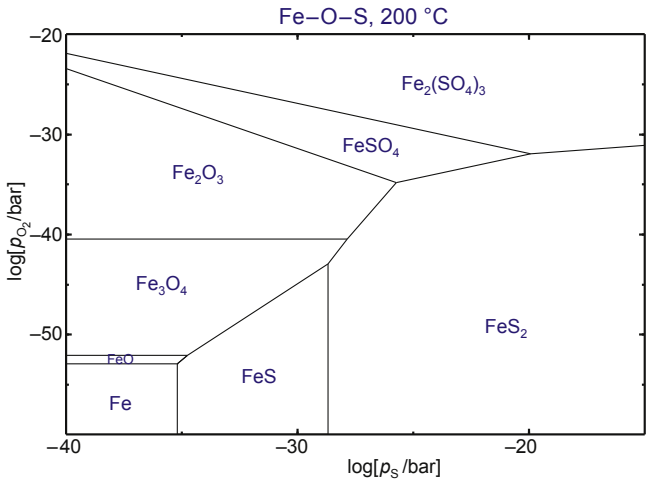


FIGURE 2.14 Predominance diagram for FeO_xS_y phases in coordinates p_S, p_{O_2} (decadic logarithms) for the constant temperature 200 °C.

cannot exist in equilibrium with both sulfates. This means that epitaxy of both sulfates on pyrrhotite, if possible at all, can create only a metastable layer.

2.1.3 Phase Transitions

If water ice is heated under ambient pressure with constant heat flux per time unit $dQ/dt = \dot{Q}$ from low temperatures $T \ll 0$ °C it first increases its temperature until the melting point (fusion point) $T_f = 0$ °C is reached. Then T remains constant for some holding time $t_f^{hold} = \Delta H_f/\dot{Q}$. The heat of fusion $\Delta H_f = 6$ kJ/mol is a thermodynamic property of the substance water. During t_f^{hold} the solid (ice) and the liquid (water) are in

equilibrium; just their mass ratio drops from unity to zero. Upon further heating, the water temperature rises again until $T_v = 100\ ^\circ C$ where water vaporizes. A second holding time $t_v^{hold} = \Delta H_v / \dot{Q}$ appears where liquid water continuously evaporates, and the heat of vaporization $\Delta H_v = 40.87$ kJ/mol is another thermodynamic property of water. The chemical substance water (H_2O) remains unchanged during the whole process, but the physical properties are changing abruptly at T_f and T_v. Melting and vaporization are phase transitions.

At phase transitions, some structural properties of a substance are changing. Usually, these are the positions of atoms, but in some cases only minor changes (electric or magnetic polarization, electronic spins) undergo variations. For the above-mentioned processes of melting and evaporation, the structural changes are very obvious; in other cases, the changes are smaller. Equation (2.15) showed that the heat $Q = \dot{Q} \cdot t$ increases the enthalpy H of the system. In Figure 2.7 the chemical elements Ar and Ag show smooth $H(T)$ functions, and the slopes $\partial H / \partial T = c_p$ are the specific heat capacities. Bromine melts at $-8\ ^\circ C$ and is liquid until $T_v = 59\ ^\circ C$ where a step with height $\Delta H_v = 29.5$ kJ/mol follows. Sulfur shows steps in the $H(T)$ functions not only at T_v and T_f, but additionally a smaller step $\Delta H_t = 400$ J/mol at $T_t = 95\ ^\circ C$. At this temperature, the crystal structure of the solid changes from α-S (low T) to β-S (high T). ΔH_t for this transition is small, because structural similarities are significant (see also Section 2.2.1.1).

Phase transitions can be classified from the thermodynamic as well as from the structural point of view; both schemes are, of course, related to each other. Paul Ehrenfest proposed a classification of phase transitions based on the discontinuity in derivatives of the Gibbs free energy. This classification says that a phase transition is of n-th order if the n-th derivative of G with respect to a system variable such as p or T is discontinuous. Figure 2.12 shows $G(T)$ for the three phases of zinc. The system will always minimize G, hence the lowermost curve is valid for the system. This is replotted in Figure 2.15 for Zn and for Ni (see also Figure 2.5). Only $G(T)$ for Zn shows significant bends at T_f and especially at T_v, consequently the first derivative $G'(T)$ has discontinuities there (arrows from top) and both transitions are of first order. $G(T)$ for Ni is much smoother, and only $G'(T)$ has a small (but hard to recognize) bend near 354 °C. In the second derivative $G''(T)$ a discontinuity appears, which resembles the shape of the Greek letter λ—the transition is of second order. The λ shape is typical for transitions from a ferromagnetic or ferroelectric state where the material has low crystal symmetry to a paramagnetic (paraelectric) state with high symmetry.

Melting and vaporization are always first-order transitions (sometimes called *transitions of the first kind*), and like all first-order transitions characterized by a "latent heat" ΔH_t that must be added to the system at T_t without changing the temperature. Second-order transitions bear no latent heat, and the introduction of heat into the system always increases its temperature. Experimentally, the discrimination between first- and second-order transitions is not straightforward, if ΔH_t is very small (below a few 10 J/mol). The ferro-/paraelectric transition of $BaTiO_3$ is sometimes described as second order, but bears nevertheless a small $\Delta H_t \approx 200$ J/mol [3].

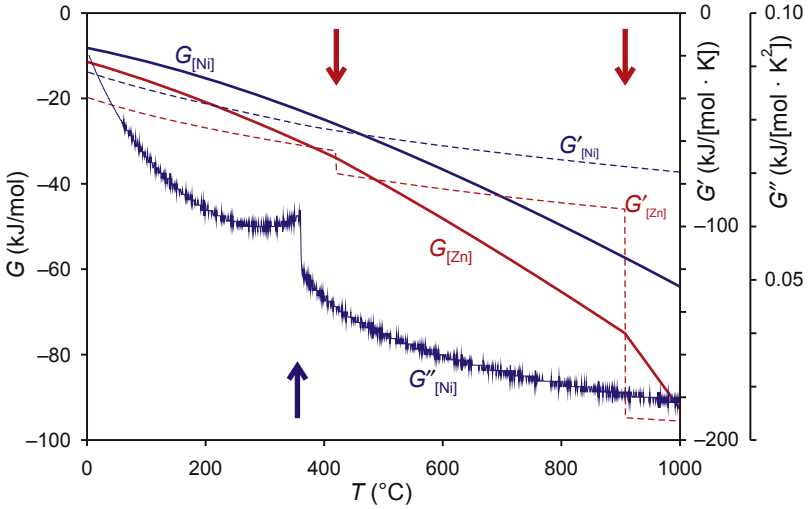


FIGURE 2.15 Solid lines: $G(T)$ for Zn and Ni at 1 bar, dashed lines: $G'(T) = \partial G/\partial T$, scattered line: $G''(T) = \partial^2 G/\partial T^2$ (for Ni only). The arrows from top mark first-order transitions of Zn (melting, evaporation). The arrow from the bottom marks a second-order transition of Ni (Curie temperature T_c where a ferromagnetic material loses intrinsic magnetization).

A thermodynamic approach for second-order transitions was introduced by Lew Landau [12], where an order parameter Q results in an excess Gibbs energy term

$$G = H - TS = \frac{1}{2}A(T - T_c)Q^2 + \frac{1}{4}BQ^4 + \frac{1}{6}CQ^6 \tag{2.24}$$

(A, T_c, B, C—constants), which describes the behavior of measurable physical quantities M such as specific heat capacity, polarization, or atomic positions near second-order phase transitions by expressions of the type

$$M \propto |T_c - T|^{-\alpha} \tag{2.25}$$

where T_c is the transition (Curie) temperature and $\alpha \approx \frac{1}{2}$ is the "critical exponent". For a comprehensive introduction to Landau theory, the reader is referred to Salje [13].

From the structural point of view, second-order transitions often proceed by smooth, continuous shifts of atoms ("displacement type"). Typically, the crystal structure in the high T phase has a high symmetry, and after the shift, at lower T, the crystal symmetry is lowered. Other second-order transitions are of the "order–disorder" type. The β/β' brass phase in Figure 2.16 is an example. At high $T \approx 800$ °C, β-brass has a wide homogeneity range from ca. 44% to ca. 64% Cu, which includes the 1:1 composition CuZn. Under these conditions, Cu and Zn atoms form a body-centered cubic structure (structural type α-Fe, space group $Im\bar{3}m$) where both atom types are distributed statistically across the corners and centers of cube unit cells. It should be noted that corners and centers of the cubic unit cells are symmetrically equivalent.

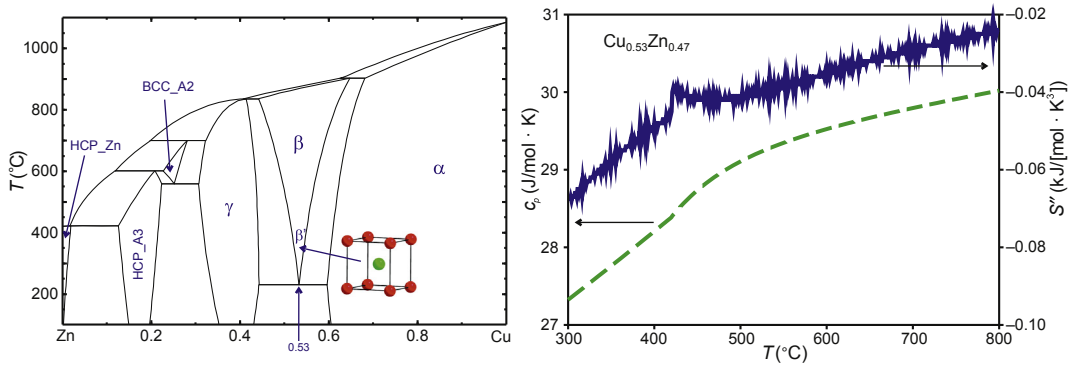


FIGURE 2.16 Left: Binary phase diagram Cu–Zn; the phase fields with one solid phase are labeled. The crystal structure of β' brass is shown in the insert. Right: Specific heat capacity (dashed line) and $S'' = \partial^2 S/\partial T^2$ of the mixed crystal 53% Cu + 47% Zn.

If the temperature becomes lower, it is energetically more favorable if every atom has more opposite type atoms as neighbors. This is reached by placing one atom type preferably on the corner positions, and the other type on the center positions. Ordering does not happen suddenly and is not complete; just the initially fully random distribution gets partially lost if T drops. The totally ordered case with different atom types at the corners or centers, respectively, is shown in the insert of the phase diagram. It corresponds to the CsCl structure type with $Pm\bar{3}m$ space group. The ordered case is called the β'-brass phase.

The phase width of β/β' brass becomes smaller for lower T and degenerates to one point $x = 0.53$ at 530 °C, which implies that ordering cannot be perfect resulting from a slight Cu excess. The right panel of Figure 2.16 shows the $c_p(T)$ function of the $Cu_{0.53}Zn_{0.47}$ alloy, which is calculated [3] between 300 and 800 °C. The broad anomaly around 450 °C results from this "continuous" phase transition [14]. The anomaly results from the change of ordering mentioned above, and this change of ordering is reflected by a minor change of slope for the $S(T)$ function, which becomes obvious in the second derivative S''.

Both panels of Figure 2.16 are calculated for equilibrium conditions. It should be noted that the redistribution of atoms on lattice sites is a diffusion process, and diffusion is time dependent. Correspondingly, the width of the transition depends not only on thermodynamic, but also on kinetic parameters. Generally, second-order and higher order transitions are characterized by a certain transition range from the low T to the high T phase. Often, only the upper limit is characterized by a critical temperature T_c (see Figures 2.15 and 2.17). Besides, the transition proceeds in a smooth way, usually without mechanical disintegration of samples such as crystals during growth. Ferroelectric lithium niobate $LiNbO_3$ is a good example: This material is grown on an industrial scale, and it grows from the melt in a paraelectric $R\bar{3}c$ phase [15]. The Li:Nb ratio is not fixed, and the formula should rather be written as $(1-x)\ Nb_2O_5 \cdot x Li_2O = Li_xNb_{1-x}O_{2.5-2x}$. Figure 2.18 shows the dependence $T_c(x)$ of the Curie temperature

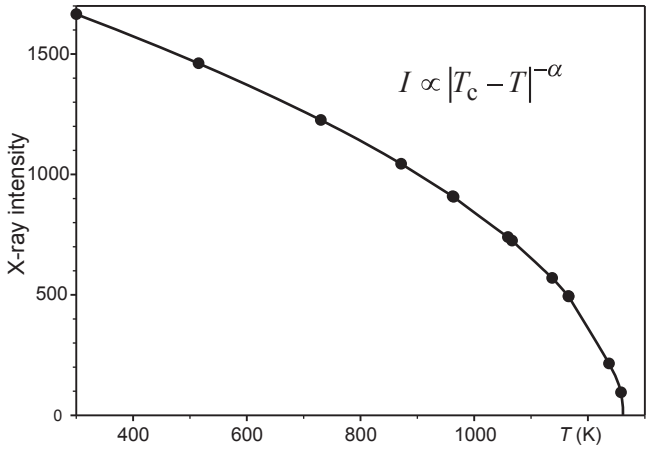


FIGURE 2.17 Measured X-ray intensity data for the $11\bar{2}3$ reflections of calcite [17] fitted with a Landau expression (2.25) gives $T_c = 1254$ °C and $\alpha = 0.51$. See also Ref. [13], p. 11.

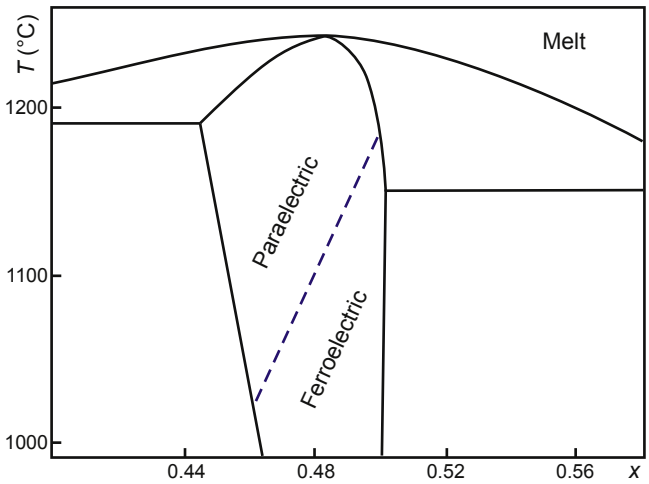


FIGURE 2.18 Dependence of the Curie temperature T_c (dashed line) across the stoichiometry width of lithium niobate $Li_xNb_{1-x}O_{2.5-2x}$ on the lithium concentration x [18]. *Reprinted with permission from AIP Publishing LLC.*

where the second-order transition to the ferroelectric $R3c$ phase occurs. Similar to the β/β' transition of brass (Figure 2.16 right), the para-/ferroelectric transition of lithium niobate is accompanied by a c_p anomaly that can be observed, e.g., by differential scanning calorimetry (DSC) [16].

Only briefly "glass transitions" shall be mentioned here, because the glass state is metastable, hence, cannot be found in (equilibrium) phase diagrams. Nevertheless, $x - T$ regions where glass states are easily formed are sometimes marked. Often, glass formation is connected with immiscibility of two liquid phases below some critical temperature T_m, which usually depends on x. If at least one of these liquids contains a "network builder" such as SiO_2, B_2O_3, or P_2O_5, a glass can be formed that initially

(at high T) has a rubberlike behavior. Upon further cooling, at the glass transition $T_g < T_m$, the glass becomes more brittle. Such behavior can be observed, e.g., in binary systems B_2O_3–A_2O (A = Li, Na, K, Rb, Cs) [19]. The softening of glass if heated above T_g can be measured by thermomechanical analysis (TMA); in addition, it is connected with a c_p anomaly, shifting a DSC curve to the endothermal direction [20].

2.1.4 Calculation of Phase Diagrams

2.1.4.1 Miscibility

If in one specific phase ϕ several components show miscibility, this mixing increases disorder hence the entropy S (Eqn (2.19)). Increasing S, however, reduces the Gibbs free energy G (Eqn (2.22)) and makes the mixed phase thermodynamically more stable, compared to the unmixed (or "mechanically mixed") state (Figure 2.19). If mixing does not occur, one has the simple case of a pure substance with fixed stoichiometry. Then c_p and the derived thermodynamic potentials H, S, G are only functions of T, and data for many substances can be found in the literature.

ϕ is a mixed phase if it has a variable composition x_1, x_2,…x_{C-1} (C—number of components). Assuming that the mixture is ideal, one has

$$G^{\phi} = \underbrace{\sum_{i}^{C} x_i \mu_i^0}_{G^0} + \underbrace{RT \sum_{i}^{C} x_i \ln x_i}_{G^{id}} \tag{2.26}$$

where the first expression G^0 results from the weighed contributions of the pure components (dashed line in Figure 2.19), and the second expression G^{id} from the ideal entropy of mixing. Often, especially if T is very high and chemical interaction forces between components are weak, Eqn (2.26) gives a realistic approximation for real systems too. If energetic (H^{ex}) or entropic (S^{ex}) excess contributions must be taken into account, a term (G^{ex}) must be added to Eqn (2.26).

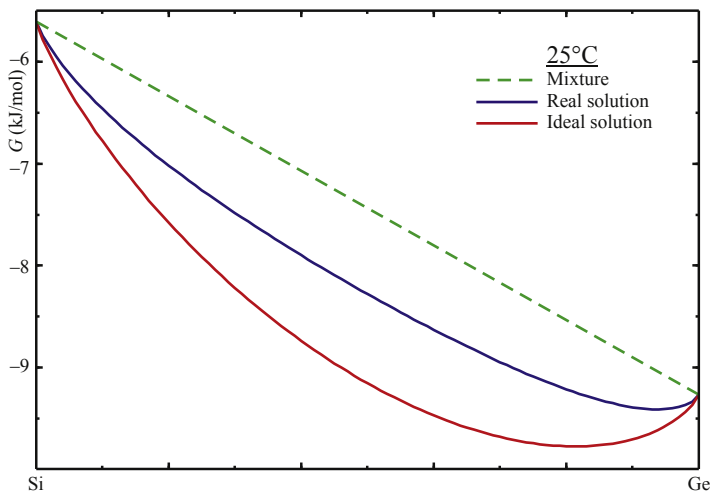


FIGURE 2.19 The functions $G(x)$ for a mechanical mixture of silicon and germanium (dashed line), and for the solid solution $Si_{1-x}Ge_x$ with real data (upper solid curve) and assuming an ideal solid solution (lowest solid curve).

Often this G^{ex} is described by the subregular solution model (Redlich–Kister [21]), which is expressed by

$$G^{ex} = x_A x_B \sum_{j=0}^{N} L_j \, (x_A - x_B)^j \tag{2.27}$$

where x_A and x_B are the molar fractions of components A and B, respectively. (Only the binary case is described here, $x_A + x_B = 1$.) L_j terms represent the interaction coefficients between the basis compounds, and they are often given as a linear function of temperature, e.g., $L_j = L_j^{(0)} + L_j^{(1)} \, T$. Other models, e.g., the "two-sublattice ionic solution" [22], take into account that anions and cations are intermixing only on separate sublattices, and an anion usually cannot jump to a cation site and vice versa.

The G^{ex} of a specific phase can be determined experimentally, e.g., by measuring phase diagrams, electrical potential differences, or by vapor pressure measurements. Such data are available not only in the original literature, but also from databases. Often, the theoretical estimation is possible if interatomic potentials between the constituents (atoms, ions) are known. Then the total energy of a pure phase (e.g., a stoichiometric crystal) can be compared to the total energy of the distorted phase (the mixed crystal). The difference is the energy of mixing. It should be noted that the determination of G^{ex} is often elaborate.

The topology of phase diagrams depends substantially on the extent to which the components can mix in the different phases. The following rules of thumb for gases, liquids, and solids can be given.

Gases: Unless under extreme conditions [23], gases are mixing in arbitrary ratio. Moreover, ideal mixing ($G^{ex} = 0$) can be assumed if the pressure does not approach the critical pressure and if the temperature is not too low. Both requirements are fulfilled for most crystal growth processes, except hydrothermal growth or high-pressure synthesis.

Liquids: Liquids (melts) of different components are miscible if chemically similar. Certainly, this statement is not very definite: It simply means that, if all components belong to one group of substances such as metals, oxides, halides, hydrocarbons, alcohols, sugars, there is a good chance that mixing is possible for all compositions. For technical applications, and especially for crystal growth, many relevant systems fulfill this condition and just one liquid phase is formed. Nevertheless, exceptions exist, and Figure 2.20 shows that, e.g., below ca. 1900 °C CaO–SiO_2 melts with ca. 85% SiO_2 decompose under formation of two liquids with different SiO_2 concentration. In a closer look, phase transition or demixing phenomena in melts seem not uncommon, especially under conditions close to the crystallization of solid phases: Crystal growth of HgTe was observed from demixed melts [24], melts of CdTe undergo a structural transition under formation of associates near T_f [25], and even if $CaCO_3$ crystallizes from aqueous $Ca(HCO_3)_2$ solutions, separation of the liquid was observed in theory and experiment [26,27].

Solids: The condition of chemical similarity for complete miscibility is valid also for crystalline materials. Yet another precondition makes the formation of mixed crystals

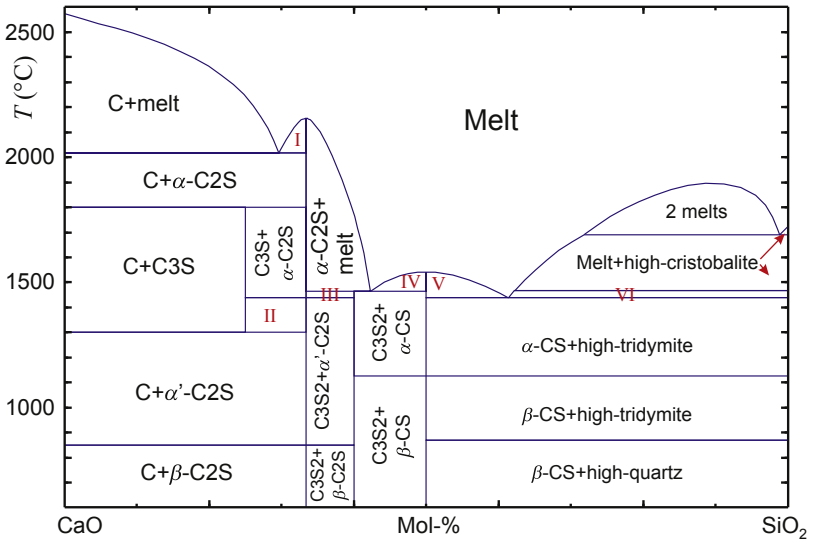


FIGURE 2.20 Phase diagram CaO–SiO_2 with a region of liquid immiscibility for SiO_2 rich melts. The abbreviations are common in the construction material industry; these are the definitions: C = CaO, C2S = Ca_2SiO_4, CS = $CaSiO_3$, C3S2 = $Ca_3Si_2O_7$, C3S = Ca_3SiO_5. All phase fields are labeled, and the roman letters stand for: I = α-C2S + melt, II = C3S + α'-C2S, III = C3S2 + α-C2S, IV = V = CS + melt, VI = high-tridymite + melt.

(=solid solutions) less probable compared to liquids: If two components A and B form a solid solution, their atoms must be able to substitute each other in the corresponding crystals in a complete chain of constitutions. This is possible only if A and B belong to the same crystal structure type. The left panel of Figure 2.21 demonstrates this for silicon and germanium. Both crystallize in the diamond structure type (space group $Fd\bar{3}m$), and the lattice constants of their cubic unit cells are $a_0^{\mathrm{Si}} = 0.543$ nm or $a_0^{\mathrm{Ge}} = 0.566$ nm, respectively. The similarities are so significant that Si and Ge can replace each other, and a continuous row of solid solutions is formed. The lens-shaped binary phase diagram shown there is typical for such "mixed crystal" systems (see Section 2.2.2.1). The right panel of Figure 2.21 shows the binary system lead–tin, both elements follow immediately after Si and Ge in the fourth group of the periodic table, and are similar from the chemical point of view. However, at ambient temperature, Sn crystallizes in a tetragonal structure (space group $I4_1/amd$, lattice constants $a_0^{\mathrm{Sn}} = 0.582$ nm, $c_0^{\mathrm{Sn}} = 0.317$ nm). The lead structure instead is a face-centered cubic lattice (space group $Fm\bar{3}m$, $a_0^{\mathrm{Pb}} = 0.495$ nm). The large Pb atoms can replace only ca. 2% of Sn in its structure; on the other side, the small tin atoms can substitute ca. 20% of Pb. A continuous solubility range cannot exist because continuous transformation between both structures is impossible. Hence, a eutectic system is formed (see Section 2.2.2.2).

As a result of the G^{id} term in Eqn (2.26), the slope of the $G^{\phi}(x)$ functions becomes infinite near the pure components at $x = 0$ or $x = 1$. This means that $G^{\phi}(x)$ drops

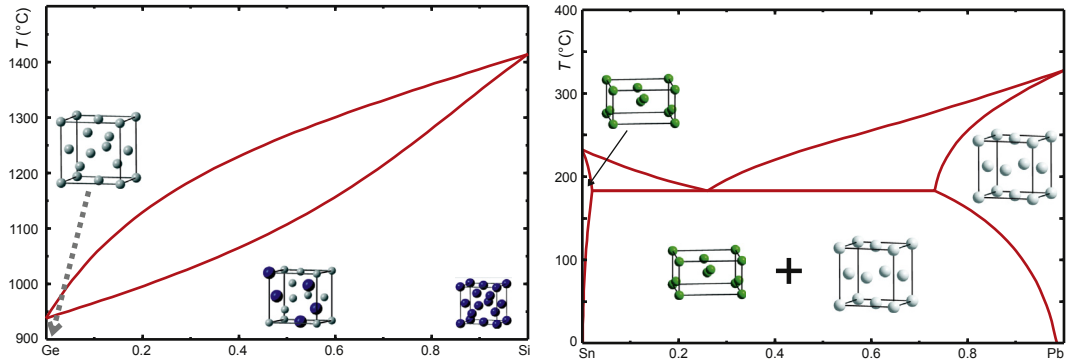


FIGURE 2.21 Left: Phase diagram Si–Ge with formation of a solid solution for $0 < x < 1$ because both end members have identical (diamond) crystal structure and can replace each other (inserts). Right: Phase diagram Pb–Sn with limited mutual solubility at the rims, resulting from different crystal structures of both end members.

considerably for every phase if minor amounts of arbitrary impurities are added. In other words, a minor "rim" solubility always exists. Minor impurity concentrations in the order of parts per million (ppm) or even less may be important, e.g., for doping semiconductors, but are irrelevant for many other applications. All solid phases in Figure 2.20 have a negligible phase width and are represented by vertical lines. Such phases with fixed stoichiometry are called "line compounds" or "daltonides." The opposite is either a mixed crystal spanning all concentrations $0 < x < 1$ of a phase diagram (Figure 2.21 left), or a "berthollide" phase where the composition is a variable within certain limits. Lithium niobate (Figure 2.18) is an example where the concentrations of Li and Nb can shift by a few percent, consequently $Li_xNb_{1-x}O_{2.5-2x}$ shows a finite phase width (="homogeneity range"). In the Cu-Zn system (Figure 2.16(left)), all solid phases are berthollides.

2.1.4.2 *Analytical Expressions*

Often, for practical purposes, binary phase diagrams with two components A, B are used. Then one has $x_B = 1 - x_A = x$ and one concentration value x describes the composition completely. x and temperature T are often used as coordinates for A–B phase diagrams, and some types of such diagrams will be presented in Section 2.2.2. It was stated in Section 2.1.4.1 that the topology of phase diagrams is determined by the mutual miscibility of A and B in different phases ϕ^i. For the three limit cases in Figure 2.22, the shape of phase boundaries can be given analytically.

There and in the following, fixed chemical compositions will be described by Latin letters or combinations of them: A, B, AB, AB_2, A_mB_n. Phases will be described by ϕ', ϕ'',…. If necessary, the chemical composition of a phase can be denoted in brackets. $A(\phi')$ means pure A in the phase state ϕ'. $A_{1-x}B_x(\phi'')$ means a mixture with molar concentration x of B in the phase state ϕ''.

For all diagrams in Figure 2.22, phase fields are separated by lines. Component A undergoes a first-order transition $\phi' \leftrightarrow \phi''$ at T_A, and B an analog transition at T_B. Often

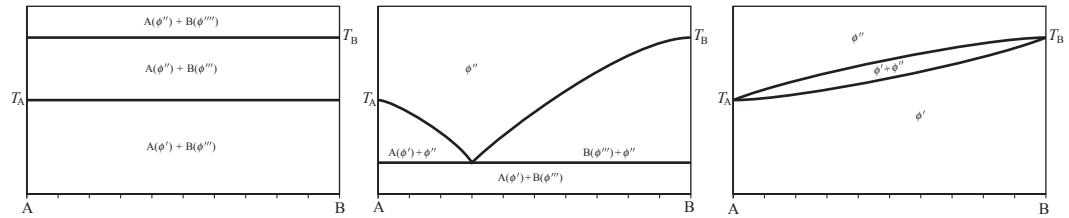


FIGURE 2.22 Three basic topologies of binary phase diagrams where both components undergo a transition from a low T phase ϕ' to a high T phase ϕ''. Left: No mixing in ϕ' and ϕ''. Middle: no mixing in ϕ', but mixing in ϕ''. Right: mixing in ϕ' and ϕ''.

ϕ'' is the liquid (melt), and then the lower phase boundary where only melt(s) appear(s) is named *liquidus*. The maximum temperature at which all phases remain solid is marked by the *solidus* line, and is below the liquidus. The shape of these phase boundaries, however, depends only on the corresponding transition temperatures T_A, T_B and on the enthalpy changes at these temperatures (=heat of transition) Q_A, Q_B. This means that diagrams such as shown in Figure 2.22 can be valid for melting transitions ($\phi'' =$ liquid) as well as for other first-order transitions, e.g., between liquids and gas, or between different solid phases. The following cases can be distinguished:

Figure 2.22 Left: No mixing at all, such with water and mercury. Then $T_A = -39\ ^\circ C$ is the melting point of Hg, and $T_B = 0\ ^\circ C$ is the melting point of H_2O. At low T we have a mixture of pure solid Hg with pure solid H_2O. At high T both components are liquid, but immiscible. Consequently, pure water and pure Hg form separate phases. At intermediate T, solid pure H_2O (ice) floats on pure liquid Hg. For the liquidus and solidus, one has

$$T^{sol} = T_A \tag{2.28}$$

$$T^{liq} = T_B \tag{2.29}$$

if B melts higher, or vice versa.

Figure 2.22 Middle: Mixing only in the high T phase ϕ'', which is often the liquid. The resulting G^{id} (Eqn (2.26)) lowers G for this phase, and makes it more stable. The left panel of Figure 2.1 is an example if one considers only the part between V_2O_5 and V_2MoO_8, or only the part between V_2MoO_8 and MoO_3. Assuming ideal and unlimited mixing of the components in phase ϕ'' and no solubility of B in A(ϕ') or A in B(ϕ'''), respectively, the bent phase lines (liquidus of A or B, respectively) can be described analytically by the expressions

$$x^{liq} = 1 - \exp\left[-\frac{Q_A}{R}\left(\frac{1}{T} - \frac{1}{T_A}\right)\right] \tag{2.30}$$

$$x^{\mathrm{liq}'} = \exp\left[-\frac{Q_\mathrm{B}}{R}\left(\frac{1}{T}-\frac{1}{T_\mathrm{B}}\right)\right] \tag{2.31}$$

where Eqn (2.30) describes the A-rich and Eqn (2.31) describes the B-rich side. Both branches meet at the eutectic point and the third phase boundary (eutectic line) is a horizontal through that point, representing the solidus. This is a *eutectic system* (see Section 2.2.2.2). Figure 2.23 shows a calculated eutectic phase diagram for two different sets Q_A, Q_B. It should be noted that a change of one Q_i influences only the slope on the side of the corresponding component *i*. Consequently, for the case $Q_\mathrm{A} = 10$ kJ/mol, $Q_\mathrm{B} = 20$ kJ/mol one had to elongate in Figure 2.23 the B liquidus (solid line) until it meets the dashed A liquidus, resulting in $x_\mathrm{eut} \approx 0.32$ and $T_\mathrm{eut} \approx 760$ K.

Figure 2.22 Right: If complete mixing of A and B is possible in phases ϕ'' and ϕ', both phases are stabilized by the Gibbs free energy of mixing. One-phase fields appear at very low and very high T. Both one-phase fields are separated by a two-phase field $\phi' + \phi''$. Assuming again ideal behavior in both phases, solidus and liquidus can be expressed analytically

$$x^{\mathrm{sol}} = \frac{\exp\left[-\frac{Q_\mathrm{A}}{R}\left(\frac{1}{T}-\frac{1}{T_\mathrm{A}}\right)\right]-1}{\exp\left[-\frac{Q_\mathrm{A}}{R}\left(\frac{1}{T}-\frac{1}{T_\mathrm{A}}\right)\right]-\exp\left[-\frac{Q_\mathrm{B}}{R}\left(\frac{1}{T}-\frac{1}{T_\mathrm{B}}\right)\right]} \tag{2.32}$$

$$x^{\mathrm{liq}} = \exp\left[-\frac{Q_\mathrm{B}}{R}\left(\frac{1}{T}-\frac{1}{T_\mathrm{B}}\right)\right] \times x^{\mathrm{sol}} \left(T_\mathrm{A} \lessgtr T \lessgtr T_\mathrm{B}\right) \tag{2.33}$$

if heat Q and temperature T of the phase transformation are known for both pure substances. The two-phase lens is symmetrical for $Q_\mathrm{A} = Q_\mathrm{B}$, and it becomes

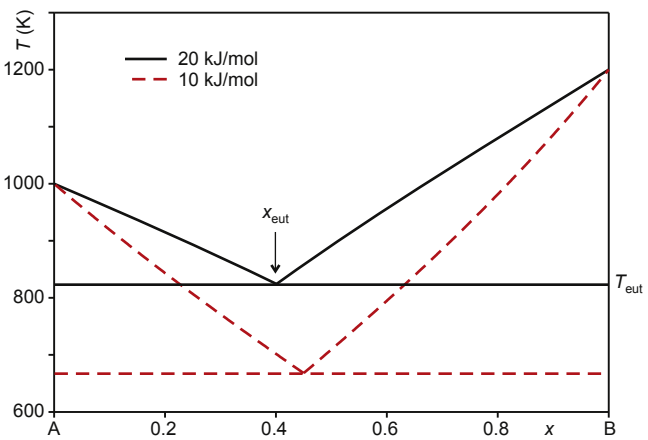


FIGURE 2.23 Calculated hypothetical eutectic phase diagram with $T_\mathrm{A} = 1000$ K, $T_\mathrm{B} = 1200$ K. The solid and dashed lines were calculated with Eqns (2.30) and (2.31), and $Q_\mathrm{A} = Q_\mathrm{B} = 20$ or $Q_\mathrm{A} = Q_\mathrm{B} = 10$ kJ/mol, respectively.

broader for large Q values. Even for ideal mixing, the lens can bend upward or downward if Q_A and Q_B are significantly different. However, both liquidus and solidus are for ideal mixtures monotonous rising or falling, respectively. Local minima may occur only if excess Gibbs free energy of mixing occurs.

2.1.4.3 Minimization of the Gibbs Energy

In Section 2.1.2.4, the Gibbs free energy G was introduced, which has a minimum for every system in thermodynamic equilibrium. The contributions of several phases $\phi'(i = 1\ldots P)$ are additive

$$G = \sum_{i=1}^{P} G_{\phi_i} \tag{2.34}$$

and each phase may consist of several constituents. The number C_ϕ of constituents of a phase ϕ may be unity—then one has a *pure phase*. For pure phases, G_ϕ is basically a function of T, and sometimes other intensive quantities such as p, but not on concentration. Instead, $G(T)$ is given by Eqn (2.22), with $H(T)$ from Eqn (2.16) and $S(T)$ from Eqn (2.21). Alternatively, formation enthalpy ΔH_f, entropy S_0 (at standard conditions) and coefficients for the $c_p(T)$ function (2.13) are a sufficient dataset for the description of pure phases.

Mixed phases bear additional contributions G^{id}, and possibly G^{ex}, to G_ϕ that depend on the concentrations of species present in this phase. Usually, the number of species is much larger than the component number of the system, as already single-phase chemical elements may form several species: gaseous oxygen exists as O, O_2, and O_3. For the whole system, the mass balance for all elementary components must be fulfilled; this means that the number of atoms of one chemical element that are present in all phases may not change.

Knowing the $G(T, x, p,\ldots)$ functions for all possible phases that can be formed from a given starting composition, and taking into account the mass balance for all chemical elements that are present, numerical minimization of G (Eqn (2.34)) allows to calculate the equilibrium state of the system for the given conditions T, p,… [28]. The task of numerical minimization of G is not straightforward, as equation systems with many unknowns (the concentrations of all species in the stable phases) have to be solved. For doing this, commercial software packages such as FactSage [3] or Thermo-Calc [29] as well as some free software, are available. The journal *Calphad* (CALculation of PHAse Diagrams) is a forum for work devoted to such calculation efforts, and an overview of available software can be found there [30]. Commercial software packages are often integrated with databases for thermodynamic data at least for pure substances, and partially also for mixture phases. Free data collections can be found in the literature (e.g., [8–10] or on the internet [31,32]). Many phase diagrams for real systems in this chapter where calculated with FactSage.

2.2 Equilibria Between Condensed Phases

2.2.1 One Component

2.2.1.1 Pressure–Temperature Diagrams

For systems with only one component ($C = 1$), Gibbs phase rule (Eqn (2.5)) gives a sum $F + P = 3$ degrees of freedom + phases. The concentration is fixed to be unity, and often pressure and temperature are used for constructing phase diagrams. Figure 2.3 shows this for sulfur as an example, where the four different phase fields are separated by lines (phase boundaries). Under ambient conditions, sulfur is a soft but brittle yellow solid, with orthorhombic crystal structure (α-S). It melts at 115 °C to a dark liquid, but if cooled quickly (e.g., by pouring into water), an almost rubberlike solid with monoclinic crystal structure is obtained (β-S), which transforms slowly back to α-S under ambient conditions. A detailed investigation reveals that at 1 bar β-S is stable for $95 \leq T\,(^{\circ}\mathrm{C}) \leq 115$.

$F + P = 3$ can be obtained in three ways, as follows, which can be seen from Figure 2.3:

$F = 0$: No degree of freedom means a fixed point with as much as three coexisting phases. This is the case in the three "triple points" T_1 ($\alpha + \beta +$ liq; 117 °C, 283 bar), T_2 ($\beta +$ liq + gas; 115 °C, 3.8×10^{-5} bar), T_3 ($\alpha + \beta +$ gas; 95 °C, 7.6×10^{-6} bar).

$F = 1$: One degree of freedom does exist along lines separating phase fields in the diagram. Along these lines, the neighboring phases are coexisting. The slope of the lines is given by the Clausius–Clapeyron equation

$$\frac{\mathrm{d}p}{\mathrm{d}T} = \frac{\Delta H}{\Delta V_m \cdot T} \tag{2.35}$$

where ΔH is the enthalpy difference between the two neighboring phases. For melting events this is, e.g., the heat of fusion ΔH_f, and first-order phase transformations (see Section 2.1.3) are characterized by the heat of transition ΔH_t. Both sulfur structures are characterized by S_8 rings and differ mainly in the arrangement of these rings. The corresponding phase transition does not rearrange too many bonds, hence $\Delta H_\mathrm{t} = 0.4$ kJ/mol is small for the $\alpha \rightleftarrows \beta$ transition of S. In other cases, where the crystal structure undergoes a complete rearrangement, the enthalpy of solid-state transitions may approach the heat of fusion. $BaCl_2$ undergoes, at 37 K below its melting point, a strong structural transition from orthorhombic to cubic [33] with $\Delta H_\mathrm{t} = 16.9$ kJ/mol, and ΔH_f (at 962 °C) is 16.0 kJ/mol. Generally, one can assume that a few kJ/mol to several 10 kJ/mol are typical values for ΔH to be entered in Eqn (2.35). ΔV_m is the difference between molar volumes in both phases, which are inverse proportional to the mass density ϱ. This is for most solids and liquids in the range $1 \leq \varrho \leq 20$ g/cm^3: Under ambient conditions extreme values are lithium (0.535 g/cm^3) and osmium or iridium with $\varrho \approx 22.6$ g/cm^3. For most transitions between condensed phases, ϱ does not change much,

and ΔV_m is small. Consequently, the slope $\mathrm{d}p/\mathrm{d}T$ is large, which is shown by the almost perpendicular phase boundaries α/β and β/liquid in Figure 2.3. For transitions between condensed phases and gases, however, the molar volume changes remarkably. The large ΔV_m results in significantly smaller slopes of the phase boundaries. If the volume of the condensed phase is neglected in comparison to the volume of the gas phase V_g, ΔV_m in (Eqn (2.35)) can be replaced by the volume of an ideal gas $V = RT/p$. This gives

$$\frac{\mathrm{d}\ln p}{\mathrm{d}T} = \frac{\Delta H}{R \cdot T^2} \tag{2.36}$$

which corresponds to the parabolic behavior of the gas phase boundary in Figure 2.3.

$F = 2$: Two degrees of freedom are available inside the four phase fields of Figure 2.3, and always only one phase is stable there. Two degrees of freedom means that two independent parameters (here p and T) can be changed within the limits of the corresponding phase field independently.

2.2.1.2 Other Fields

T and p are typical intensive parameters that are used as axes, but generally the Gibbs free energy may depend on other fields requiring more terms to be added to Eqn (2.22), e.g.,

$$G = H - TS - V_m \vec{E} \cdot \vec{D} - V_m {}^2\vec{\sigma} : {}^2\vec{\varepsilon} \tag{2.37}$$

where $W_{\mathrm{el}} = \vec{E} \cdot \vec{D}$ is the electric work density of a polarized crystal ($\vec{E}$, $\vec{D}$—electric field or displacement, respectively) and $W_{\mathrm{elast}} = {}^2\vec{\sigma} : {}^2\vec{\varepsilon}$ is the elastic energy stored in a deformed solid (${}^2\vec{\sigma}$—elastic stress, ${}^2\vec{\varepsilon}$—strain). Notations such as ${}^2\vec{\varepsilon}$ express that the corresponding physical property is described by a second rank tensor, a construct having in general 3×3 components [34].

Such alternative contributions to G can be significant, especially for epitaxial layers. Strontium titanate $SrTiO_3$ is an example: The substance crystallizes at high T in a cubic perovskite paraelectric phase, which has no permanent moments. If cooled below the Curie temperature T_C, it lowers crystal symmetry and becomes in a second-order transition (see Section 2.1.3) ferroelectric. T_C depends on the $G(T,\varepsilon)$ functions of the paraelectric and ferroelectric phases, and ε can be manipulated by growing titanate layers on substrates that are almost, but not perfectly, lattice matched. Distorted perovskite substrates such as $REScO_3$ allow "strain engineering" to manipulate the ferroelectric properties of $SrTiO_3$ layers [35].

Gibbs phase rule (Eqn (2.5)) in its classical form relies on the expression $G = H - TS = U + pV - TS$. This means that pressure p and temperature T are independent variables. If under isobar conditions $p = \text{const.}$, the number of independent variables is reduced, the phase rule has to be altered to Eqn (2.6), with 1 instead of 2 as integer constant. Contrarily, a larger number of independent variables such as shown in Eqn (2.37) can increase the integer in Eqn (2.5) from two to larger values. This is

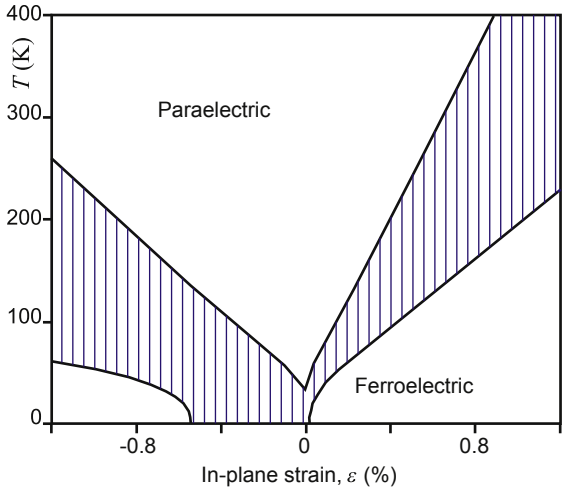


FIGURE 2.24 Phase diagram of $SrTiO_3$ with $\varepsilon = (a_{||} - a_0)/a_0$ as abscissa. $a_{||}$, a_0 is the lattice constant of $SrTiO_3$ in stressed (=epitaxially grown) or free-standing state, respectively. *After Ref. [37], reprinted by permission from Macmillan Publishers Ltd.*

demonstrated in Figure 2.24 for the one-component system $SrTiO_3$, which can be grown on several substrates, causing strain within the thin single crystalline $SrTiO_3$ layer. Only for small ε or high T the material remains in its normal paraelectric cubic phase. For large ε, the material lowers its symmetry completely and becomes ferroelectric. In the intermediate region, paraelectric and ferroelectric states coexist: a phenomenon that is not found in typical one-component phase diagrams with p, T as coordinates (Figure 2.3). For vanadium(IV) oxide VO_2 a triple point between two isolating monoclinic and a metallic rutile type phase was found in coordinates strain ε (or stress $\sigma \propto \varepsilon$, respectively) versus temperature at $T = (65.0 \pm 0.1)\ ^\circ C$, $\varepsilon = 0$ [36].

2.2.2 Two Components

The pressure dependence of phase transitions can be described by the Clausius–Clapeyron Eqn (2.35). Many phase transitions are influenced only to a small extent by the total pressure p of the system. This is the case, e.g., for the melting of solids and for transitions between different crystal structures of one chemical compound where $|\Delta V|$ is usually small. Besides, the variation of x and T can, under usual experimental conditions, be more easily performed than variations of p, which require experiments within pressure cells or sealed ampules. For both reasons, the phase composition of the system depends nearly exclusively on T and on the composition x_i of all components i. As $\sum_i x_i = 1$, one concentration value x is sufficient in the case of binary systems with only two components.

The topology of the binary phase diagram in the $x - T$ plane (often called A–B phase diagram) depends on the miscibility of A and B in the different ϕ^i. Some typical (limit) cases will be presented in the next sections. One should note that transients between these limit cases can be found often, if a limited but non-negligible miscibility

between A and B occurs in one or more phases. The same syntax that was introduced in Section 2.1.4.2 will be used for the description of phases with fixed or variable composition x.

2.2.2.1 Total Miscibility in 2 Phases

This case of total mixing in a high T and low T phase was already described analytically Eqns (2.32) and (2.33). Pure A undergoes at T_A a first-order phase transition, such as melting, from ϕ' to ϕ'', and pure B undergoes at T_B an analog first-order transition from ϕ' to ϕ''. If the necessary but not sufficient conditions

1. identical space symmetry groups (requiring at least identical point symmetry groups, which again requires at least identical crystal system)
2. similar lattice parameters a, b, c (should not be different by more than $\approx 15\%$)
3. similar nature of chemical bonds (ionic, covalent, van der Waals, or metallic)

are fulfilled for both phases, each of them can adopt any composition $0 \leq x \leq 1$ (complete miscibility). Accordingly, for low T, only one phase ϕ' is stable for all x. This means that this combination of system variables corresponds to a one-phase room (or one-phase field) in the diagram. For high T, only ϕ'' is stable, resulting in another one-phase field. For intermediate T both phases ϕ' and ϕ'' may exist in equilibrium that is represented by a two-phase field that typically has the shape of a lens. The two-phase field $\phi' + \phi''$ spreads from the composition A to the composition B in the binary phase diagram (Figure 2.25). It should be noted that the first-order transition happening at one point on the T scale for the pure substances A and B is spread over a T range which is spanned by the lines limiting the two-phase field for intermediate compositions. This T range becomes broader for larger heats of transformation (Figure 2.26).

Often, ϕ'' is simply the melt (liquidus). In this case, the first condition for miscibility is usually fulfilled, as liquids (except "liquid crystals") are isotrope. Isotropy is described by the limit point symmetry group (Curie group) $\infty \infty m$, which is the highest symmetry group at all. Condition 2 does not apply to liquids; hence, only condition 3 decides

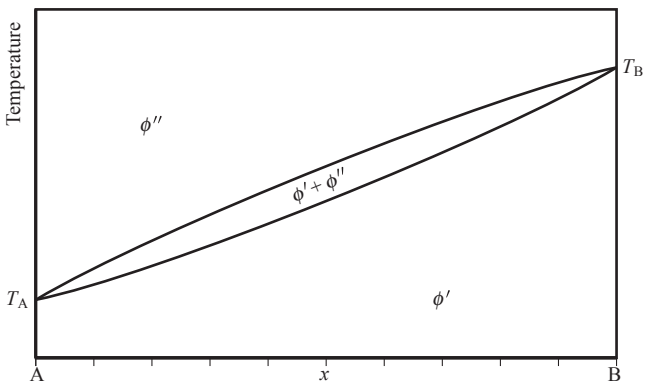


FIGURE 2.25 Binary phase diagram with ideal mixing both (low T and high T) phases ϕ' and ϕ''.

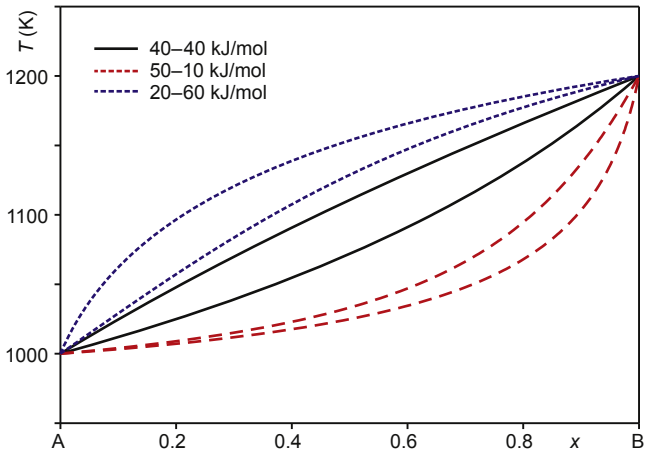


FIGURE 2.26 Calculated hypothetical mixed crystal phase diagram with $T_A = 1000$ K, $T_B = 1200$ K. The three sets of solidus and liquidus lines were calculated with Eqns (2.32) and (2.33) for the $Q_A - Q_B$ pairs given in the legend.

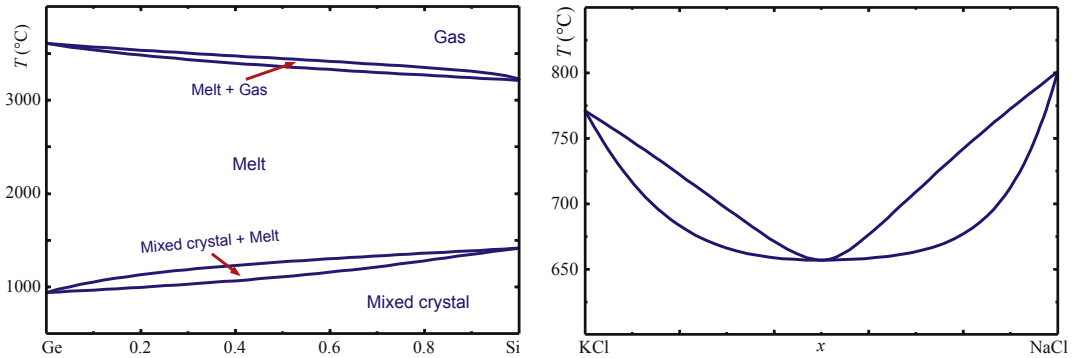


FIGURE 2.27 Left: Binary phase diagram Si–Ge for an extended T range (compare to Figure 2.20) where one-phase fields "mixed crystal," "melt," "gas," are separated by two-phase fields. Right: The binary phase diagram NaCl–KCl shows an azeotropic point where liquidus and solidus are coincident.

whether or not the phases ϕ'' of both components A or B can mix. In such cases where ϕ'' is the melt, the lower boundary line of the two-phase field $A_{1-x}B_x(sol) + melt$ is called solidus and the upper boundary line is called liquidus. Like melting, evaporation is a first-order transition and liquid–vapor equilibria are similar to solid–liquid equilibria. This is demonstrated in the left panel of Figure 2.27. Similar topologies with subsequent one-phase fields can also be observed for systems with first-order transitions between different solid phases. The zirconium–hafnium system is an example in which both components undergo transitions (h.c.p) $\rightleftarrows$ (b.c.c) $\rightleftarrows$ melt. The abbreviations in brackets stand for the identical hexagonal closed packed or body centered cubic crystal structures of both metals, respectively. For zirconium and hafnium, the solid-state transition appears at 865 or 1950 °C and melting at 1860 or 2230 °C, and complete miscibility is

observed in all phases [38]. Consequently, a narrow two-phase field appears between h.c.p. and b.c.c., and a broader one ($\Delta H_f \approx 5\Delta H_t$for both metals) between b.c.c. and melt.

If G^{ex} is small for both phases with total miscibility, the phase boundaries (e.g., liquidus and solidus; see Figure 2.26) are monotonous rising or falling. The right panel of Figure 2.27 shows in contrast the case NaCl–KCl, where the melt extends for intermediate x below the T_f of the components, forming a local minimum. If such minimum occurs, both phase boundaries must meet there in one common "azeotropic" point, where ϕ' and ϕ'' have identical composition x_{az}. This is remarkable because x_{az} is the only composition, except the pure components, for which crystal growth from the melt is possible without segregation (cf. Section 2.2.2.3). The system CaF_2 ($T_f = 1418$ °C)–SrF_2 ($T_f = 1477$ °C) has an azeotropic point at $x_{az} \approx 0.418$, $T_{az} = 1374$ °C, where homogeneous single crystals with 30 mm diameter and 50 mm length could be grown by the Czochralski method [39]. Principally azeotropic points could also be common maxima of liquidus and solidus, but this case is not very realistic: It requires a relative stabilization of the solid with respect to the liquid (negative G^{ex}). This, however, results rather in the formation of an intermediate compound. In the system Cu–Au, where both components have face-centered cubic structures, solid solutions exist for all concentrations with an azeotrope near 60% gold. Below ca. 400 °C, the interaction between Cu and Au atoms becomes so strong that their distribution on lattice sites is partially ordered. Depending on the Cu/Au ratio, compounds with approximate compositions Cu_3Au, CuAu, or $CuAu_3$ are formed, and all of them show some flexibility of composition around the ideal values 25, 50, or 75% Au.

Other examples for intermediate compounds are $Sr_xBa_{1-x}Nb_2O_6$ (SBN) with a liquidus/solidus maximum at $x = 0.61$ [40–42] or $Ca_xBa_{1-x}Nb_2O_6$ (CBN) with a liquidus/solidus maximum at $x = 0.281$ (see Figure 2.28) [43]. Sometimes, SBN and CBN are called mixed crystals, but this denomination might cause errors because it suggests mixing from the end members $EANb_2O_6$ (EA = Ca, Sr, Ba), which is wrong. Instead, both systems are pseudobinary with an intermediate congruently melting compound, and eutectics on both sides of this compound (see Section 2.2.2.4). It is just remarkable that the homogeneity width of these compound is rather large: from ca. 0.2 to 0.8 for SBN, and from ca. 0.15 to 0.4 for CBN.

The Gibbs energy gain from mixing becomes smaller for low T as the second term of Eqn (2.26) vanishes. Consequently, mixed phases do not exist in equilibrium at very low T. Of course, demixing of solid solutions (in contrast to liquid solutions) requires diffusion steps that need time, and are consequently often hindered kinetically. Not always demixing results in the formation of intermediate compounds, like in the Cu–Au system above mentioned. No intermediate compounds exist between antimony and bismuth. Both are crystallizing in identical structures belonging to the rhombohedral $R\bar{3}c$ space group with lattice constants $a_{0,rh}^{Sb} = 0.45067$ nm and $a_{0,rh}^{Bi} = 0.47458$ nm [44]. This difference of ca. 5% is rather small, and allows the formation of solid solutions $Sb_{1-x}Bi_x$ for $0 \leq x \leq 1$ at sufficiently high temperature. The binary phase diagram in Figure 2.29

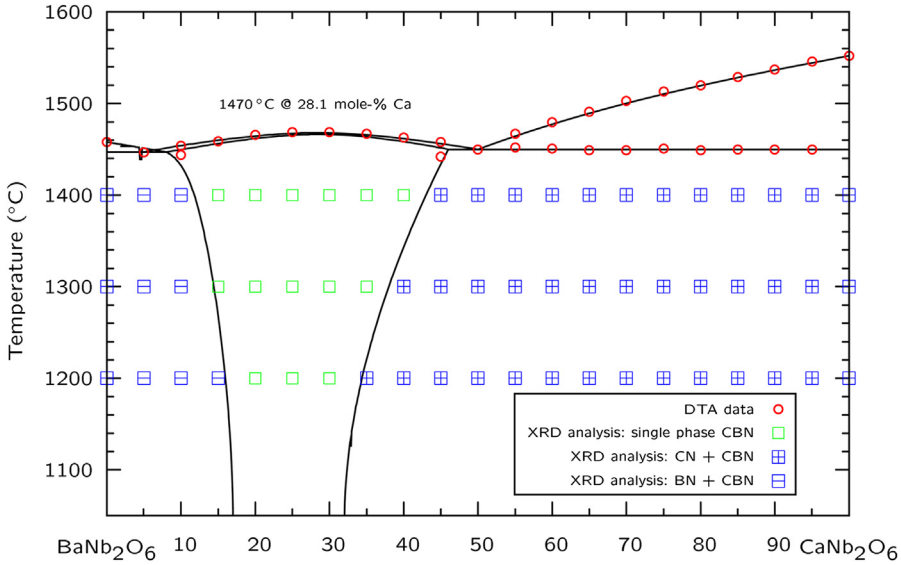


FIGURE 2.28 Pseudobinary phase diagram $CaNb_2O_6$–$BaNb_2O_6$ with congruently melting intermediate $(Ca,Ba)Nb_2O_6$ ("CBN"). The CBN phase field is single phase, in contrast to the neighboring fields. *Reprinted with permission from Elsevier [43].*

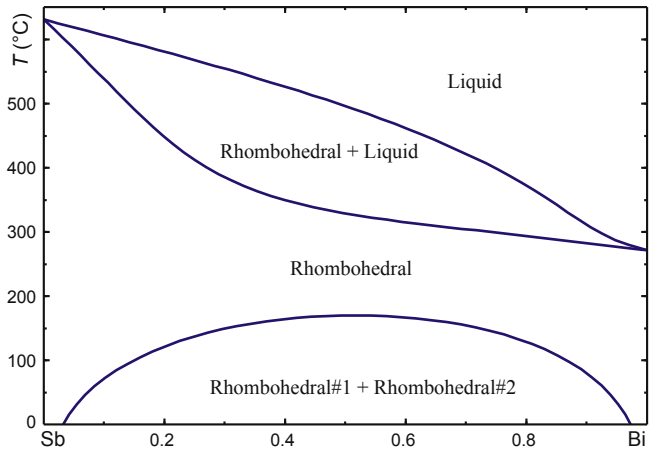


FIGURE 2.29 Phase diagram bismuth–antimony with complete miscibility at high T in the rhombohedral A7 structure. Below the miscibility gap, decomposition to a Sb-rich and a Bi-rich phase occurs, both with identical A7 crystal structure like that above the gap.

shows, however, that near $x = 0.5$, such mixed crystals become unstable at $T \lesssim 170\,^{\circ}\mathrm{C}$. Then the initially homogeneous solid solution (one phase) becomes unstable and decomposes under conservation of the crystal structure to volume elements, which are enriched in Sb or Bi, respectively (two phases). It should be mentioned that such

miscibility gaps, even if theoretically expected for all mixed crystal systems, are practically not observed very often for kinetic reasons.

2.2.2.2 Eutectics and Eutectoids

A eutectic phase diagram is formed if the components are immiscible in their solid phases, but exhibit complete miscibility in their molten (liquid) phases. This case was described analytically by Eqns (2.30) and (2.31). If a melt with composition x_{eut} in the hypothetical phase diagram Figure 2.23 is cooled, then it crystallizes completely at x_{eut} under formation of the solid phases A and B. In other words, a *eutectic reaction* liquid → A (sol) + B (sol) takes place, where one phase decomposes and forms two other different phases. The crystallization of two phases at the same time close together results typically in a fine-grained, interpenetrated structure. Such eutectic solids sometimes show metamaterial properties that might differ considerably from the properties of its constituents [45]. If in Figure 2.23 the initial composition of the cooling melt deviates from x_{eut}, the crystallization of a solid phase starts already at another $T_{liq} > T_{eut}$ and continues until T_{eut}. The first crystallizing phase left from the eutectic composition is solid A, and right from x_{eut} solid B. All phase fields are labeled correspondingly in Figure 2.30.

It is not a requirement that the high T phase is liquid. Figure 2.30 shows the more general case where, during heating, the component A undergoes at T_A the transformation $\phi' \rightarrow \phi'''$, and B at T_B a transformation $\phi'' \rightarrow \phi'''$. This means that at low T, A and B show a different phase state, and are therefore immiscible, but have identical phase states ϕ''' with unlimited miscibility at high T. If ϕ''' is a solid phase, then the composition x_{eut} undergoes at T_{eut} a eutectoid reaction.

Figure 2.31 shows as an example of high technical relevance the iron-rich part of the Fe–C system. It should be noted that this system is drawn here for true equilibrium. Under the technical conditions of iron and steel metallurgy, often cementite phase Fe_3C appears [46], which is not an equilibrium phase in the Fe–C system (but see Figure 2.32

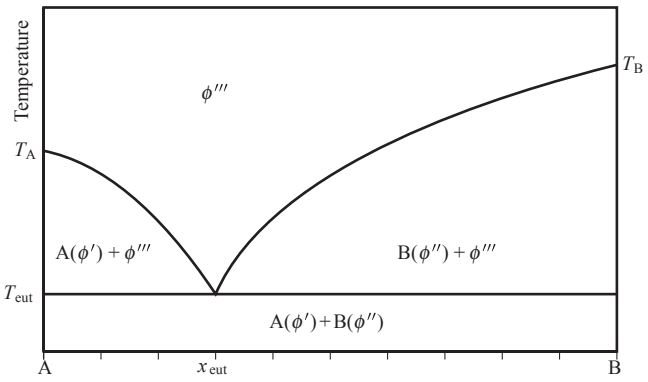


FIGURE 2.30 Binary phase diagram with ideal mixing in the high T phase ϕ''' only. Below T_A pure component A exists in the phase ϕ' and below T_B pure component B exists in the phase ϕ''. ϕ' and ϕ'' are immiscible.

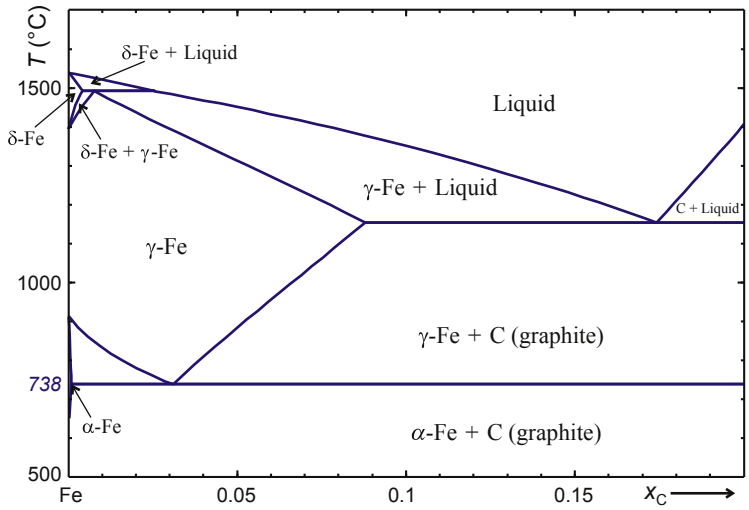


FIGURE 2.31 The iron-rich part ($x_C \leq 0.2$) of the iron–carbon phase diagram. Pure iron undergoes transitions α-Fe → γ-Fe → δ-Fe → melt. The maximum solubility of carbon in the solid phases is (in this order) 0.001, 0.09, 0.004. Single-phase γ-Fe with $x_{Fe} = 0.03$ is stable down to 738 °C, where it undergoes eutectoid decomposition to α-Fe and graphite.

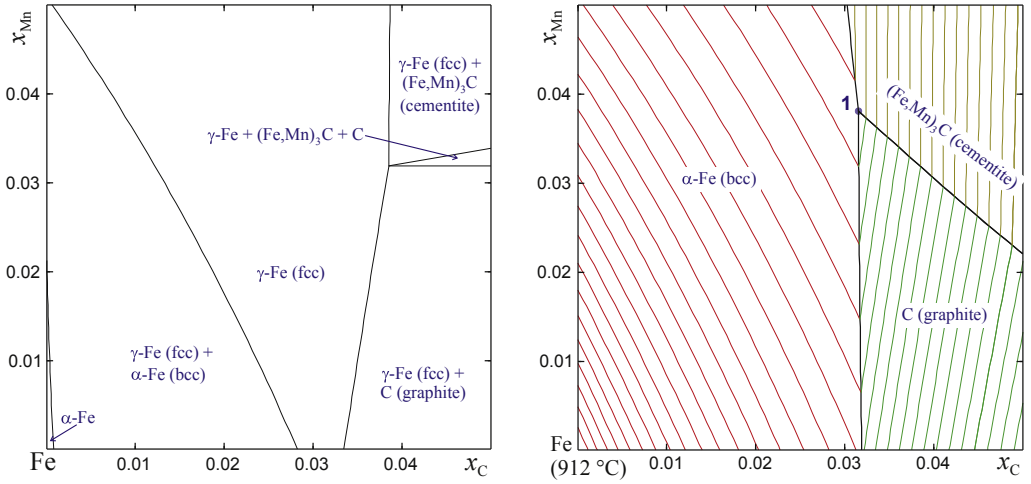


FIGURE 2.32 Left: Isothermal section $T = 750$ °C through the ternary system Fe–Mn–C near the Fe-rich corner (bottom left) up to 5 mol-% Mn and C. Only for minor concentrations of the additives ($x_C \lesssim 0.001$; $x_{Mn} \lesssim 0.02$) iron remains single-phase α-Fe; otherwise, a large range of single-phase γ-Fe exists. Right: Polythermal projection of the same concentration range on the γ-Fe (f.c.c.) phase field with 10 K isotherms. The four phase invariant point "1" is at $x = 0.032$, $y = 0.038$, $T = 693$ °C.

for Fe–Mn–C). Nevertheless, the appearance of Fe_3C results only in minor amendments to the phase diagram part that is shown in Figure 2.31: the rightmost liquidus becomes flatter—it is then the liquidus of Fe_3C instead of C, and the lowest stability limit of γ-Fe drops from 738 °C by ca. 12 K. Pure Fe has below 911 °C a body-centered cubic structure

(such as tungsten) and is comparably hard. This structure can dissolve only minor amounts of carbon. For $911 < T\,(^\circ\mathrm{C}) < 1395$, γ-Fe is stable, which has a face-centered cubic structure (such as copper) and can dissolve large amounts $\lesssim 9\%$ C. Dissolved carbon stabilizes γ-Fe down to the eutectoid temperature $T_{\mathrm{eutectoid}} = 738\ ^\circ\mathrm{C}$ at $x_{\mathrm{eutectoid}} = 0.03$. If such $Fe_{0.97}C_{0.03}$ is cooled from the γ phase field below $T_{\mathrm{eutectoid}}$, it decomposes to the α-Fe phase and graphite. This transformation from one phase to two phases has analogy with a eutectic, but proceeds only with solid phases. It is therefore called a eutectoid.

2.2.2.3 Segregation and Lever Rule

Many solidification and crystal growth processes are performed by cooling melts or melt solutions. Such crystallization is often accompanied by *segregation*. This means that a solid (e.g., the just-formed part of the crystal) and a liquid (e.g., the rest of the melt) of different composition are in equilibrium. During heating, the same observation is often called *incongruent melting*, as melting solid and created liquid have no identical composition. The appearance of segregation is very different in systems with or without mixed crystals. The left panel of Figure 2.33 shows the case of a simple mixed crystal system. If a melt with composition x_0 is cooled, the first solid is formed at the temperature T' where the liquidus touches x_0. The system is in equilibrium, and consequently every part of it is at T' now (isothermal conditions). The composition of the solid phase is shown by the solidus, and this curve gives at T' the composition $x'_{\mathrm{sol}} \neq x_0$. The dashed horizontal line at T' connecting x_0 (liquid) and x'_{sol}, which are in equilibrium, is called the *tie-line*. During further slow cooling, the system reaches T'', and now a melt x''_{liq} is in equilibrium with a solid x''_{sol}. Reaching such state, however, is an extremely time-consuming process, as all solid material that was crystallized before with $x > x''_{\mathrm{sol}}$ had to be transformed by solid-state diffusion to the new composition. Practically, this seldom happens, even during geological times: The mineral olivine is a crystal mixed between forsterite Mg_2SiO_4 and fayalite Fe_2SiO_4, and shows locally different Fe/Mg ratio in artificial as well as natural crystals [47]. Theoretically, under perfect equilibrium, the

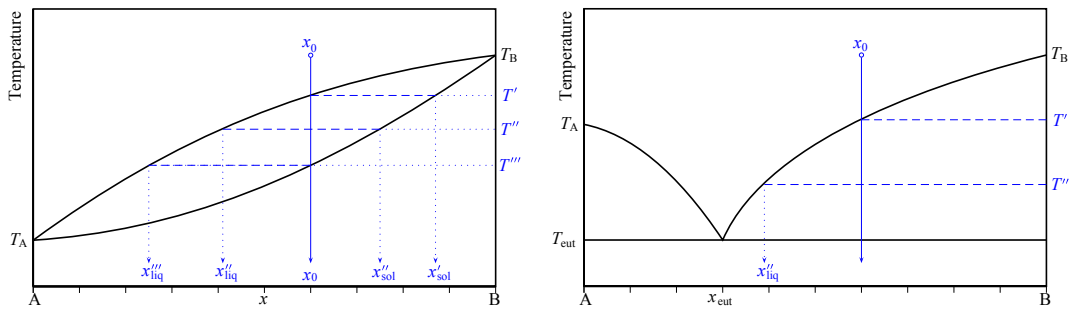


FIGURE 2.33 Crystallization of a starting composition x_0: Left: Crystallization processes in a mixed crystal system result in a solid concentration $x_{\mathrm{sol}}(T)$ that depends on the current temperature. Right: In a eutectic system, the solid body has a fixed concentration (here B with $x = 1$).

crystallization process is finished at T'''' when the two-phase field is left. As material may not disappear, the resulting crystal had the initial composition x_0 then.

Practically, the initially crystallized part will often be too far from the residual melt to reach equilibration. Depending on the geometrical conditions, the first fraction crystallized with x'_{sol} may then be surrounded by subsequently crystallized material with smaller B-composition ("gradient crystal"). In crystal-pulling processes, such as Czochralski, the B-rich material crystallizes first close to the seed, and is closer to the tail followed by material with lower B concentration. In systems such as shown in Figure 2.33 left, the higher melting component will always be enriched in the first crystallized fractions. Often, a segregation coefficient

$$k_0 = \frac{x_{sol}}{x_{liq}} \approx \text{const.} \tag{2.38}$$

is defined that approximates the solidus and liquidus to be linear near the pure components. This case was discussed first by Gulliver [48], and results in a dependence on the position given by

$$x_{sol}\left(g\right) = k_{eff} x_0 (1-g)^{k_{eff}-1} \tag{2.39}$$

where $0 < g < 1$ is the crystallized part of the melt with initial composition x_0. The effective distribution coefficient k_{eff} depends on k_0 (for crystallization that is not too quick, one often has $k_{eff} \approx k_0$) and several parameters of the crystallization process itself [49].

Segregation appears always in mixed crystal systems, except at azeotropic points (see Figure 2.27 right) and often significantly impedes crystal growth, because large volumes with constant composition cannot then be easily grown. If the concentration, hence the properties of the mixed crystal, depend too strongly on position, even a stable crystal growth process may become impractical.

Also in systems where the solid phase has a fixed composition, such as the eutectic system in Figure 2.33 right, segregation can occur. If a melt x_0 is cooled, at T' the pure solid B ($x = 1$) crystallizes. The crystallization of B (sol) continues until T_{eut} is reached where the whole system becomes solid. Such behavior may be beneficial if for some reason (too high volatility, phase transition, or another destructive process below T_B, technical restrictions for very high T_B) the crystal B (sol) cannot be grown from a pure B melt. This is the basis for melt solution growth processes, such as TSSG [50].

For one-phase fields, a point (x, T) gives the composition and temperature of this phase at this specific point. From the tie-line constructions mentioned above, it is clear that this is not the case for two-phase fields. There, the tie-line crossing this point connects two phases with different compositions that are in equilibrium at this T. Not only can both compositions, but also the shares of both phases, be read directly from the phase diagram, following the *lever rule*. According to this rule, the quantities of both equilibrium phases are indirectly proportional to the "levers" that are spanned from the given composition to the corresponding phase boundaries. For instance, in Figure 2.33

right, one has at T'' solid B ($x = 1$) in equilibrium with a melt composition x''_{liq}. The quantities are

$$\text{melt} : \frac{1 - x_0}{1 - x''_{\text{liq}}} \tag{2.40}$$

$$\text{B(sol)} : \frac{x_0 - x''_{\text{liq}}}{1 - x''_{\text{liq}}} \ . \tag{2.41}$$

With further lowering of T, x''_{liq} finally approaches x_{eut}, and this gives the maximum yield (expressed as share of the starting material) that can be reached with crystal growth processes from melt solution with

$$Y_{\max} = \frac{x_0 - x_{\text{eut}}}{x_{\text{sol}} - x_{\text{eut}}} \tag{2.42}$$

where x_{sol} is the composition of the phase which has to be grown (in Figure 2.33 right $x = 1$), and x_0 is the initial composition of the melt.

2.2.2.4 *Intermediate Compounds*

Sometimes the components of a thermodynamic system can interact so strongly that intermediate compounds are formed. This is usually the case if one component easily creates cations (such as most metals), and the other easily creates anions (such as halogens, oxygen, or sulfur). Then simple salts, oxides, or sulfides are built. Also, however, systems set up from comparably similar components can form intermediate compounds if a sufficiently large Gibbs free energy reduction can be reached; see, e.g., the CaO–SiO_2 system in Figure 2.20.

Figure 2.34 shows two similar pseudobinary systems LiF–AF (A = Cs, Rb), and both of them contain an intermediate compound $LiAF_2$. It should be noted that for A = K, Na the

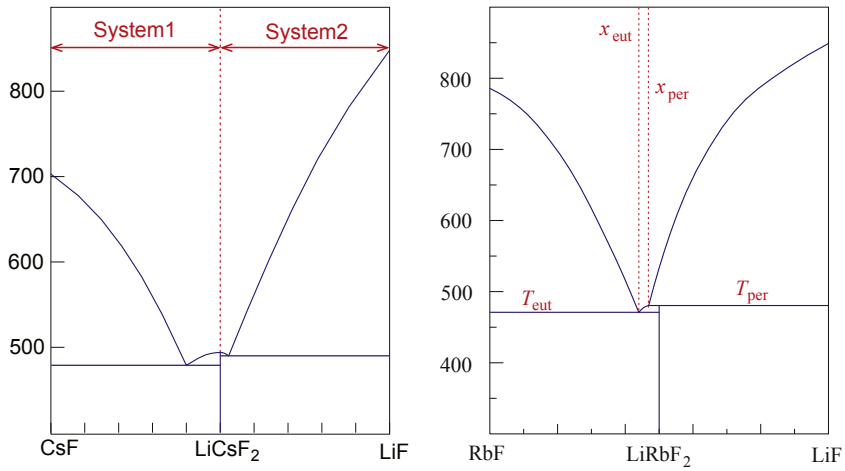


FIGURE 2.34 Both systems LiF–CsF (left) and LiF–RbF (right) contain an intermediate 1:1 compound. $LiCsF_2$ has a larger formation enthalpy, and melts congruently. This is the precondition for constructing independent partial systems.

systems form simple eutectics (for Na with some rim solubility on the NaF side), but no intermediate compounds. Obviously the larger radius difference between the A ions, in octahedral coordination Li (90), Na (116), K (152), Rb (166) Cs (181 pm), makes the Gibbs energy gain upon compound formation stronger.

From the lever rule introduced in Section 2.2.2.3, it is obvious that every melt composition from the left half of the LiF–CsF system (Figure 2.34 left, $0 \leq x_{\mathrm{LiF}} \leq 0.5$) forms during crystallization only CsF and/or $LiCsF_2$. Otherwise, from melts $0.5 \leq x_{\mathrm{LiF}} \leq 1$, only $LiCsF_2$ and/or LiF are crystallizing. This means that both halves of the LiF–CsF system are independent, and consequently form subsystems. But this is not so for the LiF–RbF system shown in the right panel of Figure 2.34: The intermediate $LiRbF_2$ at $x_{\mathrm{LiF}} = 0.5$ undergoes at T_{per} peritectic decomposition (see Section 2.2.2.5) and the lever rule shows that from melts $x_{\mathrm{per}} < x \leq 1$, LiF (sol) crystallizes first. But x_{per} is <0.5, in the left half of the system! This means that some compositions in the left half of the system have LiF from the right half as first crystallizing solid. In other words, it is impossible to define independent partial systems with peritectically melting compounds as (rim) components. Otherwise, congruently melting intermediate compounds in binary systems, such as $LiCsF_2$ in the LiF–CsF system, can be used as rim components for partial systems.

This is sometimes helpful if a complete system is on the one side sophisticated and difficult to measure, but on the other side in some regions not really interesting for a specific purpose. The CaO–SiO_2 system in Figure 2.20 is a good example: For CaO-rich compositions, the liquidus goes up beyond 2500 °C, which is considerably higher than the T limit of typical devices for thermal analysis such as differential thermal analysis (DTA). The binary phase diagram shows the congruently melting phases C2S = Ca_2SiO_4 at 33% SiO_2 (belite) and CS = $CaSiO_3$ at 50% SiO_2 (wollastonite). Consequently, C2S and CS may be used as end members of a partial system, and it becomes clear that CaO–SiO_2 mixtures between 33.33% and 50% SiO_2 may contain in equilibrium only C2S, CS, and the intermediate C3S2 ($Ca_3Si_2O_7$, rankinite). For some of them, high- and low T phases (α,α',β) do exist, but phases with different chemical composition such as "free chalk" CaO or alite C3S are not permitted in this concentration range.

2.2.2.5 *Peritectics and Peritectoids*

Every compound A_xB_y disintegrates at some specific temperature T_{AxBy}. Often, a liquid phase is the result, and then the composition of this liquid phase is the same as of the initial solid: A_xB_y (sol) → (xA + yB) (liq). This process is called *congruent melting* and is shown by $LiCsF_2$ (sol) in Figure 2.34. The behavior of $LiRbF_2$ in the right panel of the figure is different, because at its disintegration temperature T_{per} not only a liquid, but additionally solid LiF is formed. One can formulate the peritectic reaction $LiRbF_2$ (sol) → [x_{per}LiF + (1 − x_{per})RbF] (liq) + LiF (sol) with $x_{\mathrm{per}} \approx 0.47$. Peritectic decomposition (= peritectic melting) is one form of incongruent melting, because the melting compound is not in equilibrium with a liquid of the same composition. But not every incongruent melting requires a peritectic reaction—instead it was shown in

Section 2.2.2.3 that already in simple mixed crystal systems the liquid and solid phase in equilibrium have different composition, hence the solid shows incongruent melting.

Figure 2.35 shows the typical case where an intermediate compound (here assumed to have the composition AB_2) undergoes peritectic melting and solid B is the higher melting neighbor phase that is formed. At T_{per}, three phases are in equilibrium, which is the maximum for a two-component system (Eqn (2.6)), and there is no degree of freedom left. The point where B liquidus, AB_2 liquidus, and peritectic line meet is called the *peritectic point* and has the peritectic composition x_{per}. Solid AB_2 crystallizes only from melts with initial composition $x_{eut} < x_0 < x_{per}$. If one sets for the left diagram in Figure 2.35 x_0 as close as possible to x_{per} (but not beyond it!), one obtains with Eqn (2.42) the maximum yield $Y_{max} = 54.5\%$ for the growth of AB crystals from the melt ($x_0 = x_{per} = 0.5$, $x_{eut} = 0.3$, $x_{sol} = 0.6667$).

Also, compounds showing peritectic melting may have a finite phase width, and an example $A_{1\pm\delta}B_2$ is given in the right panel of Figure 2.35. The left and right phase boundaries of $A_{1\pm\delta}B_2$ are the solidus of this phase and are limiting a one-phase field. The upper limit of this field must be a point on the peritectic line; if both phase boundaries met not in one point, the composition of $A_{1\pm\delta}B_2$ would be left as a degree of freedom, which infringes the phase rule. Not only can intermediate compounds melt peritectically, but also rim compounds if a solid solution is formed that melts higher compared to the pure component. The system ZnO–MgO is an example where doping by $\lesssim$7% Mg increases the melting temperature of ZnO by ca. 45 °K [51].

In a peritectic reaction, a low T solid phase β is in equilibrium with another solid phase and a liquid: β (sol) $\rightleftarrows$ α (sol) + *liq*. A peritectoid reaction is an analog with the difference that a third solid phase replaces the liquid. Examples are $MoNi_4$ (sol), which decomposes peritectoid to $MoNi_3$ (sol) + Ni:Mo (f.c.c.), and subsequently $MoNi_3$, which decomposes to MoNi(δ) + Ni:Mo (f.c.c.), or $FeAlO_3$ in the $FeO–Fe_2O_3–Al_2O_3$ system [52].

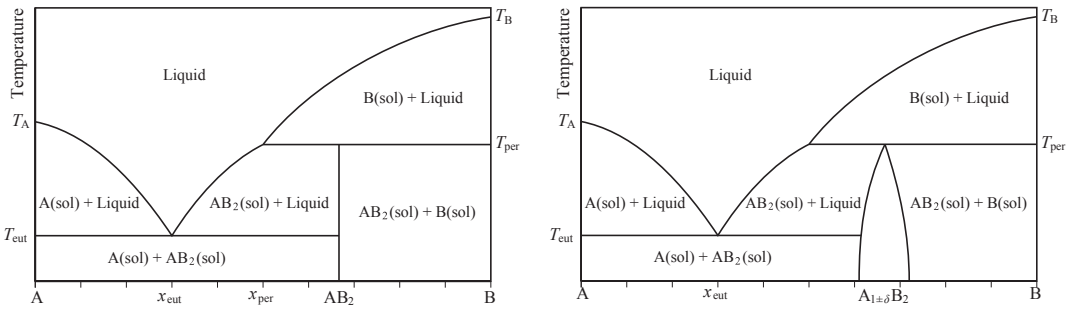


FIGURE 2.35 Two types of A–B phase diagrams with an intermediate compound AB_2 showing peritectic decomposition at T_{per}. Left: AB_2 is a line compound (daltonide, Section 2.1.4.1) with fixed composition $x_B = 0.6667$. Right: AB_2 is a berthollide with finite homogeneity width δ.

2.2.2.6 Syntectics and Monotectics

Both phenomena are related to demixing in the liquid phase. An example was given in Figure 2.20 where in the CaO–SiO_2 system SiO_2-rich melts are single phase only above ca. 1900 °C. The maximum of the "two melts" phase field near 89% SiO_2 is a critical point, and below it, more silicatelike melt is in equilibrium with another more oxidelike melt. Such behavior is called *monotectic*. The system iodine–lead even has two regions with immiscibility of liquid phases: For $x_I \lesssim 0.1$ and $T > 368$ °C, a monotectic miscibility gap occurs. Besides the intermediate phase PbI_2 ($x_I = 0.6667$) is at 406° in a syntectic equilibrium with two immiscible melts of slightly different composition ($x_I \approx 0.53$ or ≈ 0.7, respectively) [53].

2.2.3 Three and More Components

In the previous sections, thermodynamic systems with either one or two components were considered. With the Gibbs phase rule (Eqn (2.6)), one has in the easiest case of constant pressure (often $p = 1$ bar) for fields with only one phase ($P = 1$) $F = C$, hence two degrees of freedom in two-component systems, which can be well represented in 2D drawings. For systems with three or more components, this is not so straightforward, and three options remain:

1. 3D diagrams can be constructed for systems with $C = 3$, and drawn in a suitable perspective. Although such procedure is instructive, a severe drawback appears, because quantitative data are hard to read from perspective representations.
2. The number of degrees of freedom is reduced by keeping some quantities constant (e.g., T, or some x_i), or by defining dependencies between several of these parameters (e.g., $x_i/x_j = \text{const.}$). For quantitative representations, this method is preferred and will be used almost exclusively in this section.
3. Projections onto the surface of a specific phase (often the liquid) are performed.

2.2.3.1 x – y Diagrams

If x_A, x_B, x_C are the corresponding molar fractions of a ternary system, one has $x_A + x_B + x_C = 1$ and thus two independent concentrations. It is a reasonable option to set $p = \text{const.}$, $T = \text{const.}$ and choose, e.g., x_A/x_B and x_C/x_B as coordinates. This type of plot can be useful if one of the components may be distinguished, e.g., the main component of an alloy. In Figure 2.32, this is shown for iron and the two important steel additives carbon and manganese. The left panel shows an isothermal section through the ternary system near the Fe-rich corner, and the right panel is a projection of the same concentration range on the boundaries of one specific phase (here: face-centered cubic iron, γ-Fe). This means that the viewer looks from inside the γ-Fe phase field, which is not shown in the diagram, down to the boundaries of neighboring phases, which are differently shaded. For pure Fe, the f.c.c. phase is stable >912 °C and it transforms to α-Fe for lower T. By adding C (Figure 2.31) and/or Mn, the transition temperature can be

lowered. For C concentrations that are too high, free carbon separates as graphite, and for C + Mn concentrations that are too high, an iron–manganese carbide, cementite, is formed. The phase rule (Eqn (2.6)) says for $C = 3$ (Fe, C, Mn) under isobar conditions $P + F = 4$. Consequently, up to four phases can exist in equilibrium at nonvariant points ($F = 0$). "1" in the right panel of Figure 2.32 is such point where γ-Fe, α-Fe, graphite, and cementite are in equilibrium. Analogous diagrams can be constructed for $C > 3$ if for every additional component beyond three, one restriction (constant concentration, or constant concentration ratio) is implemented.

Such $x - y$ diagrams with two concentration ratios (or concentrations) as Cartesian axes are often appropriate for the presentation of systems where one component is clearly prevailing, such as alloys with a main component, or semiconductor systems with their dopants. For other systems with equivalent weighed components, Cartesian diagrams are not so well suited. The example LiCl–NaCl–KCl is shown in Figure 2.36. These substances exhibit unlimited solubility in the liquid phase and form solid solutions (halite structure). In the solid phase, however, the miscibility is unlimited only for the partial systems KCl–NaCl and LiCl–NaCl. Both partial systems have an azeotrope point, whereas KCl–LiCl is eutectic (insert). Along the abscissa of Figure 2.36, right, the composition shifts from pure NaCl (KCl/NaCl = 0) to KCl/NaCl = 5, which corresponds to $x_{KCl} = \frac{5}{6}$. The isotherms show there is a common minimum near KCl/NaCl = 1, which corresponds with the azeotrope point in Figure 2.27 right. The liquidus minimum of the ternary system appears at the composition $x_{LiCl} = 0.51$, $x_{NaCl} = 0.12$, $x_{KCl} = 0.37$, close to the binary eutectic KCl–LiCl. This composition is located inside the 360 °C isotherm close to the top rim of the ternary $x - y$ diagram (Figure 2.36). From the construction of both axes in this diagram, it is obvious that the limit system KCl–LiCl can never be reached, which is a severe drawback of this diagram type. A concentration triangle of this system, such as shown in Figure 2.37 can solve this problem.

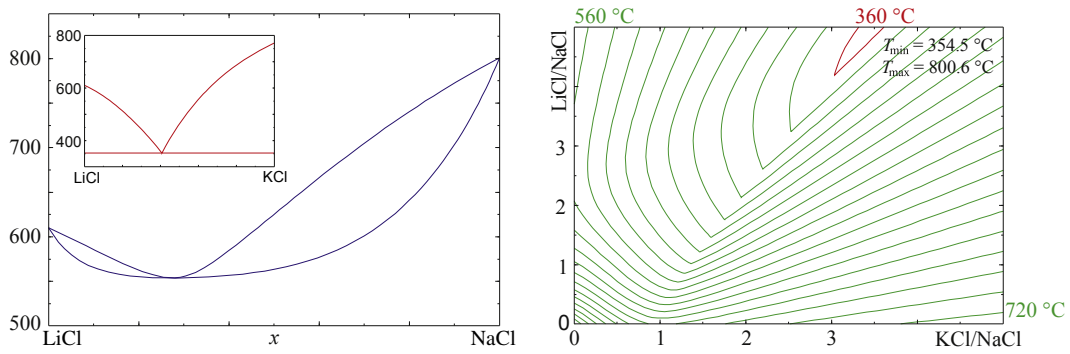


FIGURE 2.36 Left: The binary system NaCl–LiCl shows complete miscibility with an azeoptrope, such as NaCl–KCl (Figure 2.27). In contrast, KCl–LiCl form a eutectic (insert). Right: The system NaCl–LiCl–KCl with a projection onto the liquidus surface. NaCl at the origin has the highest melting point $T_f = 801$ °C. The isotherms have 20 K difference, and three of them are labeled. See also Figure 2.37.

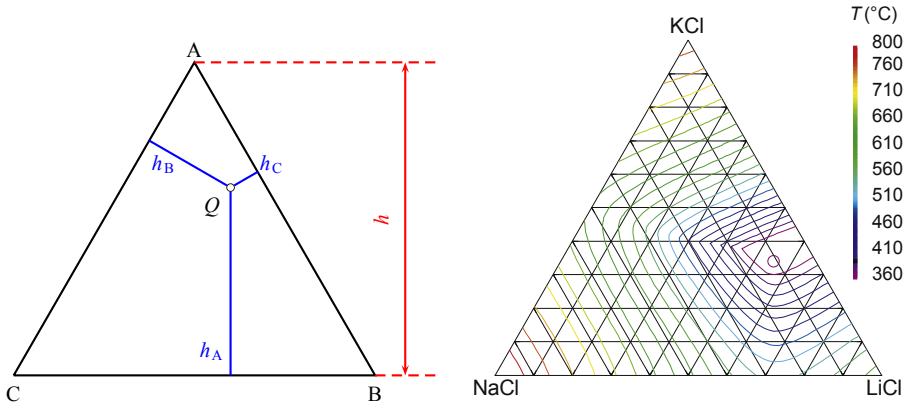


FIGURE 2.37 Left: Gibbs triangle A–B–C with one intermediate composition $x_A = 0.6$, $x_B = 0.3$, $x_C = 0.1$. For every composition one has $x_A + x_B + x_B = h$ (Eqn (2.43)). Right: The system NaCl–LiCl–KCl (see Figure 2.36) right as Gibbs triangle in polythermal projection. The eutectic point is marked by a circle, 20 K isotherms.

2.2.3.2 Concentration Triangles

The example Figure 2.36 right demonstrates a disadvantage of Cartesian presentations for ternary systems: The component on the bottom left corner of the diagram is always distinguished. This is reasonable if the other components are only minor additives to the main component, as in Figure 2.32. If, however, all components are be treated similarly, the "concentration triangle" (Gibbs triangle) is suitable for three-component systems of components A, B, C.

This construction is shown in Figure 2.37 and consists of a regular triangle where the pure components "A," "B," "C" are represented by the corners. The concentration triangle relies on Viviani's theorem that in a regular triangle, for every point Q the sum of the distances of this point from the three sides equals the height of the triangle

$$h_A + h_B + h_C = h \tag{2.43}$$

and is therefore constant. If $h = 1.0$ (100%) is chosen, h_A, h_B, h_C can represent the concentration of one component, that add up to unity. The following holds:

- All compositions where the concentration of one component is constant are represented by lines parallel to the triangle side opposite to the corner of this component. For example, in Figure 2.37, left, a horizontal line through Q represents all compositions $x_A = 0.6$, $x_B = 0 \ldots 0.4$, $x_C = 0.4 - x_B$.
- All compositions where the ratio between two components is constant are represented by straight lines that start from the corner of the third component. For example, in Figure 2.37, left, a line from the top corner "A" through Q to the bottom line represents all compositions $x_A = 1 \ldots 0$, $x_B/x_C = 3$.

For another system in Figure 2.38, right, the four regions of primary crystallization are distinguished by different shadings. From NaCl-rich compositions, NaCl crystallizes first

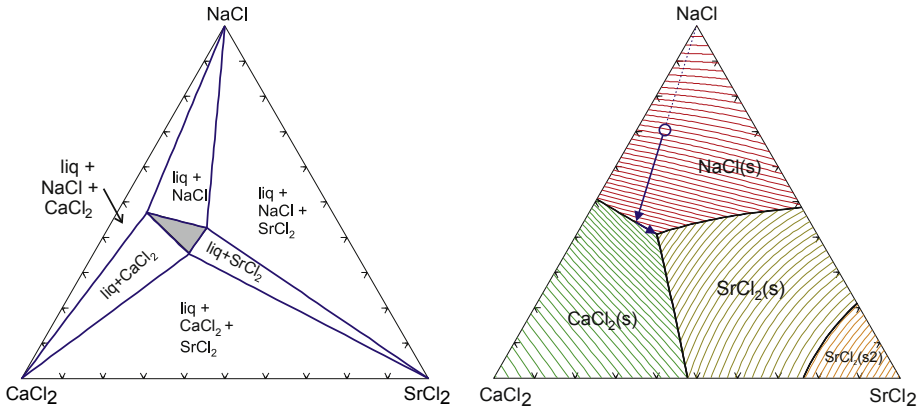


FIGURE 2.38 The ternary system NaCl–$CaCl_2$–$SrCl_2$ under the (somewhat simplifying) assumption that no mixing occurs in the solid phases. The melt shows complete mixing. NaCl: $T_f = 801$ °C, $CaCl_2$: $T_f = 772$ °C, $SrCl_2$: $T_f = 874$ °C, $T_t = 727$ °C. Left: isothermal section at 460 °C, "*liq*" field shaded. Right: Projection on the liquidus surface with 10 K isotherms and crystallization path for a NaCl-rich melt.

(top); and from $CaCl_2$-rich compositions, $CaCl_2$ crystallizes (bottom left). Solid $SrCl_2$ undergoes a solid-state phase transition s/s2 at $T_t \approx 727$ °C. Only from melts with $\gtrsim 80\%$ $SrCl_2$ does the high T phase s2 crystallize, because the liquidus temperature is $> T_t$ there (bottom right corner); for smaller $SrCl_2$ concentrations, the low T "s" phase crystallizes first.

The thick black line between the s2 and s field is the intersection of the T_t-isotherm with the liquidus surface. Most regions of primary crystallization are separated by three "eutectic valleys" represented by black lines. The eutectic valleys meet in the ternary eutectic point, where NaCl(s), $CaCl_2$(s), $SrCl_2$(s), and *liq* are in equilibrium.

If a molten sample of an arbitrary composition inside one of the primary crystallization fields is cooled, first the corresponding pure component phase crystallizes. The melt is depleted from this component, and its composition shifts to the opposite direction until it touches one of the eutectic valleys. There, the parallel crystallization of a second (pure component) phase starts, together with the previous one. This way, the melt composition moves downward (to lower T) along the valley, until it reaches the ternary eutectic point. There, the rest of the melt crystallizes isothermally. The crystallization path for a NaCl-rich melt (blue circle) is shown in the polythermal projection. In the left panel of Figure 2.38, the same system is shown in an isothermal section ca. 25 K above the eutectic temperature, where only in a small region around the eutectic composition a single phase field for the liquid remains.

2.2.3.3 Isopleth Sections

If T is requested as coordinate for systems with three or more components, instructive sections can be used where either one concentration is kept constant, or where the ratio of two concentrations is kept constant. Both cases are demonstrated in Figure 2.39 for the same system that is shown in Figure 2.38.

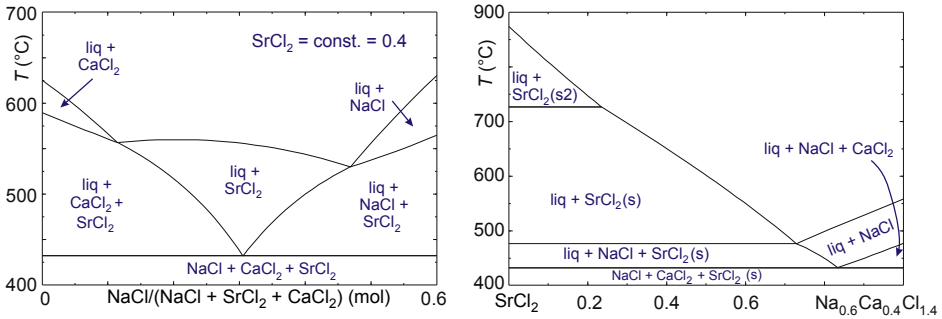


FIGURE 2.39 Sections through the ternary system NaCl–$CaCl_2$–$SrCl_2$. Left panel: for a constant $SrCl_2$ concentration $x_{Sr} = 0.4$. Right panel: for a constant concentration ratio NaCl/$CaCl_2$ = 1.5.

Figure 2.38, right, was a projection on the liquidus surface of the ternary system. The left panel of Figure 2.39 is a section through this concentration triangle perpendicular to the projection plane and parallel to the left side of the triangle, at $x_{Sr} = 0.4$. This section starts (from bottom) in the $CaCl_2$ field, crosses the $SrCl_2$ field, and terminates in the NaCl field. If the crystallization starts, e.g., inside the NaCl field, later the eutectic valley with $SrCl_2$ is reached and this phase crystallizes parallel until the ternary eutectic point. There, all three component phases are crystallizing.

The right panel of Figure 2.39 is a section through Figure 2.38 perpendicular to the projection plane and from the $SrCl_2$ corner to a point $Na_{0.6}Ca_{0.4}Cl_{1.4}$, which is a small distance above the middle of the NaCl/$CaCl_2$ side of Figure 2.38. Starting from $SrCl_2$, high T $SrCl_2$ crystallizes first. Subsequently, with lower liquidus temperature, low T $SrCl_2$ crystallizes first, and in the end NaCl crystallizes first.

It is very important to note that horizontal lines (T = const.) in such isopleth sections through ternary systems are in general not connecting phases which are in equilibrium! Practically, this means that the composition of all coexisting phases does not remain on such sections. Hence, the isothermal lines are not tie-lines, and the lever rule (Section 2.2.2.3) cannot be used.

2.2.3.4 Reciprocal Salt Pairs

Pairs of compounds with well-defined and interchangeable cations C1, C2, and anions A1, A2 are called *reciprocal salt pairs*. In such systems, C1 and C2 or A1 and A2 can replace each other in arbitrary ratio, but the relation (C1 + C2)/(A1 + A2) must be constant for maintaining charge neutrality. The presentation is typically done as square diagrams with the possible compounds C1A1, C1A2, C2A1, and C2A2 at the corners. An example is given in Figure 2.40 and shows that the mutual substitution

$$NaCl + LiF \rightleftarrows NaF + LiCl \tag{2.44}$$

leads to the presence of four compounds in the system if only two are supplied. Nevertheless, the system can be considered to be ternary, due to the dependency

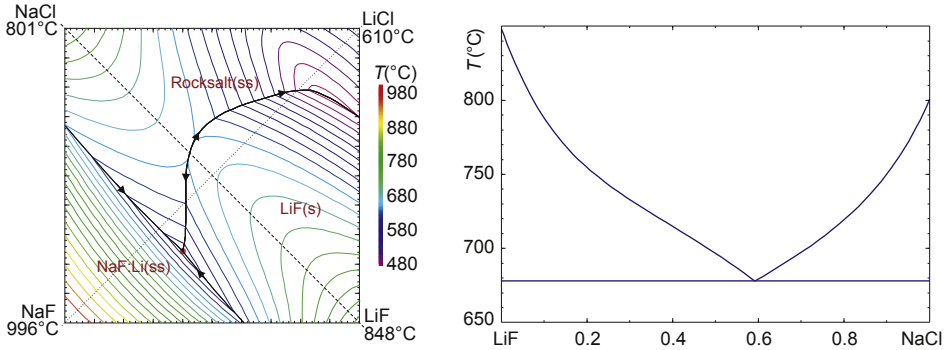


FIGURE 2.40 Left: The reciprocal salt pair (Li,Na)(F,Cl) in polythermal projection with 20 K isotherms. At $T_{eut} = 608$ °C (red dot) rocksalt(ss), NaF:Li(ss), LiF(s) and the melt are in equilibrium. Right: The eutectic of the NaCl–LiF subsystem corresponds with the saddle point in the left panel, on the dashed line.

Eqn (2.44), and according to the phase rule (Eqn (2.6)) four phases are in equilibrium at the (invariant) eutectic point.

In Figure 2.40, all rim systems except NaCl–LiCl are eutectic, which somewhat contradicts "Phase Diagrams for Ceramists" No. 3622, [6] where another eutectic and the compound $LiNaCl_2$ were claimed to exist in the upper rim system. The more recent Figure 2.36 [54], however, shows that this system forms solid solutions over the whole concentration range instead. Eutectic valleys connect the eutectic points at the rims and run to the ternary eutectic. NaCl–LiF (dashed line) can be treated as a pseudobinary system and the eutectic composition near 59% NaCl (Figure 2.40 right) is a saddle point in the liquidus projection of the left panel. This is necessary because otherwise the pseudobinary eutectic composition $(NaCl)_{0.59}(LiF)_{0.41}$ would, upon crystallization, leave the pseudobinary dashed line in the reciprocal salt pair diagram. It should be noted that the other pair NaF and LiCl (dotted line) cannot set up a pseudobinary system, because the melting point of LiCl is too low compared to the liquidus in the middle of the system (ca. 695 °C).

2.3 Equilibria Including Gas Phase

The transformations solid–gas (sublimation) as well as liquid–gas (evaporation, boiling) are first-order transitions, and consequently the gas phase can be shown in phase diagrams without further precautions. This is demonstrated in Figure 2.3 for sulfur (one component) or in Figure 2.27, left, for Si–Ge (two components). But very often the gas phase is not shown because evaporation plays no role in a specific process. In other cases, the gas phase must be taken into account, namely if at least one main species in the system has a significant volatility, if chemical or physical gas phase transport is performed, or if the gas phase equilibria determine the valency of species in the condensed phases. These cases will be considered in the following subsections.

2.3.1 Volatile Species

If a chemical element such as iodine, the heavier chalcogens (Figure 2.3), the heavier pnictogens, or even some metals such as Zn or Hg are heated, their vapor pressure quickly approaches ambient pressure and volatility becomes significant. Also, some compounds that are academically or technically interesting (e.g., NH_4Cl, AlN, SiC) can be transferred completely to the gas phase, and recrystallized at a colder seed. This process is called *sublimation* or *physical vapor transport* (PVT). For some other compounds, the sublimation temperature is so high that PVT is not straightforward (ZnO), or the fugacities of constituents are so different that only the more volatile component evaporates sufficiently (GaN). Then, transport agents may be added to create an intermediate state where the desired compound can be transported from the feed to the seed.

Figure 2.41 demonstrates the CVT growth of ZnO crystals where the chemical transport is described by the equilibrium

$$\mathrm{ZnO} + \mathrm{C} \rightleftarrows \mathrm{Zn} + \mathrm{CO} \tag{2.45}$$

which is shifted to the product side upon heating. The calculated diagram shows that mixtures of ZnO and carbon (graphite) are stable only below ca. 900 °C. For intermediate compositions, a gas phase is formed consisting mainly of Zn(g) + CO(g), which reacts at low T back to ZnO(s) (Eqn (2.45)). From the diagram, it is obvious that if thermal gradients are too large, this can lead the process to the "C + ZnO" phase field, which means that graphite inclusions may occur [56].

The left panel of Figure 2.42 shows the binary diagram gallium–arsenic with the intermediate 1:1 compound gallium arsenide. GaAs forms eutectics with both of its components; the eutectic to the Ga side is degenerate, as the melting temperature of gallium is low (30 °C). Gallium arsenide is a basic material in the semiconductor industry, mainly for solid-state lighting and high-frequency devices. Unfortunately, this x–T diagram does not

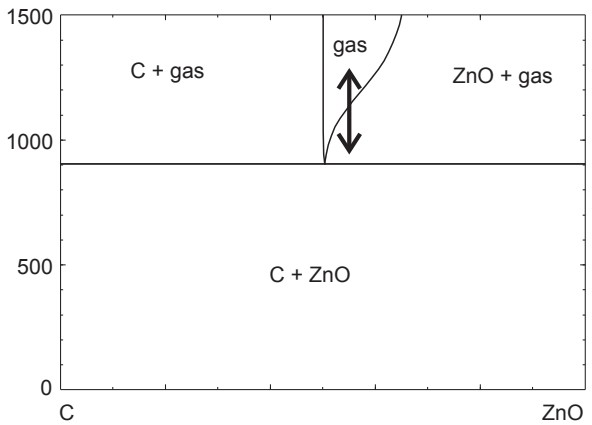


FIGURE 2.41 Phase diagram for the system ZnO–C for a system pressure $p = 1$ bar. The arrow shows the conditions under which chemical vapor transport can be performed. *Reprinted with permission from Elsevier [55].*

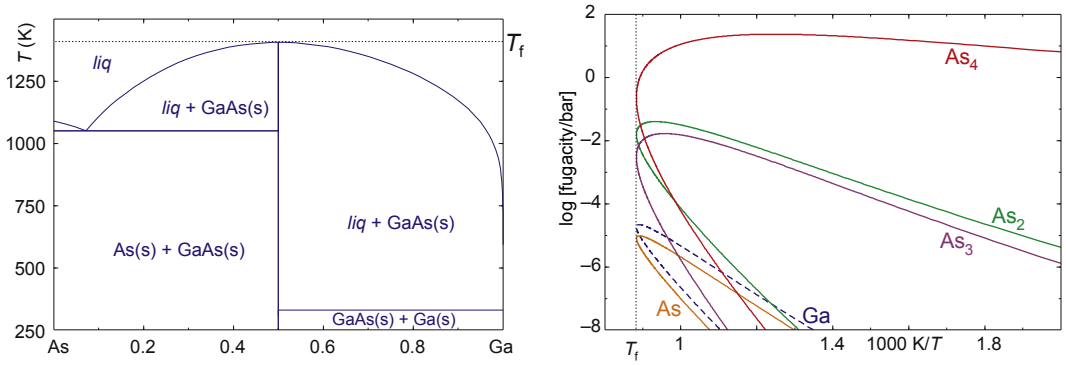


FIGURE 2.42 Left: The phase diagram Ga–As with the intermediate compound GaAs considering condensed phases only. Right: Fugacities of gas species along the GaAs liquidus.

express the high volatility, mainly of arsenic, which results in the decomposition of GaAs to arsenic gas and gallium liquid before melting under ambient pressure.

The GaAs liquidus spans almost the complete phase diagram, consequently GaAs(s) can be grown from a wide range of Ga_xAs_{1-x} compositions. The liquidus temperature at which crystallization starts will depend on x. Besides, the chemical composition can be varied slightly if the melt has an excess of one component [57]. For the control of crystal growth, it is desirable to know the equilibrium vapor pressure ($\approx$ fugacity f) of both components along the GaAs liquidus. The right panel of Figure 2.42 shows this for several species occurring in the gas phase in a plot log (fugacity) versus $1/T$. The curves there have two branches, and the upper and lower branches for the As_i ($i = 1 \ldots 4$) curves show $f_{Asi}(1/T)$ on the As-rich or Ga-rich side the GaAs liquidus, respectively. Vice versa for $f_{Ga}(1/T)$, where the higher values are reached on the Ga-rich side. All curves have an apex at the melting point T_f of GaAs.

2.3.2 Ellingham Type Diagrams

Some chemical elements, especially many transition metals, tend to form oxides, sulfides, and other compounds in multiple valency states. Then, chemical equilibria of the kind

$$2\mathrm{MeO}_{m/2} + \frac{1}{2}\mathrm{O}_2 \rightleftarrows 2\mathrm{MeO}_{(m+1)/2} \tag{2.46}$$

describe the transition from valency state m to the higher state $m + 1$. The Gibbs free energy balance $\Delta G^0 = \Delta H^0 - T\Delta S^0$ of this reaction determines which valency (here, oxidation) state is stable. If the metal oxides show no significant volatility, the chemical equilibrium (Eqn (2.46)) is determined by T and the chemical potential of oxygen (Eqn (2.23)). For the case $T =$ const. and two different volatiles V_1, V_2 in the system linear predominance diagrams were already constructed above from plots log p_{V2} versus log p_{V1} (Figure 2.14). Plots $RT \ln p_{O_2}$ versus T for redox reactions of the type (Eqn (2.46))

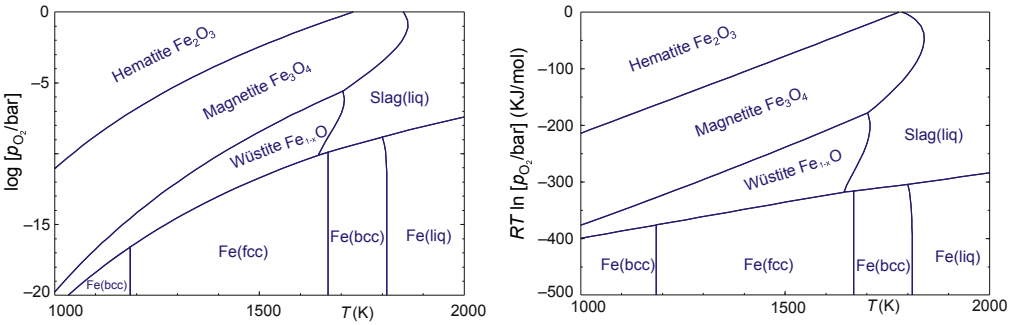


FIGURE 2.43 Predominance diagrams of the system Fe–O_2 in coordinates T, log p_{O_2} (left) or T, RTln p_{O_2} (right, Ellingham type), respectively.

are also straight lines. Such Ellingham diagrams can be used as predominance phase diagrams showing the stability fields of several oxides under the given conditions T, p_{O_2} [28].

Figure 2.43 shows the system iron–oxygen in two representations. In the right panel (Ellingham type), the phase boundaries are often almost straight lines. The "slag" phase covers a wide composition range of iron oxides where the melting temperature depends on the oxygen partial pressure, hence on the composition of condensed phases. Vertical lines at the bottom describe the phase transitions of metallic iron up to its melting (cf. Figure 2.31).

For practical purposes, often the simpler diagram type from the left panel is useful, as the important experimental parameter p_{O_2} is plotted directly. Then, phase boundaries between neighboring FeO_x are bend lines. Obviously, oxidation states with lower valency become more stable if T rises and if p_{O_2} drops, and this behavior is typical as oxidation reactions (Eqn (2.46)) are mainly exothermic. Phase boundaries with negative slope occur only in scarce cases (e.g., Re_2O_7, CrO_2).

Predominance diagrams such as shown in Figure 2.43 can easily be computed for systems with more than one metallic component. If mixed oxides

$$Me'O_x + Me''O_y \rightleftarrows Me'Me''O_{x+y} \tag{2.47}$$

are formed, these compounds appear with separate phase fields in the diagrams. In cases where $G(T)$ for $Me'Me''O_{x+y}$ are unknown, the calculation of stability fields for the simple oxides $Me'O_x$, $Me''O_y$ often gives a good approximation, because energetic contributions of reactions between oxides (Eqn (2.47)) are typically small compared to the formation enthalpies (Eqn (2.46)).

For narrow phase fields, such as $Fe_{1-x}O$ in the left panel of Figure 2.43, often no constant p_{O_2} can be found that lies totally inside this predominance field. For every process where temperature gradients occur—which is always the case for crystal growth—every specific p_{O_2} = const. then crosses several phase fields, which means that different oxides are stable at different T. Fortunately, several gaseous oxides of

Table 2.1 Oxygen Fugacities (p_{O_2} in bar) in Different Atmospheres (Total Pressure 1 bar)

Gas	500 °C	1000 °C	1500 °C	2000 °C	2500 °C
CO_2	1.24×10^{-10}	1.23×10^{-5}	1.79×10^{-3}	2.67×10^{-2}	1.14×10^{-1}
10% CO_2 +90% CO	1.76×10^{-27}	9.30×10^{-17}	2.87×10^{-10}	1.14×10^{-6}	2.13×10^{-4}
CO	1.76×10^{-27}	3.97×10^{-19}	1.87×10^{-16}	6.49×10^{-15}	5.46×10^{-13}

nonmetals, such as H_2O (as humidity in carrier gases), or CO_2 and CO, decompose upon heating, e.g., by

$$CO_2 \rightleftarrows CO + \frac{1}{2}O_2 \rightleftarrows C + O_2 \tag{2.48}$$

similarly to metal oxides, and under suitable conditions may result in a "self-adjusting oxygen partial pressure" $p_{O_2}(T)$ that runs for all T through the stability field of a desired oxide. Table 2.1 shows that p_{O_2} changes over several orders of magnitude upon heating, and often the right choice of such "reactive atmosphere" allows to keep $p_{O_2}(T)$ over the whole T range of interest inside the stability field of the desired oxide [58].

A similar approach of reactive atmospheres is feasible not only for oxides but also for other anions where a suitable carrier can be found. For sulfides, e.g., H_2S can be used to adjust the sulfur (S_2) partial pressure from 1.4 mbar (500 °C) to 270 mbar (2000 °C) under ambient total pressure. Then, a predominance diagram for the corresponding Me–S phases must be created in analogy to Figure 2.43, left, with T and $\log(p_{S_2})$ as coordinates. A suitable growth atmosphere should give a $p_{S_2}(T)$ that runs completely through the stability field of the desired sulfide.

References

[1] Alloy Phase Diagrams Center, Villars P. In: Okamoto H, Cenzua K, editors. ASM alloy phase diagram database. Materials Park, OH: ASM International; 2006–13. http://www1.asminternational.org/AsmEnterprise/APD/.

[2] Villars P. Pauling file – inorganic materials database; 2013. http://paulingfile.com/.

[3] GTT Technologies, Kaiserstr. 100, 52134 Herzogenrath, Germany, FactSage 6.4; 2013. http://www.factsage.com/.

[4] Magnéli A, Blomberg B. Studies on the vanadium pentoxide—molybdenum trioxide system. II. Phase analysis. Acta Chem Scand 1951;5:585–9.

[5] Eick HA, Kihlborg L. The crystal structure of V_2MoO_8. Acta Chem Scand 1966;6:1658–66.

[6] American Ceramic Society. NIST, phase equilibria diagrams 3.4; 2013. http://ceramics.org/publications-and-resources/phase-equilibria-diagrams.

[7] Paufler P. Phasendiagramme (in German). Braunschweig; Wiesbaden: Vieweg; 1981.

[8] Alcock CB, Chase MW, Itkin VP. Thermodynamic properties of the group Ia elements. J Phys Chem Rev Data 1994;23:385–497.

[9] Barin I. Thermochemical data of pure substances. Weinheim: Wiley-VCH; 2008.

[10] Chase JMW, Curnutt JL, Downey JJR, McDonald RA, Syverud AN, Valenzuela EA. JANAF thermochemical tables, 1982 supplement. J Phys Chem Rev Data 1982;11:695–940.

[11] Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27:379–656.

[12] Landau L. The theory of phase transitions. Nature 1936;138:840–1.

[13] Salje EKH. Phase transitions in ferroelastic and co-elastic crystals. Cambridge: University Press; 1990.

[14] Yeomans JM. Statistical mechanics of phase transitions. Oxford: University Press; 1992.

[15] Boysen H, Altorfer F. Neutron powder investigation of the high-temperature structure and phase transition in $LiNbO_3$. Acta Cryst B 1994;50:405–14.

[16] Dohnke I, Trusch B, Klimm D, Hulliger J. A study on the influence of ytterbium and impurities on lattice parameters and the phase transition temperature of Czochralski-grown $LiNbO_3$. J Phys Chem Solids 2004;65:1297–305.

[17] Dove MT, Powell BM. Neutron diffraction study of the tricritical orientational order/disorder phase transition in calcite at 1260 K. Phys Chem Miner 1989;16:503–7.

[18] Carruthers JR, Peterson GE, Grasso M, Bridenbaugh PM. Nonstoichiometry and crystal growth of lithium niobate. J Appl Phys 1971;42:1846–51.

[19] Shaw RR, Uhlmann DR. Subliquidus immiscibility in binary alkali borates. J Am Ceram Soc 1968;51:377–82.

[20] Nitsch K, Cihlář A, Klimm D, Nikl M, Rodová M. Na–Gd phosphate glasses – preparation, thermal and scintillating properties. J Therm Anal Calor 2005;80:735–8.

[21] Redlich O, Kister AT. Algebraic representation of thermodynamic properties and the classification of solutions. Ind Eng Chem 1948;40:345–8.

[22] Hillert M, Jansson B, Sundman B, Ågren JA. A two-sublattice model for molten solutions with different tendency for ionization. Met Trans A 1985;16:261–6.

[23] Gray CG, Gubbins KE, Joslin CG. Theory of molecular fluids, vol. 2. Oxford: University Press; 2011 [Chapter 7].

[24] Delves RT. Constitutional supercooling and two-liquid growth of HgTe alloys. Br J Appl Phys 1965;16:343–51.

[25] Mühlberg M, Rudolph P, Laasch M, Treser E. The correlation between superheating and supercooling in CdTe melts during unseeded Bridgman growth. J Cryst Growth 1993;128:571–5.

[26] Pouget EM, Bomans PHH, Goos JACM, Frederik PM, de With G, Sommerdijk NAJM. The initial stages of template-controlled $CaCO_3$ formation revealed by cryo-TEM. Science 2009;323:1455–8.

[27] Wallace AF, Hedges LO, Fernandez-Martinez A, Raiteri P, Gale JD, et al. Microscopic evidence for liquid–liquid separation in supersaturated $CaCO_3$ solutions. Science 2013;341:885–9.

[28] Cahn RW, Haasen P, Kramer EJ. Materials science and technology. In: Cahn, et al., editors. Thermodynamics and phase diagrams of materials (Arthur D. Pelton), vol. 5. Multivolume ed. Weinheim: VCH: 1991. p. 1–73.

[29] Thermo-Calc Software. Norra Stationsgatan 93, Plan 5, 113 64 Stockholm, Sweden, Thermo-Calc 3.0, 2013. http://www.thermocalc.se/.

[30] Hallstedt B, Liu Z-K. Software for thermodynamic and kinetic calculation and modelling. Calphad 2009;33:265.

[31] NIST Standard Reference Database Number 69. NIST chemistry webbook; 2011. http://webbook.nist.gov/chemistry/.

[32] Springer materials—the Landolt-Börnstein database. Heidelberg: Springer-Verlag GmbH; 2013. http://www.springermaterials.com.

[33] Hull S, Norberg ST, Ahmed I, Eriksson SG, Mohn CE. High temperature crystal structures and superionic properties of $SrCl_2$, $SrBr_2$, $BaCl_2$ and $BaBr_2$. J Solid State Chem 2011;184:2925–35.

[34] Nye JF. Physical properties of crystals. Oxford: Clarendon Press; 1951.

[35] Schlom DG, Chen L-Q, Eom C-B, Rabe KM, Streiffer SK, Triscone J-M. Strain tuning of ferroelectric thin films. Annu Rev Mater Res 2007;37:589–626.

[36] Park JH, Coy JM, Kasirga TS, Huang C, Fei Z, Hunter S, et al. Measurement of a solid-state triple point at the metal-insulator transition in VO_2. Nature 2013;500:431–4.

[37] Haeni JH, Irvin P, Chang W, Uecker R, Reiche P, Li YL, et al. Room-temperature ferroelectricity in strained $SrTiO_3$. Nature 2004;430:758–61.

[38] Fast JD. The allotropic transformation of hafnium and a tentative equilibrium diagram of the system zirconium–hafnium. J Appl Phys 1952;23:350–1.

[39] Klimm D, Rabe M, Bertram R, Uecker R, Parthier L. Phase diagram analysis and crystal growth of solid solutions $Ca_{1-x}Sr_xF_2$. J Cryst Growth 2008;310:152–5.

[40] Carruthers JR, Grasso M. Phase equilibria relations in the ternary system BaO–SrO–Nb_2O_5. J Electrochem Soc 1970;117:1426–30.

[41] Megumi K, Nagatsuma N, Kashiwada Y, Furuhata Y. The congruent melting composition of strontium barium niobate. J Mater Sci 1976;11:1583–92.

[42] Ulex M, Pankrath R, Betzler K. Growth of strontium barium niobate: the liquidus–solidus phase diagram. J Cryst Growth 2004;271:128–33.

[43] Muehlberg M, Burianek M, Joschko B, Klimm D, Danilewsky A, Gelissen M, et al. Phase equilibria, crystal growth and characterization of the novel ferroelectric tungsten bronzes $Ca_xBa_{1-x}Nb_2O_6$ (CBN) and $Ca_xSr_yBa_{1-x-y}Nb_2O_6$ (CSBN). J Cryst Growth 2008;310:2288–94.

[44] Schiferl D, Barrett CS. The crystal structure of arsenic at 4.2, 78 and 299 K. J Appl Cryst 1969;2:30–6.

[45] Pawlak DA, Turczynski S, Gajc M, Kolodziejak K, Diduszko R, Rozniatowski K, et al. How far are we from making metamaterials by self-organization? The microstructure of highly anisotropic particles with an SRR-like geometry. Adv Funct Mater 2010;20:1–9.

[46] Gulyaev AP. On the iron–carbon diagram. Metal Sci Heat Treat 1990;32:493–4.

[47] Ganschow S, Klimm D. Shaped crystals—growth by micro-pulling-down technique. In: Fukuda T, Chani VI, editors. Ch. Growth of olivine and Wüstite crystals. Heidelberg: Springer, Berlin; 2007. p. 265–74.

[48] Gulliver GH. The quantitative effect of rapid cooling upon the constitution of binary alloys. J Inst Met 1913;9:120–57.

[49] Burton JA, Prim RC, Slichter WP. The distribution of solute in crystals grown from the melt. J Chem Phys 1953;21:1987–91. 1991–1996.

[50] Elwell D, Scheel HJ. Crystal growth from high-temperature solutions. London: Academic Press; 1975.

[51] Schulz D, Bertram R, Klimm D, Schulz T, Thiede E. Segregation of Mg in $Zn_{1-x}Mg_xO$ single crystals grown from the melt. J Cryst Growth 2011;334:118–21.

[52] Muan A, Gee CL. Phase equilibrium studies in the system iron oxide–Al_2O_3 in air and at 1 atm. O_2 pressure. J Am Ceram Soc 1956;39:207–14.

[53] Okamoto H. I–Pb (iodine–lead). J Phase Equilibria Diffusion 2010;31:320–1.

[54] Sangster J, Pelton AD. Phase diagrams and thermodynamic properties of the 70 binary alkali halide systems having common ions. J Phys Chem Ref Data 1987;16:509–61.

[55] Klimm D, Schulz D, Ganschow S. Growth of bulk ZnO. In: Bhattacharya P, Fornari R, Kamimura H, editors. Comprehensive semiconductor science and technology. Amsterdam: Elsevier; 2011. p. 302–38.

[56] Ntep J-M, Hassani SS, Lusson A, Tromson-Carli A, Ballutaud D, Didier G, et al. ZnO growth by chemical vapour transport. J Cryst Growth 1999;207(1–2):30–4.

[57] Wenzl H, Dahlen A, Fattah A, Petersen S, Mika K, Henkel D. Phase relations in GaAs crystal growth. J Cryst Growth 1991;109:191–204.

[58] Klimm D, Ganschow S, Schulz D, Bertram R, Uecker R, Reiche P, et al. Growth of oxide compounds under dynamic atmosphere composition. J Cryst Growth 2009;311:534–6.

3 Atomistic Calculation of Defect Thermodynamics in Crystalline Silicon

Talid Sinno

DEPARTMENT OF CHEMICAL AND BIOMOLECULAR ENGINEERING, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PA, USA

CHAPTER OUTLINE

3.1 Introduction 138
3.2 Theoretical Infrastructure for Analysis of Point Defect and Cluster Thermodynamics 139
3.2.1 Point Defects 139
3.2.2 Point Defect Clusters 140
3.3 Theoretical Estimation of Ground State Point Defect Formation Properties 142
3.3.1 Empirical Potential Atomistic Simulations 142
3.3.2 Quantum Mechanical Estimates for Point Defect Thermodynamics 145
3.4 Ground State Point Defect Cluster Thermodynamics 147
3.4.1 Vacancy Clusters 147
3.4.2 Small Compact Self-Interstitial Clusters 149
3.4.3 Large Planar Self-Interstitial Clusters 152
3.5 Inherent Structure Theory and Potential Energy Landscapes 153
3.5.1 Inherent Structure Landscape (ISL) Theory for Defects in Crystals 155
3.5.2 Sampling Inherent Structures with Molecular Dynamics Simulation 156
3.6 High Temperature Defect Thermodynamics 158
3.6.1 The Single Vacancy 158
3.6.2 Vacancy Clusters 159
3.6.3 Self-Interstitial Clusters 163
3.6.3.1 Landscape Roughness and the Effect of Pressure 165
3.7 Conclusions 167
References 169

Handbook of Crystal Growth. http://dx.doi.org/10.1016/B978-0-444-56369-9.00003-4

3.1 Introduction

Intrinsic point defects, namely vacancies and self-interstitials, in crystalline silicon have been the subject of an enormous number of studies over the last several decades. The driving force for this intense research effort stems from the fact that intrinsic point defects are responsible for a staggering range of fundamentally interesting phenomena and play important roles in the formation of almost all known microstructural features found in dislocation-free, single-crystal silicon. This material, most commonly in the form of thin wafers sliced from melt-grown ingots, underpins much of the (silicon-based) microelectronics and photovoltaics industries.

To this day, almost all single-crystal silicon is still grown by the venerable Czochralski (CZ) technique in which a seed of single-crystal is dipped into a highly pure silicon melt contained in a rotating, quartz-lined crucible and then slowly pulled away from the melt under tightly controlled thermal conditions to produce a cylindrical ingot [1]. Currently, silicon boules 300 mm in diameter are routinely grown weighing well over 250 kg, and next-generation 450 mm ingots remain under development. While dislocations and grain boundaries are completely suppressed in the CZ growth process, the formation of a range of point defect–impurity complexes and intrinsic point defect aggregates remains technologically relevant in spite of decades of research. With that said, current emphasis is principally on process optimization because while much of the basic mechanisms are well understood, established defect microstructure reduction measures usually come at a cost of reduced throughput and increased process expense.

Unlike impurities and intentionally introduced dopants, intrinsic point defect generation in growing crystals is a consequence of fundamental thermodynamic forces that cannot be avoided. Simply put, the minimum free energy state of any crystal at finite temperature does not correspond to crystalline perfection, but rather is one that includes a (temperature-dependent) distribution of point defects and small defect clusters. Fortunately, intrinsic point defects and very small clusters comprised of them are not directly associated with deleterious effects such as reduced charge carrier lifetimes; rather, it is how they interact with each other and with other species that ultimately leads to reduced material quality. The aim for most silicon crystal growers, therefore, is not to circumvent point defect formation altogether, but rather to cleverly design process conditions so that point defects do not subsequently lead to undesirable microstructure in the form of large aggregates. This target is not a static one: the very definition of "large" is continuously evolving due to improvements in detection techniques and increased sensitivity to ever-smaller defects as CMOS scaling continues and feature lengths decrease. Nonetheless, for this endeavor to be successful, a quantitatively accurate mechanistic picture of intrinsic defect thermodynamics must be available.

The goal of this chapter is to provide a (highly selective and necessarily incomplete) overview of intrinsic point defect and defect cluster thermodynamics that is relevant to the formation of vacancy and self-interstitial microdefects during silicon crystal growth and wafer annealing. The present emphasis is on predictions from molecular

simulations based on empirical interaction potentials. When appropriate, comparisons are made with results from more accurate quantum mechanical methods. Note that this chapter does not seek to provide a comprehensive quantitative summary of point defect and cluster formation energies and entropies, although whenever appropriate, literature references to such studies will be provided. The remainder of the chapter is structured as follows. In Section 3.2, the basic theoretical elements for describing point defect and cluster thermodynamics are briefly described. In Section 3.3, approaches for estimating ground state point defect formation energies are summarized; these include atomistic simulations based on quantum mechanical and empirical potential descriptions. In Section 3.4, analysis of ground-state defect cluster thermodynamics is presented. A more general theoretical framework for finite temperature defect thermodynamics then is presented in Section 3.5 and applied to single point defects and clusters in Section 3.6. Finally, conclusions and outlook are presented in Section 3.7.

3.2 Theoretical Infrastructure for Analysis of Point Defect and Cluster Thermodynamics

3.2.1 Point Defects

The materially open system equilibrium concentration of a particular point defect species is one of the most essential properties underlying any quantitative analysis of defect formation and evolution. This is because the equilibrium concentration directly dictates the extent of any point defect supersaturation (or undersaturation) present, which in turn dictates the extent of clustering. Here, "open system" denotes that the crystal domain is connected to a sink/source for point defects that allows an unconstrained equilibrium to be established everywhere in the domain—the crystal surface typically serves this purpose.

Consider the free energy of a crystal domain with N distinct lattice sites containing one type of point defect species,

$$G = G_0 + nG^f - k_B T \ln \frac{N!}{(N-n)!n!}, \tag{3.1}$$

where n is the number of point defects, G^f is the free energy of formation of a single point defect, and G_0 is the perfect crystal (reference) free energy. The last term in Eqn (3.1) represents the contribution of translational entropy to the total free energy, where it is assumed that each lattice site can accommodate a single point defect and no proximity effects are present, i.e., point defects do not interact with each other. The equilibrium concentration of point defects is obtained by minimizing Eqn (3.1) with respect to n, i.e.,

$$C^{eq}(T) = gC_s \exp\left(-\frac{G^f}{k_B T}\right) = gC_s \exp\left(\frac{S^f}{k_B}\right)\exp\left(-\frac{H^f}{k_B T}\right), \tag{3.2}$$

where, C_s is the density of lattice sites ($\sim 5 \times 10^{22}/\text{cm}^3$ for Si), g is a configurational degeneracy factor, and H^f, S^f represent the formation enthalpy and entropy, respectively, of a single point defect. The configurational degeneracy factor is usually taken to be the symmetry of the point defect structure and therefore is a number that is O(1) in magnitude; it is often ignored in analyses of point defect thermodynamics. We will return to the issue of defect degeneracy in Section 3.4 and show that the entropy arising from configurational degeneracy can in fact be quite significant. The remaining entropic term in Eqn (3.2) then corresponds to the vibrational entropy of formation. The conventional approach for calculating the equilibrium concentration of a point defect species using atomistic simulations is to first find the lowest energy configuration and then compute its formation energy and vibrational entropy, usually under the assumption that these are constants, i.e., that the equilibrium concentration is an Arrhenius function of temperature.

3.2.2 Point Defect Clusters

The thermodynamics of defect clusters can be formulated in a similar manner as for point defects. The total free energy of a system containing a distribution of a single point defect species and its aggregates is given by

$$G(T) = G_0 + \sum_i n_i G_i^f(T) - k_B T \ln \Omega, \tag{3.3}$$

here $G_i^f(T)$ is the formation free energy of a cluster of size i, n_i is the number of such aggregates, and Ω is the number of ways that a distribution of clusters can be distributed across lattice sites. For an open system (i.e., one with accessible surfaces) in which each cluster size can reach equilibrium independently of the others and assuming that the cluster site density is equal to lattice site density, Ω is the same as for the point defect case (Eqn (3.1)) and the cluster equilibrium concentrations are given by

$$C_i^{eq}(T) = g_i C_s \exp\left(-\frac{G_i^f}{kT}\right), \quad \forall i. \tag{3.4}$$

In many situations the crystal cannot be assumed to be materially open and the cluster size distribution is established via a constrained free energy minimization subject to a fixed total number of point defects. Such a situation is typical in crystal growth where the nearest surface is much further away than the point defect diffusion distance. Now, Ω is given by [2]

$$\Omega = \prod_i \frac{(i)^{n_i}(N/i)!}{(N/i - n_i)!(n_i)!} \tag{3.5}$$

which represents the number of ways of distributing $\{n_i\}$ clusters over the N lattice sites. The closed system equilibrium cluster concentration at each size is now obtained by minimizing the augmented free energy function

$$\widehat{G} = G_0 + \sum_i n_i G_i^f - k_B T \sum_i \left\{ n_i \ln i + \frac{N}{i} \ln \frac{N}{i} - \left[\frac{N}{i} - n_i \right] \ln \left[\frac{N}{i} - n_i \right] - n_i \ln n_i \right\}, \\ + \lambda \left(\sum_i i \cdot n_i - n_{tot} \right) \tag{3.6}$$

with respect to the number of each cluster size [3], i.e.,

$$\frac{\partial \widehat{G}}{\partial n_i} = G_i^f - k_B T \ln \frac{N - i \cdot n_i}{n_i} + i\lambda = 0, \quad \forall i. \tag{3.7}$$

The last term in Eqns (3.6) and (3.7) represents the constraint of a fixed total number of point defects and λ is a Lagrange multiplier. An immediate implication of Eqn (3.7) is that the equilibrium concentrations of all cluster sizes are coupled to each other. Under the physically reasonable assumption that clusters exist in dilute concentrations, $i \cdot n_i \ll N$, the closed system cluster equilibrium concentrations are given by solving Eqn (3.7) for each size

$$\widehat{C}_i^{eq}(T) = g_i C_s \exp\left(-\frac{\widehat{G}_i^f}{k_B T} \right) = g_i C_s \left(\frac{C}{C^{eq}} \right)^n \exp\left(-\frac{G_i^f}{k_B T} \right), \quad \forall i \tag{3.8}$$

where

$$\widehat{G}_i^f(T) = -ik_B T \ln\left(\frac{C}{C^{eq}(T)} \right) + G_i^f(T), \quad \forall i \tag{3.9}$$

and C and C^{eq} represent the actual and equilibrium concentrations of point defects, respectively. Note that the assumption of diluteness effectively decouples the equilibrium concentration of each cluster size from other sizes—the concentration is only a function of the extent of point defect local supersaturation, C/C^{eq}. Finally, applying Eqn (3.7) to the monomer point defect species gives an expression for the equilibrium monomer point defect concentration in the presence of a cluster of size i, i.e.,

$$C^{eq}|_i = C^{eq} \exp\left(\frac{\partial G_i^f / \partial i}{k_B T} \right), \quad \forall i. \tag{3.10}$$

The enhancement of the point defect equilibrium concentration in the vicinity of a cluster as specified in Eqn (3.10) is a statement of the Gibbs–Thompson effect [4]; alternatively, this equilibrium concentration can be thought of as the cluster "vapor pressure" in which the vapor phase consists of a monomer point defect "fluid" surrounding a cluster.

All equilibrium concentrations defined above, whether for point defects or for clusters, require the estimation of free energies of formation. There are two principal approaches for carrying out this task. The first is to directly compute free energies using atomistic calculations based either on quantum mechanical or empirical descriptions for the interatomic interactions. The second is to measure experimentally some property or phenomenon and extract the free energies using a model. Both approaches have

inherent limitations and advantages. Atomistic scale calculations are the most direct path for free energy estimation: the system is completely defined and the free energy can be unambiguously accessed. On the other hand, empirical atomistic calculations, which are computationally very efficient, are completely dependent on the specification of a good interatomic potential function. The issue of accuracy is mostly (but not completely) resolved with quantum mechanical calculations, but the immense computational requirements of these methods limit the system sizes that can be considered and also largely preclude the use of statistical sampling approaches commonly employed for free energy estimation. In the following section, we describe the salient features of each approach for computing point defect thermodynamic properties.

3.3 Theoretical Estimation of Ground State Point Defect Formation Properties

3.3.1 Empirical Potential Atomistic Simulations

Empirical potentials are constructed beginning with an ansatz for the functional form of the interaction energy between pair, triplet, and higher-order (many-body) interaction functions. The resulting potential function is usually characterized by several parameters that are regressed to a number of material properties that are either experimentally measured or computed using more accurate quantum mechanical calculations. The challenge for accurate modeling of covalently bonded materials such as silicon is the directionality of the bonds, leading to structures (such as the diamond lattice) that typically require many-body interaction terms to stabilize. Arguably, silicon has served as the prototypical system for the development of potential models for covalently bonded materials and there exists a large number of potentials and parameterizations in the literature.

In an excellent early review, Balamane et al. [5] defined two main types of empirical formalisms, “cluster potentials” in which pair and triplet interactions terms were separately defined, and “cluster functionals” or bond-order formulations, in the which the pair interaction is modified by many-body (usually three) environment terms. The Stillinger–Weber (SW) potential [6] was the first “comprehensive” model to address both solid and liquid phases, interfaces and surfaces, and defects. The potential was represented by separate two- and three-body terms with the tetrahedral angle built explicitly into the latter. Similar cluster potentials include the Pearson, Takai, Halicioglu, Tiller (PTHT) [7] and Biswas–Hamann (BH) [8] potentials, although these are less popular than the SW. The Tersoff model [9,10] followed soon after with similar capabilities as the SW potential but was based on a bond-order formulation. The SW and Tersoff potentials have since been applied in countless studies, being reparameterized several times [11–14] and extended to Group IV alloy systems [10,12,15,16] and continue to be popular to this day. Another popular class of potentials is the modified embedded-atom model (MEAM) [17–20], in which the embedded-atom model for metals was extended to

include directionality for covalent materials such as silicon. The MEAM potential framework is attractive because parameters have been generated for a wide range of materials and compounds making it quite versatile. The Lenosky potential (LP) is a derivative of the MEAM class of potentials that was specifically optimized for silicon [21]. Finally, the so-called environment-dependent interatomic potential (EDIP) [22,23] represents another formulation in the bond-order class of potentials.

The most common metrics for potential validation include crystal phase stability order, elastic constants, surface energetics and reconstructions, point defect thermodynamics and structure, dislocation and stacking-fault energetics, and amorphous/liquid phase structure [14,16] and melting characteristics [11]. Finally, although isolated clusters have also been considered in potential evaluation, the energetics of such structures are generally quite poorly reproduced unless the potential parameters are specifically tuned using these highly undercoordinated structures [20]. There is no single clear winner among silicon potentials and each has been shown to have strengths and weaknesses, some of which may be traced to the database that was used to establish the parameters. Even different parameterizations of the same potential function have been shown to provide rather different results.

Before proceeding with a comparative analysis of point defect formation energies, which we focus on here, it is important to note that empirical potentials cannot account for different point defect charge states. As mentioned previously, the lowest energy configuration for a given point defect is not necessarily the electrically neutral one, making a comparison between DFT and empirical potential results somewhat ambiguous. Here, we take the point of view that an empirical potential is a coarse-grained representation of atomic bonding in which the electronic states have been somehow (not formally) "averaged out." In this view, comparisons between empirical potential and DFT predictions for point defect formation energies should be made on the basis of the lowest energy charge state at zero temperature, and a Boltzmann-weighted average of the different charge states at finite temperatures. In other words, there is nothing in the empirical potential formulation to necessarily suggest that the "equivalent" configuration in the DFT case is the neutral state.

A representative list of empirical potential predictions for the single vacancy formation energies at zero temperature is provided in Table 3.1. Apart from some outliers, there is general consensus in the range of $E_V^f \sim 3 - 4$ eV, in good agreement with the DFT range 3.2–3.6 eV (see Section 3.2). In all cases, the basic ground state configuration is a missing atom with some localized relaxation of the surrounding atoms. Note that few of these studies report the precise configurational details of the relaxed vacancy such as inward/outward relaxation, symmetry, and bond angles and lengths of neighboring atoms. This is largely due to the fact that such high-resolution information is not likely to be meaningfully reproduced by most empirical potentials. Again, the notion that an empirical potential function represents some kind of a coarse-grained representation of the full bonding environment can be invoked to suggest that empirical potentials may be good at describing certain, but not all, features of a point defect configuration.

Table 3.1 Zero-Temperature Vacancy Formation Energies Predicted by Selected Empirical Potentials for Silicon

Potential Model	Vacancy Formation Energy (eV)	Ref.
SW	2.59, 2.82	[5,24]
SW-Pizzagalli	3.27	[13]
BH	2.12	[5,8]
Tersoff (T3)	3.70	[5]
Tersoff-ARK	3.70	[11]
MEAM-Baskes	3.19	[25]
MEAM-Timonova	3.14	[20]
MEAM-Lenosky	3.30	[21]
EDIP	3.22	[21]
Bond-order	2.76	[26]

Table 3.2 Selected Zero-Temperature Formation Energies for Ground State Configurations of Self-Interstitials, as Predicted by Selected Empirical Potentials for Silicon

Potential Model	Self-interstitial Formation Energy (eV)	Lowest Energy Configuration	Ref.
SW	3.61	<110>-dumbbell	[24]
SW-Pizzagalli	5.92	<110>-dumbbell	[13]
BH	1.56	Tetrahedral	[5,8]
Tersoff (T3)	3.45	Tetrahedral	[5,27]
Tersoff-ARK	2.20	Tetrahedral	[11]
MEAM-Baskes	4.81	Tetrahedral	[25]
MEAM-Timonova	3.84	Tetrahedral	[20]
MEAM-Lenosky	3.00	Tetrahedral distorted	[21]
EDIP	3.35	<110>-dumbbell	[21]
Bond-order	2.64	Tetrahedral	[26]

The assessment of self-interstitial formation energies is more nuanced, where several distinct configurations may exist in a relatively narrow energy range. The most reported configurations include the tetrahedral, hexagonal bond-centered, and some form of the <110>-dumbbell configuration. Shown in Table 3.2 are formation energies for the reported lowest-energy configurations found for several empirical potentials. It should be noted that, unlike the vacancy case, significant "scatter" exists in the reported self-interstitial formation energy values for a given configuration and with a given potential model. In addition, the lowest-energy configuration is not always the same across potentials, with the tetrahedral and <110>-dumbbell geometries most often being reported as the ground state configurations. These observations are likely to be at

least partially due to the presence of multiple closely spaced local minima that correspond to different interstitial configurations and even different relaxations of a single configuration. In fact, as we will show in Section 3.5, the notion of a "rough energy landscape" with multiple local minima is a generic feature that introduces a significant amount of configurational entropy at elevated temperature, necessitating a somewhat different approach for thinking about finite temperature point defect and cluster thermodynamics.

Comparison of the self-interstitial formation energy predictions in Table 3.2 to the results of DFT calculations, whose consensus gives a range of about 3.3–3.8 eV for the <110>-dumbbell, highlights the limitations of using (currently available) empirical potentials for quantitative estimation of point defect parameters—it is not yet possible to do this without strict a posteriori validation of the results against more accurate approaches. More importantly, the potentials that appear to provide the best agreement with specific DFT calculations were fitted to the DFT results, and the quantitative accuracy of the potential for any other properties therefore is in question (until explicitly tested). Nonetheless, the most commonly employed potentials such as SW, Tersoff and EDIP all appear to provide good estimates in the range of 3.3–3.9 eV.

3.3.2 Quantum Mechanical Estimates for Point Defect Thermodynamics

Atomistic simulations that explicitly consider electronic interactions between ions, frequently referred to as ab initio or *first-principles* calculations, are indisputably the gold standard for theoretically estimating silicon defect thermodynamic and transport properties. The vast majority of quantum mechanical atomistic simulations are based on the application of density function theory (DFT), in which the many-electron problem is reduced to that of a single electron moving in an effective potential specified either through the local density approximation (LDA) or generalized gradient approximation (GGA) [28]. More accurate treatments of the many-electron problem are available (e.g., quantum Monte Carlo and Hartree-Fock), but these are currently very infrequently applied to defects in silicon because of their extreme computational intensity [29].

Not only are DFT calculations obviously important in their own right, but they are also crucial for (1) parameterizing empirical potentials for silicon, and (2) validating the predictions of these empirical potentials by providing reference calculations. On the other hand, despite the relative efficiency of DFT, such calculations still remain *much* more computationally intensive than corresponding empirical potential simulations. This is particularly true when considering larger defects such as clusters and when free energies at finite temperature are required. The computational expense of DFT has led to continued interest in simplified quantum mechanical approximations, such as the tight-binding method [30], in addition to empirical potential development. Although tight-binding simulations have in some instances been shown to be demonstrably better than some empirical potentials [31], it is difficult to unequivocally make the case that

their added computational expense necessarily improves accuracy relative to empirical potentials.

A detailed treatise of quantum mechanical models is well beyond the scope of this chapter. Instead, the aim here is to provide a brief discussion of the principal outcomes and outstanding issues related to such calculations, specifically in the context of point defect formation thermodynamics in silicon. The history of silicon point defect formation energy calculations with electronic DFT dates back over three decades. Arguably, the formation energy and relaxation structure of the neutral single vacancy have served as the principal benchmarks during this period [32–43]. The reason for this stems from the deceptive simplicity of the calculation; indeed, definitive convergence criteria for the calculation have only recently emerged [41–43]. Remarkably, the numerous reported DFT calculations of the neutral vacancy formation energy have ranged from over 4 eV to just below 3 eV. Even the qualitative nature of the surrounding atomic relaxation (inward vs. outward, symmetry) has been the subject of debate. The reasons for the long-standing challenge associated with this particular calculation are in fact quite straightforward. First, several choices and parameter selections must be made when performing any DFT simulation, some of which are constrained by computational expense. These choices include (i) supercell size, (ii) exchange-correlation functional, (iii) basis set type (plane waves or orbitals) and energy cutoff, and (iv) Brillouin zone integration mesh. As shown recently in refs. [41,43], the particular combination of these parameters and functionals can lead to errors that interact in nonlinear ways and make it difficult to perform convergence tests without a complete (and very expensive) parametric analysis. For example, it appears that a system size of about 256 atoms is required to sufficiently isolate the neutral vacancy and allow it to relax into the D_{2d} symmetry via the expected Jahn-Teller distortion [43]. This is not surprising given the long-ranged elastic relaxation fields around vacancies found in both DFT and empirical potential simulations [44,45]. In addition, electrostatic and wave–function interactions across the periodic boundaries are likely to play roles in the slow convergence with respect to system size. One example of the nonlinear coupling among the various parametric selections was observed in Ref. [43] in which a sparser sampling of the Brillouin zone was more likely to exhibit the correct relaxed vacancy symmetry when the system size was small.

The long history of convergence notwithstanding, there is little doubt that the state-of-the-art DFT calculations have now reached the point where they can be reliably used to compute a variety of point defect and small cluster thermodynamics. For the neutral vacancy formation energy, the range between 3.2 and 3.6 eV almost certainly includes the “correct” value. The neutral self-interstitial is much less well studied but the <110>-split dumbbell configuration is well established as the ground state with the formation energy range 3.3–3.8 eV [36,46–48].

Note that the neutral charge state is simply used here as a basis for discussion—it is not necessarily the charge state with the lowest formation energy. In fact, under extrinsic conditions corresponding to finite doping levels, charged states are usually lower in formation energy and may even possess different structures than the neutral

configuration. This poses some ambiguity in the comparison of defect structures and energies between DFT and empirical potential calculations. While it is customary to compare to the neutral defect structure when validating empirical potential predictions, one may also consider the structures predicted by empirical potentials as coarse-grained entities that represent some kind of "average" over the various charge states. In other words, given the complete absence of explicit charge consideration in most empirical potentials for silicon, the use of the neutral configuration as a reference is not necessarily meaningful—one could just as well use the lowest energy configuration, irrespective of its charge, as a basis for validation.

3.4 Ground State Point Defect Cluster Thermodynamics

There are far fewer atomistic simulation studies of point defect cluster thermodynamics due to the larger simulation cells and longer relaxation times required. The principal difficulty in computing cluster thermodynamics is the identification of the ground state configuration. For quantum mechanical calculations, in particular, a good initial guess for the cluster structure is required before static relaxation can be applied. Nonetheless, certain features of small vacancy and self-interstitial clusters are now well established. For example, both types of clusters exhibit magic sizes that are particularly stable relative to neighboring sizes [49–51]. In addition, both cluster species are characterized by overall decreasing trends in their per-point defect formation energies. The latter trend is a consequence of the ability of larger clusters to achieve reconstructions that are not possible for very small ones.

3.4.1 Vacancy Clusters

Vacancy cluster ground state configurations are relatively straightforward to derive on the basis of broken-bond minimization. Using this approach, Chadi and Chang [52] were able to predict magic sizes for clusters up to size 12. The 6-vacancy hexagonal ring structure and 10-vacancy "adamantine cage" configuration were found to be the key building blocks for all cluster sizes, see Figure 3.1. These conclusions were subsequently supported by quantum mechanical (LDA-DFT) [53], tight-binding [50], and empirical potential (EDIP) atomistic calculations [51].

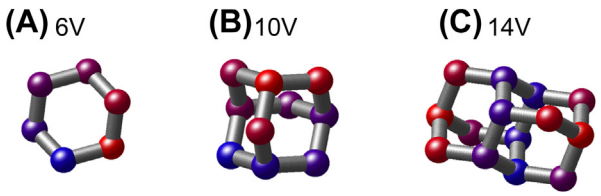


FIGURE 3.1 Ground state configurations of vacancy clusters generated on the basis of closed ring and adamantine cage configurations. From left: (A) 6-vacancy ring, (B) 10-vacancy adamantine cage, and (C) 14-vacancy cluster comprised entirely of closed rings and cages. All three configurations correspond to specially stable "magic" sizes.

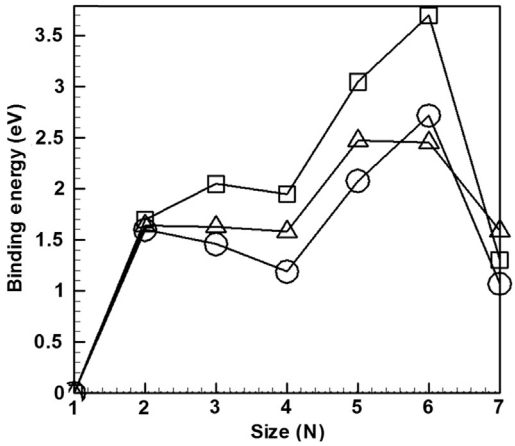


FIGURE 3.2 Binding energies as a function of size for vacancy clusters computed using LDA-DFT (squares), tight-binding (circles), and the empirical EDIP potential (triangles). Binding energies, E_n^b, are defined according to the relationship $E_n^b = E_{n+1}^f - (E_n^f + E_1^f)$, where E_n^f is the formation energy of a cluster of size n.

Shown in Figure 3.2 are zero-temperature binding energies for vacancy clusters up to size seven for the three different interaction models. Overall, the agreement between the three representations is very good, clearly showing the special stability of the 6-vacancy hexagonal ring configuration. The agreement across calculations also demonstrates the validity of using empirical potentials such as EDIP for defect thermodynamic property estimation. More recently, divacancy and hexavacancy formation energies computed with GGA-DFT and several empirical potentials (SW, EDIP and Tersoff) were compared and once again shown to be quite consistent across the various potentials [49].

Calculations for larger cluster sizes with tight-binding and EDIP potentials also show similar special stability for sizes 10, 14, and 18—all sizes that correspond to completed hexavacancy rings and/or adamantine cages [45,51,54]. Interestingly, the hexavacancy ring and adamantine cage building blocks naturally lead to octahedral cluster shapes bounded by (111) planes at larger cluster sizes—the octahedral geometry being the most common experimentally observed one for large vacancy aggregates present in CZ-grown silicon crystals [1]. The formation energies for vacancy clusters that follow the octahedral motif have been calculated using the EDIP potential for sizes as large as 1000 [45,51,54]. The overall vacancy cluster energetics were found to be described very well by a power-law function that scales as $n^{2/3}$, where n is the number of vacancies in the cluster. In other words, vacancy clusters, as expected, are energetically defined by their surfaces. Deviation from this power-law scaling was only observed for very small clusters, which are unable to arrange into closed rings and cages. Approximating the cluster surface area by assuming a spherical shape, the EDIP potential was found to predict a zero-temperature effective surface energy of 1.24 J/m^2. This value is in excellent agreement with several experimental measurements of the Si(111) surface energy at cryogenic temperatures [55,56].

Finite temperature formation free energies (including vibrational entropy) for vacancy clusters computed with the EDIP potential suggest that the overall trends observed with the zero-temperature calculations are preserved with increasing temperature [57]. Near the melting temperature, the effective surface free energy including vibrational entropy is reduced to about 0.9 J/m^2. Unfortunately, the validity of this estimate is more difficult to verify against experimental measurements, although one indirect experimental measurement places the melting temperature (111) surface energy right around 0.89 J/m^2 [58]. The issue of effective void surface energies at finite temperature is revisited more comprehensively in Section 3.6, where the influence of configurational entropy is also included.

3.4.2 Small Compact Self-Interstitial Clusters

The analysis of self-interstitial clusters is complicated by bonding reconstructions associated with the presence of multiple self-interstitials in close proximity. While vacancy clusters are characterized by the octahedral morphology across all sizes, no single morphological motif describes self-interstitial clusters. Nonetheless, several broad features of self-interstitial clusters are now well established both experimentally and by various types of calculations. Small clusters containing up to about 15 self-interstitials are compact and three-dimensional and exhibit magic sizes, particularly at sizes that are integer multiples of four (i.e., 4, 8, and 12) [46,50,59]. Evidence for special stability of these sizes has been demonstrated not only theoretically but also experimentally. Cowern et al. [60], in particular, used inverse modeling of boron diffusion profiles to extract effective formation free energies for interstitial clusters as a function of size and found that the best agreement was obtained when 4- and 8-interstitial clusters were assumed to be particularly stable relative to adjacent sizes. Larger interstitial clusters are observed experimentally in several types of planar structures including {113}-oriented rod-like defects, {111} partial and perfect dislocation loops, and possibly also {100}-oriented plate-like structures; example studies include refs. [61–72]. These structures are most commonly observed in ion-implanted samples. Evidence for the {100} planar defects is not as well established as for the other types of planar defects. The transitions between small compact clusters and the various larger planar structures are still not fully understood, and the dependence of these transitions on temperature and other variables (e.g., local stress) even less so.

Here, we summarize a selection of theoretical estimates for the ground state formation thermodynamics of self-interstitial clusters and defer discussion of finite temperature thermodynamics to Section 3.5. A large number of computational studies of interstitial clusters have focused specifically on the 4-interstitial (I_4) cluster, which is the smallest magic cluster size. The ground state structure predicted by most of these studies consists of 5- and 7-membered bond rings with no dangling bonds; see Figure 3.3 (taken from Ref. [73]). We henceforth refer to this structure as the Humble/Arai configuration after refs. [74,75]. Various estimates for the formation energy of this cluster are

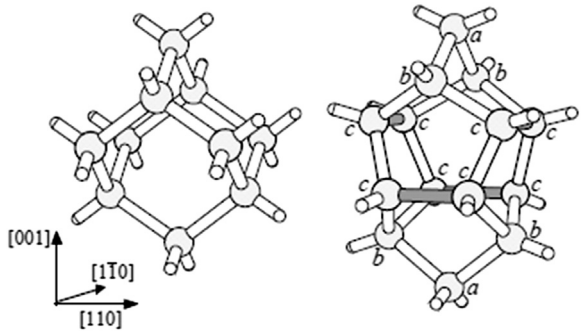


FIGURE 3.3 Schematic representation of the 4-interstitial cluster (right) relative to the perfect silicon lattice (left). The four [001] split–interstitial pairs are represented by atoms labeled "c" and the reconstructed bonds which link the interstitial pairs are shaded. *Taken from Ref. [73].*

Table 3.3 Ground State 4-Interstitial (I_4) Formation Energies Predicted by Various Interaction Models

Model	I_4 Formation Energy (eV)	Ref.
LDA-DFT	8.7	[73]
LDA-DFT	5.96	[76]
LDA-DFT	7.28	[46]
GGA-DFT	7.40	[46]
Tight-binding	9.41	[50]
Tight-binding	9.84	[75]
Empirical SW	8.40	[75]
Empirical EDIP	8.75	[77]

summarized in Table 3.3. With few exceptions, the formation energy estimates are remarkably consistent across the various studies, and the per-interstitial formation energy is tightly clustered in the range 1.8–2.4 eV/atom. It is notable, however, that not all studies show the I_4 to be a magic size. For example, the tight-binding calculations in Ref. [50] and the LDA-DFT results in Ref. [48] both indicate that the per-interstitial formation energy varies monotonically with size and that I_4 is no more stable than adjacent sizes. Moreover, the tight-binding results generally lead to systematically higher formation energies than the DFT predictions.

Shown in Figure 3.4 are formation energies per interstitial for a range, $2 \leq n_I \leq 9$, of small compact self-interstitial clusters computed using the EDIP empirical potential and two different sets of DFT calculations [46,48,77]. Overall the agreement between the three sets of calculations is quite good. The EDIP formation energies are higher than both of the DFT results although the difference in the two sets of DFT predictions is of similar magnitude as the difference between the EDIP results and those in Ref. [46].

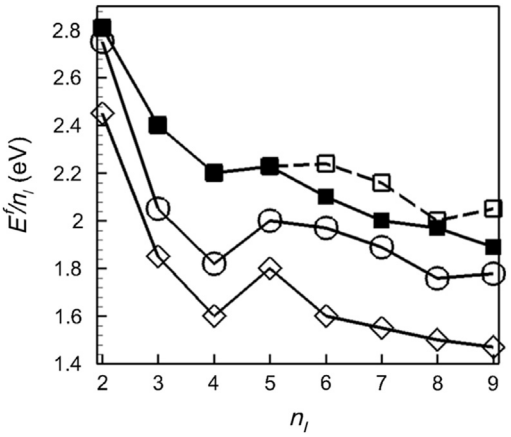


FIGURE 3.4 Formation energies per interstitial as a function of cluster size, n_I. Squares—EDIP results for compact (open) and elongated (filled) [77]; circles—DFT [46]; diamonds—DFT [48].

While the special stability of the 4-interstitial cluster (Humble/Arai configuration) is found in all three studies, the picture is somewhat murkier for the 8-interstitial cluster. The interpretation of the formation energy calculations is complicated by the fact that two different morphologies for $n_I \geq 5$ can be identified. The first, which is denoted as "compact" is based on the Humble/Arai motif; thus a 5-interstitial cluster is a Humble/Arai 4-mer plus an additional interstitial and so on. The second structural motif, "elongated", refers to [110]-aligned configurations that serve as the building blocks for the {113}-oriented planar configurations (see next section).

The relative stability of compact and elongated configurations in the size range $n_I < 12$ is not conclusive: Refs [77] and [46] seem to indicate that the elongated structures are more stable but ref. [48] finds the opposite trend. In either case, the difference is not large and the 8-interstitial cluster is only weakly favored over neighboring sizes, if at all. A possible explanation for this apparent discrepancy with the experimental inference in Ref. [60] is provided in Figure 3.5. Here, cluster formation free energies that include vibrational entropy are estimated using the EDIP potential. Interestingly, the vibrational entropies are such that the free energies of compact and elongated configurations become almost identical. Moreover, for the compact 8-interstitial cluster, which is comprised of two Humble/Arai 4-interstitial cluster, a rather large configurational entropy for the compact structure also is suggested due to the large number of (almost degenerate) ways that the two 4-interstitial clusters can be placed [77]. With these entropic factors included, strong "magicness" at size eight emerges. These results suggest that the consideration of entropic contributions might be necessary to fully understand cluster thermodynamics and that conclusions based on ground state configurations may be misleading or at least incomplete. In Sections 3.5 and 3.6, the calculation of entropic contributions is generalized and further evidence is provided for the importance of entropy in defect thermodynamic analysis.

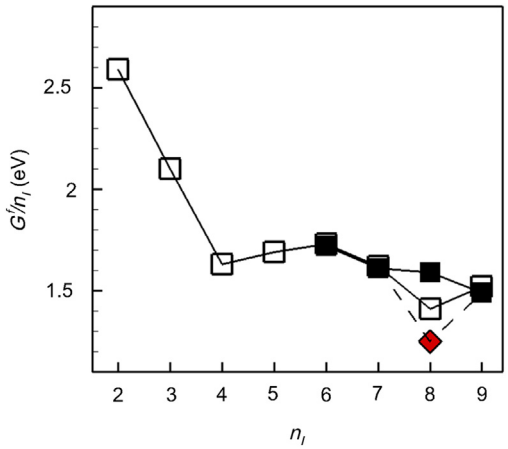


FIGURE 3.5 EDIP-predicted formation free energies as a function of interstitial cluster size at 1100 K (EDIP melting temperature is 1520 K). Open squares—compact structures; filled squares—elongated structures. Single diamond symbol shows free energy including estimated configurational entropy for 8-interstitial compact cluster. *Adapted from Ref. [77].*

3.4.3 Large Planar Self-Interstitial Clusters

Formation energy estimates have also been computed for the various types of planar defects found in ion-implanted silicon wafers. The most intensively studied is the {113}-oriented family of planar structures, which are characterized according to how densely interstitial chains are packed along the [332] direction. Briefly, the notation /I/ represents a sequence of adjacent self-interstitial chains (highest density packing) while /IO/ and /IIO/ represent sequences along the defect (in the [332] direction) in which some chains are missing—a missing chain is denoted by "O"; see Ref. [62] for more notational details. A summary of per-interstitial formation energies for the {113} family of planar is shown in Table 3.4 assuming infinitely long interstitial chains (in the [110] direction). It is notable that the lowest energy configuration on a per-interstitial basis is not the densest (/I/) but rather one that consists of about 66–75% occupancy of interstitial chains along the [332] direction, i.e., /IIO/ or /IIIO/.

It has been suggested that {113}-oriented defects grow and eventually transform into {111}-oriented dislocation loops, namely Frank partial loops (FDLs) and perfect dislocation loops (PDLs). There is also some direct experimental evidence for such a transformation [66]. The transformation size has been estimated to be approximately 20 nm, or when clusters contain on the order of 500 self-interstitials [67]. The formation energies for large {111}-oriented planar defects can be estimated on the basis of continuum mechanical arguments because of the relatively simple structures. For small loop sizes, FDLs have lower per-interstitial formation energy than PDLs but the trend is reversed when the loop radius is greater than about 40 nm [79]. The limiting FDL formation energy is 0.027 eV/atom (the stacking-fault energy), while PDL formation energies tend

Table 3.4 Ground State Formation Energies for {113}-Oriented Planar Self-Interstitial Clusters

Model	Configuration	Formation Energy (eV/atom)	Ref.
LDA-DFT	/IIO/	0.68	[75]
SW	/IIO/	0.88	
LDA-DFT	/IO/	0.49	[76]
LDA-DFT	/I/	0.63	[68]
	/IO/	0.72	
	/IIO/	0.46	
	/IIIO/	0.48	
Tight-binding	/IO/	1.35	[78]
LDA-DFT	/I/	0.71	[48]
	/IO/	0.76	
	/IIIO/	0.55	

to zero (on a per-interstitial basis) in the limit of large loop size. These very small energies suggest that if sufficient self-interstitials are present to allow growth and coarsening to produce large cluster sizes, dislocation loops will be the predominant structures. Claverie et al. [67,70,79–81] have summarized comprehensively self-interstitial cluster morphologies as a function of annealing history in a series of papers.

3.5 Inherent Structure Theory and Potential Energy Landscapes

As described in the previous section, the "conventional" strategy for computing point defect (or cluster) free energies is to locate the ground state (minimum energy) configuration, compute its formation energy and vibrational entropy, and if possible estimate a degeneracy factor based on symmetry considerations. These elements are then combined in Eqn (3.2) or (3.4) to provide the "open-system" equilibrium concentration for that particular defect. The implicit assumption in this picture is that no other configurations of the defect are energetically close to the ground state and thus do not contribute in any way to the stability of the defect. In this section, a framework is described for including the contributions of higher-energy configurations into the defect free energy. It will be shown that the configurational entropy contribution from higher formation energy configurations can be significant at the high temperatures associated with silicon crystal growth and wafer annealing. As a result, defect free energies based on ground state analyses can overestimate the true free energy and, in some cases, lead to qualitatively incorrect predictions for defect thermodynamics.

In the following, the notion of the Inherent Structure Landscape (ISL) is used to develop a quantitative theory for describing high-temperature defect thermodynamics. While ISL theory has since been applied successfully to a variety of systems, including

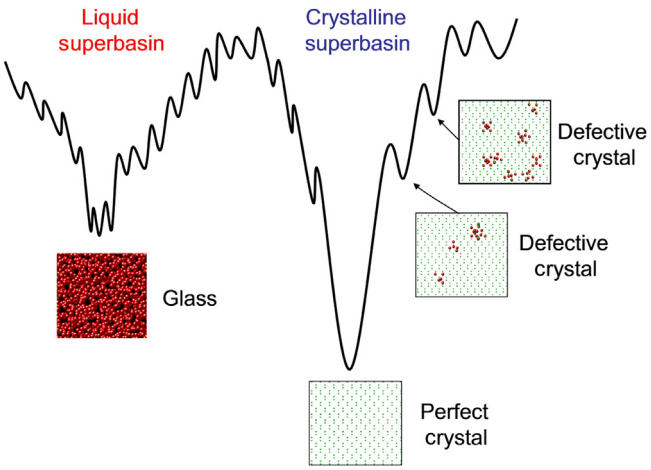


FIGURE 3.6 One-dimensional schematic representation of the potential energy landscape for a model atomic material showing two "superbasins" corresponding to liquid/glass/amorphous solid and crystal macro-states. Within each superbasin, many local minima correspond to distinct inherent structures—the global minimum corresponds to the perfect crystal configuration. Insets show example configurations in which large red spheres qualitatively correspond to disordered (i.e., noncrystalline) atoms and small green ones denote atoms in crystalline positions. *Taken from Ref. [92].*

atomic clusters [82–84], glasses [85–87] and polymers [83,88], it has not received much attention in the arena of crystal defect thermodynamic analysis since the early work of Stillinger and Weber [89]. Inherent structures, as first introduced by Stillinger and Weber [89,90], correspond to local minima in the 3N-dimensional potential energy (or enthalpy) landscape defined by the three-dimensional coordinates of an N-atom system [91]. Shown in Figure 3.1 is a schematic one-dimensional projection of a potential energy landscape for some hypothetical material. On the left is a "superbasin" that contains all micro-configurations that are macroscopically noncrystalline, including the liquid phase (higher energy) and amorphous solid phase (lower energies). On the right is the crystal superbasin that includes the perfect crystal ground state and higher-energy levels that correspond to configurations that include defects. In all cases, "inherent structures" correspond to mechanically stable configurations and are represented by local minima in Figure 3.6. In turn, each local minimum is surrounded by a basin that defines the region of phase space from which a local minimum is always reached upon downhill energy minimization (e.g., steepest descent or conjugate-gradient) [82].

Under certain conditions, the total phase space of the system is well approximated by the collection of basins surrounding inherent structures, i.e., the system spends most of its time in the vicinity of one or more of the inherent structures and very little time in transition between them. Under these conditions, the partition function that describes the system becomes much simpler to approximate and enables a direct route for estimating various thermodynamic properties. In the following section, a summary of the

ISL framework is provided in a form that is appropriate for the analysis of defect thermodynamics in crystalline materials.

3.5.1 Inherent Structure Landscape (ISL) Theory for Defects in Crystals

Consider the constant particle-volume-temperature (NVT) statistical ensemble, in which the Helmholtz free energy is related to the classical canonical partition function, Z, by the relation

$$\beta F = -\ln Z(N, \beta, V), \tag{3.11}$$

where $\beta = 1/kT$, V is the system volume, and

$$Z = \frac{1}{N!}\frac{1}{\Lambda^{3N}}\int \exp\left(-\beta E\left(\mathbf{r}^N\right)\right) d\mathbf{r}^N. \tag{3.12}$$

In the above equation, Λ is the thermal de Broglie wavelength and $E(\mathbf{r}^N)$ represents the potential energy of the N-atom system. Within the ISL approximation, the partition function is rewritten in terms of an integral over inherent structure basins so that [90]

$$Z = \frac{1}{\Lambda^{3N}}\int g(E_\alpha)\exp(-\beta E_\alpha)\exp(-\beta F_{vib}(\beta, E_\alpha))dE_\alpha, \tag{3.13}$$

where E_α is the (local ground state) potential energy of inherent structure α and $g(E_\alpha)$ is the configurational density-of-states (DOS), or degeneracy function, for the distribution of basins within the landscape. In other words, $g(E_\alpha)$ is the number of basins present that possess minimum potential energy E_α. Note that the configurational degeneracy function is a property of the potential energy landscape and is independent of temperature. The vibrational free energy, $F_{vib} = -TS_{vib}$, of each basin is a measure of the basin size where the vibrational entropy of a basin, $S_{vib} \equiv klnN_{vib}$, can be regarded as representing the number of vibrational states contained in that basin.

We now define a combined vibrational and configurational degeneracy in a single function, i.e.,

$$G'(\beta, E_\alpha) \equiv g(E_\alpha)\exp(-\beta F_{vib}(\beta, E_\alpha)). \tag{3.14}$$

Noting that the perfect crystal has unit configurational degeneracy, i.e., $g(E^P) = 1$ where E^P is the perfect crystal energy, the combined DOS function may be expressed in terms of formation energies, i.e.,

$$G''(\beta, \Delta E_\alpha) \equiv \frac{G'(\beta, E_\alpha)}{G'(\beta, E^P)} = g(E_\alpha)\exp(-\beta \Delta F_{vib}(E_\alpha)), \tag{3.15}$$

where $\Delta E_\alpha \equiv E_\alpha - E^P$ is the formation energy of basin α and

$$\Delta F_{vib}(E_\alpha) \equiv F_{vib}(\beta, E_\alpha) - F_{vib}\left(\beta, E^P\right), \tag{3.16}$$

is the formation vibrational free energy. Note that all reference perfect crystal properties are scaled to refer to the same number of atoms as in the defective crystal. Within the

harmonic approximation [93], the combined DOS function in Eqn (3.15) can be further simplified into the form [92,94]

$$G''(\beta, \Delta E_\alpha) = G(\Delta E_\alpha)\Omega^P(\beta), \tag{3.17}$$

where

$$\Omega^P(\beta) = G'\left(\beta, E^P\right)\exp\left(-\beta E^P\right), \tag{3.18}$$

is a temperature-dependent constant defined only on the perfect crystal configuration. The DOS distribution, $G(\Delta E_\alpha)$, is now independent of temperature and will be used as the basis for computing defect thermodynamic properties in the following sections. Using Eqns (3.17) and (3.18), the partition function in Eqn (3.13) can be written entirely in terms of formation energy as

$$Z = \frac{\Omega^P(\beta)}{\Lambda^{3N}} \int G(\Delta E_\alpha)\exp(-\beta \Delta E_\alpha) d\Delta E_\alpha. \tag{3.19}$$

Finally, the formation (Helmholz) free energy for a defect is given by

$$\Delta G = -k_B T \ln \int G(\Delta E_\alpha)\exp(-\beta \Delta E_\alpha) d(\Delta E_\alpha). \tag{3.20}$$

The above formalism readily can be extended to the isobaric-isothermal (NPT) ensemble which is characterized by an *enthalpy* landscape [95,96]. The key result is that the isothermal-isobaric partition function can be written in an analogous form to Eqn (3.13), i.e.,

$$Y(N, P, T) \propto \int g(H_\alpha)\exp(-\beta H_\alpha)\exp\left(-\beta \tilde{F}_{vib}(\beta, H_\alpha)\right) dH_\alpha \tag{3.21}$$

where P is the pressure, H_α is the enthalpy of inherent structure α, and $\tilde{F}_{vib}(\beta, H_\alpha)$ is the vibrational free energy of basin α within the NPT ensemble. Similar considerations used to derive Eqn (3.19) can be applied to give

$$Y \sim \int G(\Delta H_\alpha)\exp(-\beta \Delta H_\alpha) d\Delta H_\alpha. \tag{3.22}$$

In Eqn (3.22), the formation enthalpy is calculated as $\Delta H_\alpha = \Delta E_\alpha + P\Delta V_\alpha$, where $\Delta V_\alpha \equiv V_\alpha - V^P$ is the formation volume of a particular configuration that corresponds to basin α relative to the volume of the perfect crystal, V^P.

3.5.2 Sampling Inherent Structures with Molecular Dynamics Simulation

In order to apply ISL theory to defect thermodynamics, the potential energy landscape (PEL) must be appropriately sampled. Standard molecular dynamics (MD) is used to perform all PEL sampling described in this chapter, although any suitable variant of the Monte Carlo method is equally applicable. For example, prior studies have employed a number of highly efficient methods, notable examples being the basin hopping

technique [82] and the minima hopping method [97]. Although certainly not the most efficient approach, the principal advantage of direct MD sampling is that it automatically accounts for all configurational and vibrational degrees of freedom.

The procedure for locating potential energy basins for the various defect configurations is as follows. First, the defect is inserted into the simulation cell and a short MD simulation is used to relax the structure before data is collected. Next, the probability distribution of the inherent structure minimum energies is accumulated by periodically taking snapshots of the MD trajectory and quenching them using conjugate-gradient energy minimization (or any other quenching method). Here, minimization was performed every 200 MD time steps. The result from each quench gives the local inherent structure formation energy, ΔE_α, of the current basin. A formation energy (or enthalpy) histogram is then collected into energy/enthalpy bins (width is typically 0.1–1.0 eV). The result of this procedure is a temperature-dependent probability distribution of basin energies, which is related to the configurational-vibrational density-of-states distribution by [87,88,98]

$$P(\Delta E, \beta_{sim}) = G(\Delta E)\exp(-\beta_{sim}\Delta E) \tag{3.23}$$

where β_{sim} is the temperature at which the simulation is performed. Note that in Eqn (3.23) and in all subsequent discussion, the subscript "α" is omitted for notational brevity.

In the following sections, all MD simulations for PEL sampling were carried out in either the constant atom-volume-temperature (NVT) or the constant atom-pressure-temperature (NPT) ensembles using the EDIP empirical potential unless otherwise stated. It should be noted here that there is recent evidence [99] that potential energy landscapes predicted by empirical potentials such as EDIP tend to be significantly rougher than landscapes generated by more accurate representations such as DFT and tight-binding potentials. While such discrepancies, which are not yet fully characterized, can alter the quantitative estimates of entropic contributions obtained from inherent structure theory, that the qualitative conclusions generated on the basis of empirical potentials are useful for understanding the mechanisms of high-temperature microstructural evolution.

The LAMMPS software [100] was used for all MD simulations and conjugate-gradient energy minimizations. In the case of NVT simulations, the system volume was chosen using short NPT simulations to provide the desired value of the hydrostatic pressure. MD time steps of 1.0 fs were used to integrate the particle trajectories. Generally, O(10^5) inherent structures are needed to produce a converged probability distribution at a specified temperature. The exact number of inherent state visits required for convergence depends on the least visited states within the distribution; the 10^5 estimate is sufficient to converge distributions down to probabilities in the 10^{-3}–10^{-4} range (with increasing error at lower probabilities). Convergence was checked by comparing distributions as the number of inherent structure samples increased.

3.6 High Temperature Defect Thermodynamics

In this section, the ISL framework outlined in Section 3.5 is used to compute point defect and cluster thermodynamics at finite temperature. The single vacancy is considered first to illustrate the key features of the theoretical framework followed by vacancy and self-interstitial cluster analyses.

3.6.1 The Single Vacancy

The inherent structure probability distribution for the single vacancy near the melting temperature (1600 K) was computed using the EDIP potential according to the method outlined in Section 3.5.2. Shown in Figure 3.7(A–F) are example mechanically stable structures with increasing formation energy. Several configurations (e.g., (a), (b), and (d))

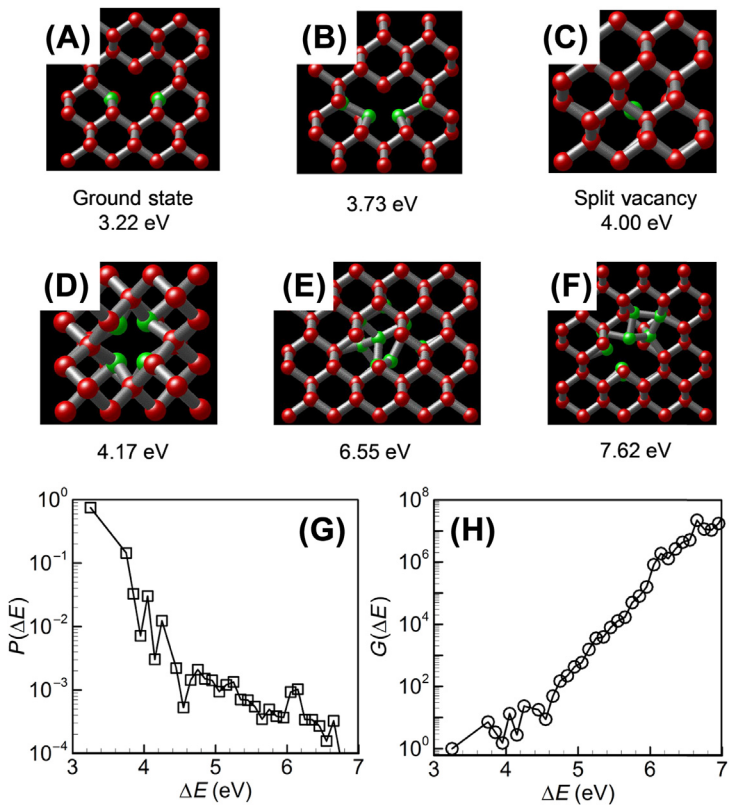


FIGURE 3.7 (A–F) Example inherent structures of the single vacancy including the ground state (A), the split vacancy (C) and several other higher energy configurations. Red spheres denote silicon atoms that are within a threshold distance of their perfect lattice positions, while green spheres correspond to atoms that are displaced by more than 3% of a bond length. (G) Probability density distribution (PDF) for the EDIP vacancy; (H) Corresponding density-of-states (DOS) distribution.

are qualitatively similar but, in fact, correspond to completely distinct relaxations of the atoms surrounding the vacant site. Higher energy configurations (e.g., (e) and (f)) correspond to increasingly disordered structures.

The plots in (g) and (h) show the PDF and DOS for the vacancy. The PDF is normalized to unit probability, while the DOS, computed from the PDF using Eqn (3.23), is shifted so that $G(\Delta E) = 1$ for the ground state configuration. While the probability of observing the ground state is highest, the system quenches to other configurations approximately 20% of the time at the temperature considered here. Note that there is no fundamental limit on the formation energies that can be accessed in these simulations; the range of energies accessed is limited by the extent of sampling. In other words, larger inherent structure samples would show progressively higher energy (and rarer) configurations. More frequent access to even higher energy configurations can be achieved simply by increasing the simulation temperature. In fact, as with almost all simulations of crystalline systems with periodic boundary conditions, significant superheating is possible and much higher energy inherent structures can be found. While such configurations do not play much of a role at temperatures below the melting point, they are important in the context of melting [88,92,94,101].

Most significantly, application of Eqn (3.20) at the melting temperature of EDIP silicon results in a formation free energy that is about 20% lower than that obtained by only considering the ground state configuration. In other words, at or near the melting point of EDIP silicon the (open system) equilibrium concentration of single vacancies is actually about a factor of two higher than would be expected based only on the ground state. This result starkly demonstrates the potential impact of configurational entropy at high temperature and suggests that it must be considered when high-temperature properties are required. As will be shown in the following sections, this is even more the case for point defect clusters.

3.6.2 Vacancy Clusters

The extension of the preceding analysis to vacancy clusters is straightforward. In this case, several vacancies are placed in close proximity, the system allowed to equilibrate at some temperature, and periodic quenches applied to isolate inherent structures. The only additional consideration here is that any configurations that correspond to "broken" clusters, in which not all vacancies are connected to each other in a single cluster, are discarded because these are not relevant to the PDF or DOS distributions for the single cluster. Here, the atomic energy is used to identify atoms that are near vacancies, and then the Stillinger criterion [102] is used to assess whether or not they are connected.

Shown in Figure 3.8 are PDFs (A) and DOS distributions (B) for several vacancy cluster sizes at 1600 K. Most significantly, the PDFs (A) for clusters larger than size three become peaked at formation energies that are substantially higher than the ground state configuration. In fact, at the temperature considered here, the ground state

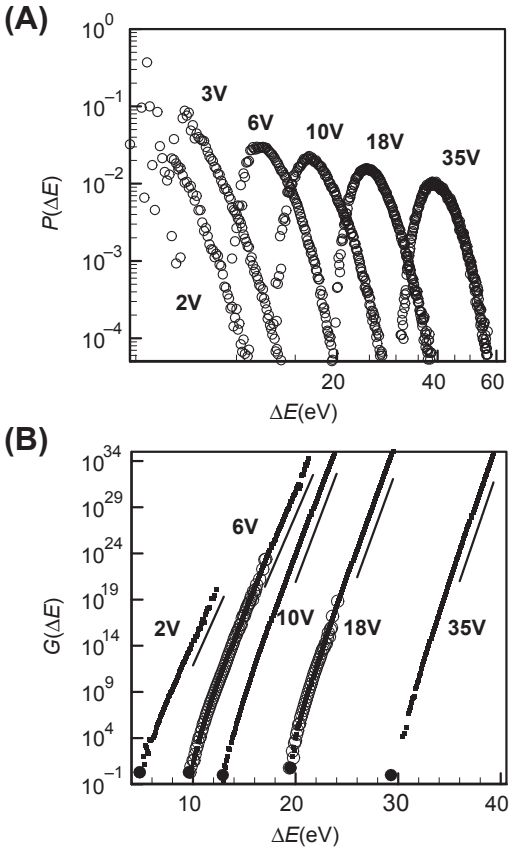


FIGURE 3.8 (A) Probability distribution function for EDIP vacancy clusters at 1600 K. (B) Selected corresponding anchored density-of-states functions—thick lines; exponential fits—thin short lines; density-of-states for hexagonal ring configurations (see Section 3.4) used to anchor the distributions—solid circles. *Adapted from Ref. [98].*

configurations are never observed. Simply, the enormous configurational and vibrational degeneracy present at higher energy levels overwhelms the energetic favorability of the ground state at elevated temperature. This phenomenon becomes progressively more pronounced for larger clusters, whereby the difference between the energies of the ground state and the most likely states increases with cluster size.

The PDFs in Figure 3.8, which have been normalized to unit area, are known only to within a multiplicative constant, which is required for computing cluster free energies of formation using Eqn (3.20). The process by which these constants are obtained is as follows. First, the PDFs are converted to the corresponding DOS distributions as shown in Figure 3.8(B). The DOS curves are formed using simulations at multiple temperatures in order to properly sample low energy states and in particular the ground states (shown by the large solid circles). Recall that the ground state configurations for vacancy clusters are derived from hexagonal rings and adamantine cages as shown in Figure 3.1. Once the

DOS distributions are obtained, the degeneracy for the bin containing the ground state (and only the ground state) is calculated. To do this, the vibrational spectrum for the ground state configuration is computed using the quasi-harmonic approximation (QHA) [103,104] or some equivalent approach. The configurational degeneracy of the ground state then is estimated using symmetry arguments (see configurations in Figure 3.1) and the product of the two contributions (vibrational and configurational) is used to obtain $G(\Delta E_{GS})$, where "GS" represents the ground state. The rest of the DOS is then shifted accordingly for each cluster size.

The impact of configurational and vibrational entropy on vacancy cluster free energies is shown as a function of temperature and size in Figure 3.9. Shown in Figure 3.9(A) are cluster free energies as a function of size and temperature expressed as effective surface energies, i.e., $\sigma = \Delta G(n,T)/2.224n^{2/3}$, where the numeric factor represents the surface area of a sphere with volume equal to n vacancies. The top surface in Figure 3.9(A) represents the free energy for the ground state configuration and includes the vibrational entropy which leads to the linear decrease of the free energy with

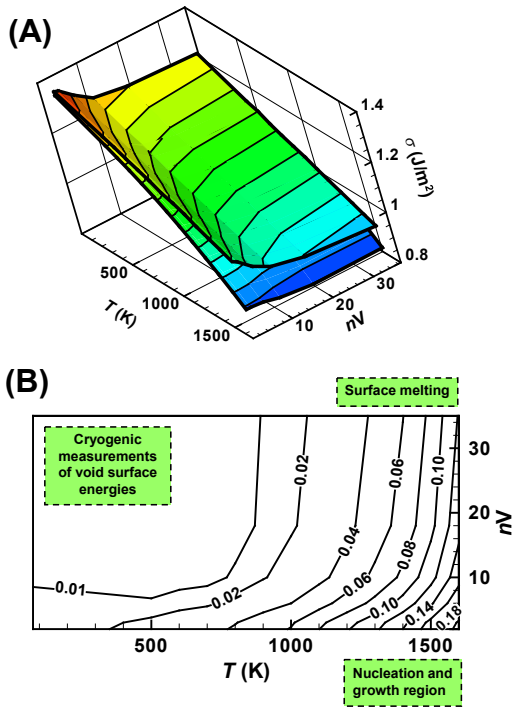


FIGURE 3.9 (A) Temperature–size dependence of the formation free energy for small vacancy clusters. The free energy is expressed as an effective surface free energy for convenience (see text). Top surface corresponds to ground state (GS) analysis (vibrational entropy only) while lower surface corresponds to ISL analysis (ISL) that includes both vibrational and configurational entropic contributions. (B) Normalized difference, $|\sigma_{GS} - \sigma_{ISL}|/\sigma_{ISL}$, between the two analyses. *Adapted from Ref. [98].*

temperature. The lower surface was obtained using Eqn (3.20) and the data shown in Figure 3.8, i.e., including both vibrational and configurational entropy. While the agreement between the two curves is excellent at low temperature, they diverge significantly at higher temperature due to the rapid increase in configurational entropy.

A more transparent view of the difference between the two results is shown in Figure 3.9(B) in which the normalized difference between the two free energies is plotted on a line plot. Several important points are highlighted. The largest difference is apparent in the lower right corner, corresponding to small clusters at high temperature. These clusters are significantly different in morphology from the expected compact octahedral structures observed experimentally (which are large and have cooled to low temperature). Instead of assuming the compact hexagonal ring and adamantine cage constructs depicted in Figure 3.1, these clusters tend to be in loosely bound, extended configurations with capture radii that are significantly larger than those of the compact configurations. Given that this small size-high temperature regime is crucial in the initial stages of void nucleation during silicon crystal growth, the omission of configurational entropy in any nucleation-growth model for voids is likely to cause large errors.

In the opposite corner, i.e., larger clusters at low temperature, the agreement between the ground state and ISL analyses is perfect because configurational entropy does not play a role. Recall from Section 3.4.1 that in this region, where void formation energies are well described by the (111) surface energy, the EDIP predictions are in excellent agreement with experimental measurements. Finally, in the top right corner of Figure 3.9(B), corresponding to larger clusters at high temperature, a moderate but persistent error between the ground state and ISL analysis exists. This size–temperature regime corresponds to surface melting of voids, in which the overall void shape may be compact, but configurational entropy is sufficient to lead to a melted surface layer at the void–matrix interface. In fact, this feature is directly connected to the phenomenon of defect-induced melting, which is beyond the scope of the present discussion.

In summary, the ISL picture for vacancy cluster thermodynamics shows surprisingly rich behavior. At high temperatures, configurational entropy drives small clusters into mobile, loosely bound structures that have lower free energy than corresponding compact configurations. These effects play major roles in the nucleation and early growth phase of small vacancy clusters. As the crystal cools, the impact of configurational entropy gradually decays, leading to the well-known compact octahedral structures observed experimentally. The action of configurational entropy is therefore somewhat insidious—although it materially impacts the final distribution of voids in silicon crystals, it is hard (or even impossible) to observe it directly in action. However, there is some indirect evidence for the validity of this overall picture. In ref. [57], the ISL-based free energy function obtained in Figure 3.9 was employed in a continuum model for void nucleation and growth during CZ growth of silicon crystals [1,58,105,106]. While these models have been quite successful at predicting void size distributions as a function of crystal growth conditions, they have relied on some degree of empirical parameter fitting in order to attain quantitative agreement with experimental

measurements of void size, void density, and the nucleation temperature at which voids suddenly nucleate during crystal cooling. A long-standing issue in particular was the fact that the (111) surface model for octahedral voids, while intuitively appealing, gave free energies that were too high and resulted in insufficient void formation. The ISL free energy picture resolves this issue with no empirical parameter fitting. At high temperatures where nucleation and growth take place, the free energy is much lower than the octahedral model would predict, and at low temperatures it naturally becomes consistent with it.

In concluding this section, it should be noted once again that the above results are strictly only valid for the EDIP potential. As mentioned earlier, there is some evidence that PEL roughness is somewhat overpredicted by empirical potentials as compared to DFT calculations [99]. If this is indeed the case, the results presented here are likely to overestimate the impact of configurational entropy. On the other hand, similar calculations with multiple empirical potentials (data not shown) seem to show that the presence of a noisy PEL with many closely spaced, mechanically stable inherent structures is fairly ubiquitous, with only quantitative deviations across the various potential models. Clearly, further work aimed at analyzing the nature of empirical potential and DFT PELs is required before more formal conclusions regarding this issue can be made.

3.6.3 Self-Interstitial Clusters

The ISL-based analysis of self-interstitial clusters is essentially identical to that for vacancy clusters. Given the higher degree of morphological complexity associated with self-interstitial clusters as mentioned earlier in Sections 3.4.2 and 3.4.3, only a few of the major features are described here and the reader is referred to refs. [92,107,108] for more details. Shown in Figure 3.10 are PDFs for several interstitial cluster sizes computed

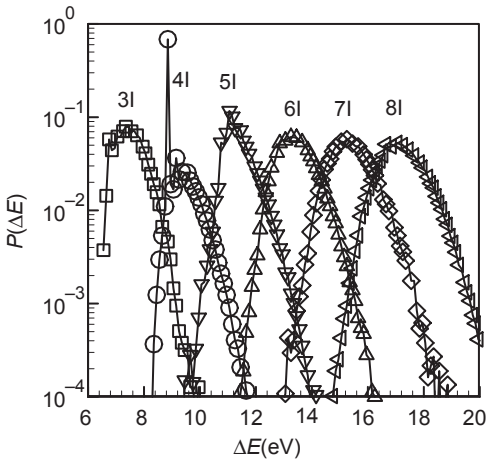


FIGURE 3.10 Probability distribution functions for EDIP self-interstitial clusters at 1100 K. *Adapted from Ref. [108].*

using the EDIP potential at 1100 K. Most of the distributions are qualitatively similar to the vacancy cluster PDFs, except for a striking feature in the 4-intersitial case. Here, a sharp peak is observed in the probability distribution at 8.75 eV formation energy that corresponds to the Humble/Arai configuration discussed in Section 3.4.2. Note that the peak does not correspond to the lowest energy configuration, i.e., the EDIP potential predicts slightly lower formation energy configurations, although these are observed quite rarely. Thus, while the Humble/Arai configuration, which is the known ground state configuration for the 4-interstitial cluster, is not predicted to be the absolute EDIP ground state, it is by far the most favored configuration. Other cluster sizes do not appear to exhibit similar special configurations.

The special favorability of the Humble/Arai configuration arises from a single unusual feature—it possesses very high vibrational entropy relative to energetically neighboring configurations [108]. Shown in Figure 3.11 are vibrational entropies computed using the QHA for large numbers of inherent structures for two vacancy clusters (left) and two interstitial clusters (right). Qualitatively, all clusters exhibit the same behavior in which the vibrational entropy increases roughly linearly with increasing inherent structure energy, although the dispersion is larger for the interstitial clusters. The vibrational entropy of formation corresponding to the Humble/Arai configuration is denoted by the large shaded sphere in Figure 3.11(C) and is notably higher (by about 5–6 k_B) than the surrounding values. This anomalous vibrational entropy readily accounts for the approximately O(10^2) increase in probability for the Humble/Arai configuration. Thus, although some configurations are predicted by the EDIP potential to be slightly more

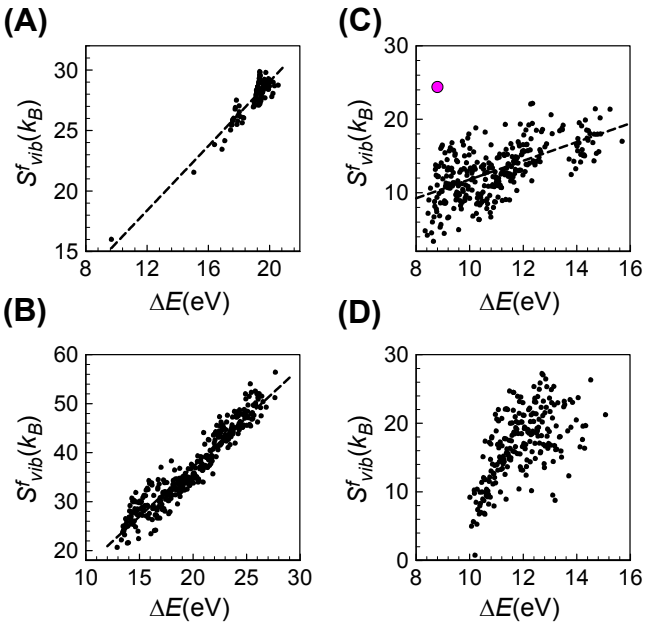


FIGURE 3.11 Vibrational entropy of formation for (A) 6-vacancy, (B) 10-vacancy, (C) 4-interstitial, and (D) 5-interstitial clusters as a function of formation energy. Each symbol represents a QHA calculation for a single inherent structure of a given cluster species. The large circle in (C) represents the Humble/Arai configuration. Dashed lines are guides and do not represent linear fits. *Taken from Ref. [108].*

energetically favorable than the Humble/Arai structure, it is entropically stabilized at finite temperature. In the language of potential energy landscapes, the basin that corresponds to the Humble/Arai configuration is particularly "wide" rather than being exceptionally "deep".

The fact that the source of the stability is entropic leads to fundamentally different temperature dependence. Recall that for the vacancy cluster case, the ground state and other low energy/low entropy configurations become progressively less important as the system temperature rises. This is because as the temperature increases, the higher configurational entropy associated with higher energy configurations begins to dominate the free energy shifting the system toward sampling higher energy configurations. Here, however, due to the fact that the Humble/Arai configuration is vibrationally stabilized, it remains important even at elevated temperature.

3.6.3.1 Landscape Roughness and the Effect of Pressure

The prior discussion of small interstitial clusters in Section 3.4.2 suggests that similar entropic effects should be observed for the 8-interstitial cluster whose ground state is essentially two interacting Humble/Arai clusters in close proximity. However, no anomalous peak is obviously apparent in Figure 3.10 in the PDF for 8-interstitial case; this was also found to be the case at all temperatures at which ISL sampling was performed [108]. While temperature appears to have a simple and predictable influence on the PDFs shown in Figure 3.10—they shift to the right and become wider with increasing temperature—the impact of hydrostatic pressure was found to be much more subtle and provides insight into the complex interplay between energy and configurational and vibrational entropies in these systems.

PDFs at 1100 K are shown in Figure 3.12 for the 4- and 8-interstitial clusters at different applied hydrostatic pressures. The base-case zero pressure PDFs are denoted

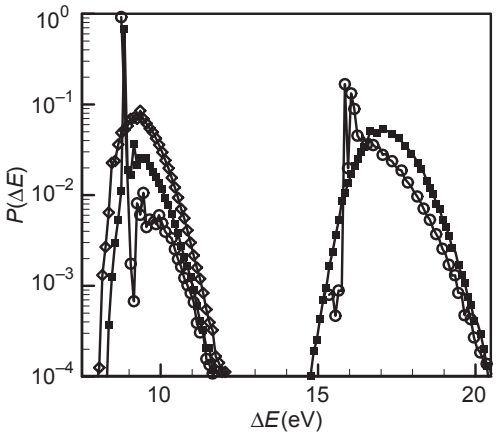


FIGURE 3.12 PDFs for the 4 (left) and 8 (right) interstitial clusters at 1100 K. Filled squares—zero pressure; open circles—3 GPa tension; open diamonds (4-interstitial only)—3 GPa compression.

by the filled square symbols and correspond to the PDFs shown previously in Figure 3.10. The open circles show the PDFs computed under 3 GPa of applied tension, which corresponds to about 1% tensile strain. The 4-interstitial PDF is modified in two important ways—first the dominance of the Humble/Arai peak increases, and second, the inherent structures previously found to the left of the peak are no longer mechanically stable and disappear so that the Humble/Arai configuration becomes the lowest energy configuration. The change in 8-interstitial PDF is even more striking. Whereas the zero pressure distribution was the usual smoothly rounded distribution, the PDF under tension now exhibits sharp peaks near the minimum energy, much like the 4-interstitial case. In fact, these spikes correspond to the well-known "compact" and "elongated" configurations identified in refs. [46,48]. Once again, their probabilities are enhanced by their high vibrational entropies relative to neighboring configurations. Finally, the 4-interstitial PDF obtained under compression is shown by the open diamond symbols. Here, the peak corresponding to the Humble/Arai configuration is no longer visible.

It is tempting to interpret the trends in Figure 3.12 by considering the effect of hydrostatic pressure on the Humble/Arai configuration. However, as shown in ref. [108] applied pressure (compressive or tensile) does not alter the formation energy or vibrational entropy of the Humble/Arai configuration. Instead, the behavior in Figure 3.12 arises from a collective phenomenon across the entire PEL of the system. In Figure 3.13, the DOS distributions for the 4-interstitial cluster at each of the three applied pressures are plotted so that they are equal at the energy bin that contains the Humble/Arai configuration. The effect of hydrostatic pressure is immediately evident. Simply put, compression tends to increase the density of states while tension reduces it. Given that

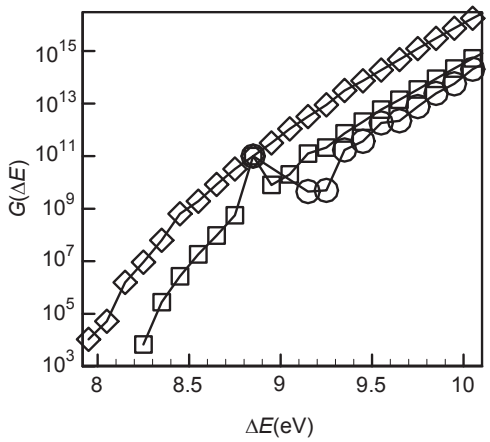


FIGURE 3.13 4-interstitial DOS as a function of pressure. Squares—zero pressure; circles—3 GPa tension; diamonds—3 GPa compression. *Taken from Ref. [108].*

the vibrational entropy of a given configuration is unaltered by hydrostatic pressure (at least at the levels applied here), the effect on the DOS is largely configurational. Thus, when compression is applied, the Humble/Arai configuration does not become altered; it simply loses its dominance over other configurations because the total number of alternative configurations is now much greater. The opposite effect is present under tension—as inherent structures are lost, the dominance of the Humble/Arai configuration becomes even stronger. Thus, whether the Humble/Arai configuration is dominant or not, depends on a competition between its high vibrational entropy and the collective configurational entropy of all other configurations.

A very simple, yet conceptually appealing, interpretation of the effect of hydrostatic pressure is that compression acts to roughen the potential energy landscape, while tension acts to smooth it, much like pushing and pulling on a sheet of paper. Subsequent analysis with other potentials such as Tersoff has shown that this is not unique to EDIP, and appears to be a common feature of empirical potentials (or at least bond-order type potentials). One possible explanation for this behavior is that as the system is compressed, the number of neighbors within interaction range of a given atom increases, leading to increased landscape roughness.

Finally, it is interesting to note that while the EDIP potential predicted the existence of apparently spurious stable configurations with formation energies below that of the Humble/Arai configuration, these minima in the PEL quickly disappear under the application of tension. Consider a situation in which the EDIP potential is being tested using the 4-interstitial cluster. Preforming the defect in the Humble/Arai configuration would show that, indeed, this is a stable structure with a predicted formation energy that is in good agreement with DFT values. On the other hand, a global minimization over the PEL, with no initial knowledge of the Humble/Arai configuration, would identify a ground state that looks nothing like the Humble/Arai configuration with slightly lower formation energy. In this case, the potential would be considered as invalid when compared to DFT. The present analysis shows that the potential is largely in excellent agreement with DFT results but appears to predict some spurious local minima some of which lie at lower energy than the Humble/Arai configuration. The collective impact of these spurious minima on the overall configurational entropy predicted by EDIP is an open question. Answering it may help identify new ways to think about empirical potential design and validation.

3.7 Conclusions

The aim of this chapter was to provide evidence, through several vignettes, for the utility of atomistic simulations of intrinsic defect physics in silicon. The identification of the energetic ground state and the subsequent calculation of its structure and formation thermodynamics has been the subject of many studies, yet the ability to do this with complete quantitative confidence remains elusive. In the second portion of the chapter,

an example application of atomistic simulation to the study of entropic effects at elevated temperature was used to demonstrate the potential for such simulations to provide mechanistic information that is generally very difficult to come by with experimental measurements.

Overall, it is safe to say that atomistic calculations are, as a collective tool kit, extremely powerful and versatile tools for studying the thermodynamics of point defect and defect clusters in crystalline silicon. Indeed, there are now countless examples of successful studies in the literature and there is little doubt that they have greatly increased our understanding of defect physics, thermodynamics and transport in silicon. However, numerous challenges remain and these must be clearly identified when deciding on how to apply this tool kit to a particular situation. First and foremost, it is important to state at the outset that the notion of employing atomistic calculations in a "plug-and-play" fashion to compute defect properties on demand has not yet been realized. Ironically, the temptation to treat simulations in this manner has been exacerbated by the increased availability of sophisticated, user-friendly and very powerful open-source codes. However, in both empirical potential and electronic structure calculations, there are many parameters that need to be carefully specified (and are problem specific) and many validation steps that need to be undertaken each time. This is especially true when the properties being computed are not well studied. In the case of empirical potential calculations, it is clear that the choice of potential is crucial—while there is no one single potential that works best for everything, some are clearly better for particular aspects. Even for so-called ab initio calculations, the proper choice of model parameters such as sampling points for integration, system size, and exchange-correlation functional can make surprisingly large impacts on the computed property. This can lead to an overestimate of the accuracy of the results.

On the whole, and particularly for silicon, it can therefore be argued that much of the success achieved with atomistic simulation of defect physics in silicon has not been related to the calculation of quantitative estimates for specific thermophysical properties. Instead the success has stemmed in large part from an improved mechanistic understanding of processes that are otherwise extremely difficult to probe experimentally. As mentioned above, part of the reason for this is the large scatter in the predicted properties with respect to a number of model parameters and choices. Moreover, coupled with this is the fact that many observable defect-related phenomena in silicon crystal growth and wafer annealing arise from a highly sensitive balance between self-interstitials and vacancies requiring tremendous (and presently unattainable) accuracy to predict without some amount of parametric regression. As a result, the most successful applications of this tool kit have been to establish a semi-quantitative picture, which is then further refined by detailed comparisons to experimental measurements. Given the tremendous progress in the quality and scope of atomistic simulations over the last few decades, there is no doubt that the role of such calculations will only grow.

References

[1] Sinno T, Dornberger E, von Ammon W, Brown RA, Dupret F. Defect engineering of Czochralski single-crystal silicon. Mater Sci Eng R 2000;28:149–98.

[2] Katz JL, Wiedersi H. Nucleation of voids in materials supersaturated with vacancies and interstitials. J Chem Phys 1971;55:1414.

[3] Prasad M, Sinno T. Internally consistent approach for modeling solid-state aggregation. I. Atomistic calculations of vacancy clustering in silicon. Phys Rev B 2003;68:045206.

[4] Lifshitz IM, Slyozov VV. The kinetics of precipitation from supersaturated solid solutions. J Phys Chem Solids 1961;19:35–50.

[5] Balamane H, Halicioglu T, Tiller WA. Comparative-study of silicon empirical interatomic potentials. Phys Rev B 1992;46:2250–79.

[6] Stillinger FH, Weber TA. Computer-simulation of local order in condensed phases of silicon. Phys Rev B 1985;31:5262–71.

[7] Pearson E, Takai T, Halicioglu T, Tiller WA. Computer modeling of Si and Sic surfaces and surface processes relevant to crystal-growth from the vapor. J Cryst Growth 1984;70:33–40.

[8] Biswas R, Hamann DR. Interatomic potentials for silicon structural energies. Phys Rev Lett 1985;55:2001–4.

[9] Tersoff J. Empirical interatomic potential for silicon with improved elastic properties. Phys Rev B 1988;38:9902–5.

[10] Tersoff J. Modeling solid-state chemistry – interatomic potentials for multicomponent systems. Phys Rev B 1989;39:5566–8.

[11] Agrawal PM, Raff LM, Komanduri R. Monte Carlo simulations of void-nucleated melting of silicon via modification in the tersoff potential parameters. Phys Rev B 2005;72.

[12] Erhart P, Albe K. Analytical potential for atomistic simulations of silicon, carbon, and silicon carbide. Phys Rev B 2005;71.

[13] Pizzagalli L, Godet J, Guénolé J, Brochard S, Holmstrom E, Nordlund K, et al. A new parametrization of the Stillinger–Weber potential for an improved description of defects and plasticity of silicon. J Phys Condens Matter 2013;25:055801.

[14] Vink RLC, Barkema GT, van der Weg WF, Mousseau N. Fitting the Stillinger–Weber potential to amorphous silicon. J Non-Cryst Solids 2001;282:248–55.

[15] Kelires PC, Tersoff J. Equilibrium alloy properties by direct simulation – oscillatory segregation at the Si–Ge(100) 2x1 surface. Phys Rev Lett 1989;63:1164–7.

[16] Cook SJ, Clancy P. Comparison of semiempirical potential functions for silicon and germanium. Phys Rev B 1993;47:7686–99.

[17] Baskes MI. Atomistic potentials for the molybdenum-silicon system. Mat Sci Eng A 1999;261:165–8.

[18] Thijsse BJ. Silicon potentials under (ion) attack: towards a new MEAM model. Nucl Instrum Meth B 2005;228:198–211.

[19] Lee BJ. A modified embedded atom method interatomic potential for silicon. Calphad-Computer Coupling Phase Diagrams Thermochem 2007;31:95–104.

[20] Timonova M, Thijsse BJ. Optimizing the MEAM potential for silicon. Model Simul Mater Sci Eng 2011;19.

[21] Lenosky TJ, Sadigh B, Alonso E, Bulatov VV, de la Rubia TD, Kim J, et al. Highly optimized empirical potential model of silicon. Model Simul Mater Sci Eng 2000;8:825.

[22] Bazant MZ, Kaxiras E, Justo JF. Environment-dependent interatomic potential for bulk silicon. Phys Rev B 1997;56:8542–52.

[23] Justo JF, Bazant MZ, Kaxiras E, Bulatov VV, Yip S. Interatomic potential for silicon defects and disordered phases. Phys Rev B 1998;58:2539–50.

[24] Sinno T, Jiang ZK, Brown RA. Atomistic simulation of point defects in silicon at high temperature. Appl Phys Lett 1996;68:3028–30.

[25] Baskes MI, Nelson JS, Wright AF. Semiempirical modified embedded-atom potentials for silicon and germanium. Phys Rev B 1989;40:6085–100.

[26] Gillespie B, Zhou X, Murdick D, Wadley H, Drautz R, Pettifor D. Bond-order potential for silicon. Phys Rev B 2007;75:155207.

[27] Marqués LA, Pelaz L, Castrillo P, Barbolla J. Molecular dynamics study of the configurational and energetic properties of the silicon self-interstitial. Phys Rev B 2005;71:085204.

[28] Argaman N, Makov G. Density functional theory: an introduction. Am J Phys 2000;68:69–79.

[29] Estreicher SK. Structure and dynamics of point defects in crystalline silicon. Phys Status Solidi B 2000;217:513–32.

[30] Colombo L. Tight-binding molecular dynamics: a primer. Riv Nuovo Cimento 2005;28:1–59.

[31] Lenosky TJ, Kress JD, Kwon I, Voter AF, Edwards B, Richards DF, et al. Highly optimized tight-binding model of silicon. Phys Rev B 1997;55:1528.

[32] Antonelli A, Bernholc J. Pressure effects on self-diffusion in silicon. Phys Rev B 1989;40:10643–6.

[33] Scheffler M, Vigneron JP, Bachelet GB. Total-energy gradients and lattice-distortions at point-defects in semiconductors. Phys Rev B 1985;31:6541–51.

[34] Car R, Kelly PJ, Oshiyama A, Pantelides ST. Microscopic theory of impurity-defect reactions and impurity diffusion in silicon. Phys Rev Lett 1985;54:360–3.

[35] Nichols CS, Vandewalle CG, Pantelides ST. Mechanisms of dopant impurity diffusion in silicon. Phys Rev B 1989;40:5484–96.

[36] Blochl PE, Smargiassi E, Car R, Laks DB, Andreoni W, Pantelides ST. 1st-principles calculations of self-diffusion constants in silicon. Phys Rev Lett 1993;70:2435–8.

[37] Pino AD, Rappe AM, Joannopoulos JD. Abinitio investigation of carbon-related defects in silicon. Phys Rev B 1993;47:12554–7.

[38] Zhu J, delaRubia TD, Yang LH, Mailhiot C, Gilmer GH. Ab initio pseudopotential calculations of B diffusion and pairing in Si. Phys Rev B 1996;54:4741–7.

[39] Mercer JL, Nelson JS, Wright AF, Stechel EB. Ab initio calculations of the energetics of the neutral Si vacancy defect. Model Simul Mater Sci Eng 1998;6:1–8.

[40] Puska MJ, Poykko S, Pesola M, Nieminen RM. Convergence of supercell calculations for point defects in semiconductors: vacancy in silicon. Phys Rev B 1998;58:1318–25.

[41] Probert MIJ, Payne MC. Improving the convergence of defect calculations in supercells: an ab initio study of the neutral silicon vacancy. Phys Rev B 2003;67.

[42] Wright A. Density-functional-theory calculations for the silicon vacancy. Phys Rev B 2006;74:165116.

[43] Corsetti F, Mostofi AA. System-size convergence of point defect properties: the case of the silicon vacancy,". Phys Rev B 2011;84.

[44] Dai JG, Kanter JM, Kapur SS, Seider WD, Sinno T. On-lattice kinetic Monte Carlo simulations of point defect aggregation in entropically influenced crystalline systems. Phys Rev B 2005;72.

[45] Prasad M, Sinno T. Internally consistent approach for modeling solid-state aggregation. I. Atomistic calculations of vacancy clustering in silicon. Phys Rev B 2003;68.

[46] Lee S, Hwang GS. Structure and stability of small compact self-interstitial clusters in crystalline silicon. Phys Rev B 2008;77:085210.

[47] Kim J, Kirchhoff F, Aulbur WG, Wilkins JW, Khan FS, Kresse C. Thermally activated reorientation of di-interstitial defects in silicon. Phys Rev Lett 1999;83:1990–3.

[48] Kim J, Kirchhoff F, Wilkins JW, Khan FS. Stability of Si-interstitial defects: from point to extended defects. Phys Rev Lett 2000;84:503–6.

[49] Wang C, Wang Z, Meng Q. Comparative study of the empirical interatomic potentials and density-functional simulations of divacancy and hexavacancy in silicon. Phys B: Condens Matter 2011;406:467–70.

[50] Bongiorno A, Colombo L, Cargnoni F, Gatti C, Rosati M. Evolution of energetics and bonding of compact self-interstitial clusters in Si. Europhys Lett 2000;50:608–14.

[51] Prasad M, Sinno T. Atomistic-to-continuum description of vacancy cluster properties in crystalline silicon. Appl Phys Lett 2002;80:1951–3.

[52] Chadi DJ, Chang KJ. Magic numbers for vacancy aggregation in crystalline Si. Phys Rev B 1988;38:1523–5.

[53] Estreicher SK, Hastings JL, Fedders PA. The ring-hexavacany in silicon: a stable and inactive defect. Appl Phys Lett 1997;70:432–4.

[54] Prasad M, Sinno T. Internally consistent approach for modeling solid-state aggregation. II. Mean-field representation of atomistic processes. Phys Rev B 2003;68.

[55] Jaccodine RJ. Surface energy of germanium and silicon. J Electrochem Soc 1963;110:524–7.

[56] Gilman JJ. Direct measurements of the surface energies of crystals. J Appl Phys 1960;31:2208–18.

[57] Frewen TA, Kapur SS, Haeckl W, von Ammon W, Sinno T. A microscopically accurate continuum model for void formation during semiconductor silicon processing. J Cryst Growth 2005;279:258–71.

[58] Voronkov VV, Falster R. Vacancy-type microdefect formation in Czochralski silicon. J Cryst Growth 1998;194:76–88.

[59] Kapur SS, Sinno T. Entropic origins of stability in silicon interstitial clusters. Appl Phys Lett 2008;93:221911.

[60] Cowern NEB, Mannino G, Stolk PA, Roozeboom F, Huizing HGA, van Berkum JGM, et al. Energetics of self-interstitial clusters in Si. Phys Rev Lett 1999;82:4460–3.

[61] Madden PK, Davidson SM. Nature of rod-like defects observed in boron-irradiated silicon. Radiat Eff 1972;14:271–3.

[62] Kohyama M, Takeda S. Atomic-structure and energy of the (113) planar interstitial defects in Si. Phys Rev B 1992;46:12305–15.

[63] Jones KS, Liu J, Zhang L, Krishnamoorthy V, DeHoff RT. Studies of the interactions between (311) defects and type I and II dislocation loops in Si+ implanted silicon. Nucl Instrum Methods Phys Res Sect B 1995;106:227–32.

[64] Pan GZ, Tu KN. Transmission electron microscopy on {113} rodlike defects and {111} dislocation loops in silicon-implanted silicon. J Appl Phys 1997;82:601–8.

[65] Muto S, Takeda S. New stable defect structure on the (001) plane in germanium formed by deuteron irradiation. Philos Mag Lett 1995;72:99–104.

[66] Li JH, Jones KS. {311} defects in silicon: the source of the loops. Appl Phys Lett 1998;73:3748–50.

[67] Claverie A, Colombeau B, De Mauduit B, Bonafos C, Hebras X, Ben Assayag G, et al. Extended defects in shallow implants. Appl Phys A 2003;76:1025–33.

[68] Goss JP, Eberlein TAG, Jones R, Pinho N, Blumenau AT, Frauenheim T, et al. Planar interstitial aggregates in Si. J Phys: Condens Matter 2002;14:12843–53.

[69] Fedina L, Gutakovskii A, Aseev A, Van Landuyt J, Vanhellemont J. On the mechanism of {111}-defect formation in silicon studied by in situ electron irradiation in a high resolution electron microscope. Philos Mag A 1998;77:423–35.

[70] Cristiano F, Colombeau B, Bonafos C, Altibelli A, Ben Assayag G, Claverie A. Thermal evolution of extrinsic defects in ion implanted silicon: current understanding and modelling. Gettering and defect engineering semiconductor technology, vol. 82–84; 2002. 201–206.

[71] Chaudhry S, Liu J, Jones KS, Law ME. Evolution of dislocation loops in silicon in an inert ambient .2. Solid-State Electron 1995;38:1313–9.

[72] Claverie A, Colombeau B, Cristiano F, Altibelli A, Bonafos C. Modeling of the Ostwald ripening of extrinsic defects and transient enhanced diffusion in silicon. Nucl Instrum Methods Phys Res Sect B 2002;186:281–6.

[73] Coomer BJ, Goss JP, Jones R, Oberg S, Briddon PR. Identification of the tetra-interstitial in silicon. J Phys-Condens Matter 2001;13:L1–7.

[74] Humble P. The structure and mechanism of formation of platelets in natural type Ia Diamond. Proc R Soc London Ser A 1982;381:65.

[75] Arai N, Takeda S, Kohyama M. Self-interstitial clustering in crystalline silicon. Phys Rev Lett 1997; 78:4265–8.

[76] Kohyama M, Takeda S. First-principles calculations of the self-interstitial cluster I-4 in Si. Phys Rev B 1999;60:8075–80.

[77] Kapur SS, Sinno T. Entropic origins of stability in silicon interstitial clusters. Appl Phys Lett 2008;93.

[78] Alippi P, Colombo L. Lattice-strain field induced by {311} self-interstitial defects in silicon. Phys Rev B 2000;62:1815–20.

[79] Cristiano F, Grisolia J, Colombeau B, Omri M, de Mauduit B, Claverie A, et al. Formation energies and relative stability of perfect and faulted dislocation loops in silicon. J Appl Phys 2000;87:8420–8.

[80] Claverie A, Giles LF, Omri M, de Mauduit B, Ben Assayag G, Mathiot D. Nucleation, growth and dissolution of extended defects in implanted Si: impact on dopant diffusion. Nucl Instrum Methods Phys Res Sect B 1999;147:1–12.

[81] Cristiano F, Hebras X, Cherkashin N, Claverie A, Lerch W, Paul S. Clusters formation in ultralow-energy high-dose boron-implanted silicon. Appl Phys Lett 2003;83:5407–9.

[82] Bogdan TV, Wales DJ, Calvo F. Equilibrium thermodynamics from basin-sampling. J Chem Phys 2006;124.

[83] Calvo F, Doye JPK, Wales DJ. Collapse of Lennard-Jones homopolymers: size effects and energy landscapes. J Chem Phys 2002;116:2642–9.

[84] Farrell JD, Lines C, Shepherd JJ, Chakrabarti D, Miller MA, Wales DJ. Energy landscapes, structural topologies and rearrangement mechanisms in clusters of dipolar particles. Soft Matter 2013;9: 5407–16.

[85] Sciortino F, Kob W, Tartaglia P. Inherent structure entropy of supercooled liquids. Phys Rev Lett 1999;83:3214–7.

[86] Debenedetti PG, Stillinger FH. Supercooled liquids and the glass transition. Nature 2001;410: 259–67.

[87] Buchner S, Heuer A. Potential energy landscape of a model glass former: thermodynamics, anharmonicities, and finite size effects. Phys Rev E 1999;60:6507–18.

[88] Nakagawa N, Peyrard M. The inherent structure landscape of a protein. Proc Natl Acad Sci USA 2006;103:5279–84.

[89] Stillinger FH, Weber TA. Point-defects in Bcc crystals – structures, transition kinetics, and melting implications. J Chem Phys 1984;81:5095–103.

[90] Stillinger FH, Weber TA. Hidden structure in liquids. Phys Rev A 1982;25:978–89.

[91] Wales DJ. The energy landscape as a unifying theme in molecular science. Philos Trans R Soc A 2005;363:357–75.

[92] Nieves AM, Chuang CY, Sinno T. Inherent structure analysis of defect thermodynamics and melting in silicon. Mol Simul 2012;38:659–70.

[93] Mishin Y, Sorensen MR, Voter AF. Calculation of point-defect entropy in metals. Philos Mag A 2001;81:2591–612.

[94] Nieves AM, Sinno T. An enthalpy landscape view of homogeneous melting in crystals. J Chem Phys 2011;135.

[95] Mauro JC, Loucks RJ, Balakrishnan J, Raghavan S. Monte Carlo method for computing density of states and quench probability of potential energy and enthalpy landscapes. J Chem Phys 2007;126.

[96] Potuzak M, Mauro JC, Kiczenski TJ, Ellison AJ, Allan DC. Communication: resolving the vibrational and configurational contributions to thermal expansion in isobaric glass-forming systems. J Chem Phys 2010;133.

[97] Goedecker S. Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. J Chem Phys 2004;120:9911–7.

[98] Kapur SS, Prasad M, Crocker JC, Sinno T. Role of configurational entropy in the thermodynamics of clusters of point defects in crystalline solids. Phys Rev B 2005;72.

[99] Ghasemi SA, Amsler M, Hennig RG, Roy S, Goedecker S, Lenosky TJ, et al. Energy landscape of silicon systems and its description by force fields, tight binding schemes, density functional methods, and quantum Monte Carlo methods. Phys Rev B 2010;81:214107.

[100] Plimpton SJ. Fast parallel algorithms for short-range molecular dynamics. J Comp Phys 1995;117:1.

[101] Wales DJ, Berry RS. Melting and freezing of small argon clusters. J Chem Phys 1990;92:4283–95.

[102] Stillinger FH. Rigorous basis of Frenkel-Band theory of association equilibrium. J Chem Phys 1963; 38:1486.

[103] Bian Q, Bose SK, Shukla RC. Vibrational and thermodynamic properties of metals from a model embedded-atom potential. J Phys Chem Solids 2008;69:168–81.

[104] Della Valle RG, Venuti E. Quasiharmonic lattice-dynamics and molecular-dynamics calculations for the Lennard-Jones solids. Phys Rev B 1998;58:206–12.

[105] Sinno T, Brown RA. Modeling microdefect formation in Czochralski silicon. J Electrochem Soc 1999;146:2300–12.

[106] Sinno T. A bottom-up multiscale view of point-defect aggregation in silicon. J Cryst Growth 2007; 303:5–11.

[107] Kapur SS, Sinno T. Detailed microscopic analysis of self-interstitial aggregation in silicon. I. Direct molecular dynamics simulations of aggregation. Phys Rev B 2010;82.

[108] Kapur SS, Nieves AM, Sinno T. Detailed microscopic analysis of self-interstitial aggregation in silicon. II. Thermodynamic analysis of single clusters. Phys Rev B 2010;82.

4 Stoichiometry of Oxide Crystals

Satoshi Uda

INSTITUTE FOR MATERIALS RESEARCH, TOHOKU UNIVERSITY, KATAHIRA, AOBA-KU, SENDAI, MIYAGI, JAPAN

CHAPTER OUTLINE

4.1 General Overview of Conventional Stoichiometry and Related Point Defects 176
4.1.1 Brief Overview of Stoichiometry and Point Defects in Oxide Compounds 176
4.1.2 Crystal Sites in Oxide Crystals 179
4.1.3 Partitioning Behavior of Ionic Solutes in an Oxide Melt 180
4.2 Extended Concept of Stoichiometry in Oxide Crystals 182
4.2.1 The Thermodynamic Definition of Stoichiometry Associated with the Activity of Constituent Elements 183
4.2.2 Conventional Stoichiometric Compositions and the Law of Multiple Proportions 184
4.2.2.1 Activities of the Constituent Elements in a Conventional Stoichiometric Crystal 185
4.2.3 Extended Stoichiometric Compositions Including Impurities and Vacancies 188
4.2.3.1 Degrees of Freedom in a Crystal Site 188
4.2.3.2 Activity of the Constituent Elements in an Extended Stoichiometric Crystal 189
4.3 Growth Characteristics of Stoichiometric Crystals 192
4.3.1 Crystallization Electromotive Force 193
4.3.2 Activities of Constituent Elements in the Melt and Solid States for an Oxide Crystal that is Both Stoichiometric and Congruent 196
4.4 Oxide Crystals Having a Stoichiometric Composition Coincident with the Congruent Point 200
4.4.1 MgO-Doped $LiNbO_3$ 201
4.4.1.1 Distribution of the Melting Temperatures of MgO-Doped $LiNbO_3$ 201
4.4.1.2 Crystallization Electromotive Force 202
4.4.1.3 Bulk Crystal Growth of cs-MgO:$LiNbO_3$ and Its Nonlinear Optical Characterization 203
4.4.2 MgO-Doped $LiTaO_3$ 207
4.4.2.1 Distribution of Melting Temperatures of MgO-Doped $LiTaO_3$ 207
4.4.2.2 Crystallization Electromotive Force 208
4.4.2.3 Bulk Crystal Growth and Curie Temperature of cs-MgO:$LiTaO_3$ 208
4.4.3 Thermodynamic Requirements for Impurity Doping 210
4.5 Summary 212
References 212

Handbook of Crystal Growth. http://dx.doi.org/10.1016/B978-0-444-56369-9.00004-6

4.1 General Overview of Conventional Stoichiometry and Related Point Defects

4.1.1 Brief Overview of Stoichiometry and Point Defects in Oxide Compounds

A conventional stoichiometric compound is characterized as one that has an exact and fixed composition with a small integer ratio among its atomic components. This concept is related to Dalton's so-called law of multiple proportions [1]. The atoms in such a compound are located precisely at the appropriate crystal lattice points, and neither impurities nor defects are allowed to exist. Nonstoichiometry represents a deviation from the well-defined stoichiometry, by the introduction of point defects such as impurities and vacancies, when forming a solid solution. As such, the definition of a nonstoichiometric oxide material often refers to a structure that may be formed within a certain range of anion-to-cation ratios. Stoichiometric materials are used in electronic, optic, and optoelectronic devices [2,3] because of their excellent uniformity and high efficiency, which are attributed to the perfect ordering of the atoms in the crystal lattice. For an ideal stoichiometric structure, defects should be avoided. On the other hand, because their compositional variation enables the optimization of piezoelectric, ferroelectric, ferromagnetic, and other physical properties, nonstoichiometric materials with certain ranges of solid solutions often attract more attention than stoichiometric materials from the viewpoint of potential applications, given their greater ability to meet the requirements for different devices. Therefore, point defects, such as impurities or vacancies, are often intentionally added to a host material for such a purpose. Semiconducting materials, in which only a small amount of dopant in the range of 10^{14}–10^{18} /cm^3 is sufficient to cause drastic changes in the electronic properties, are typical examples [4–7]. Color change in the oxides [8] is another. However, it is difficult to experimentally show the exactness of a stoichiometric composition; it is also challenging to precisely determine the solid solution range in nonstoichiometric materials.

When discussing the characteristics of point defects, it is vital to have an understanding of the structure of the host material. Conventionally, the bonding state between the constituent elements, that is, the major components (the cations and anions) and any impurities and vacancies, forms the basis for discussing the generation, morphology, and mode of distribution of defects. Pauling [9] introduced the theoretical concepts of bond ionicity and bond strength parameters for a stable coordination polyhedron based on his second principle, the electrostatic valence rule. Since then, many modifications to the second principle have been suggested, including those of Brown and Shannon [10] and Brown [11,12]. The result has been the development of a concept proposing a bond-valence model in which static bond energy is a function of bond length, i.e., the distance between an oxygen ion and an associated cation. However, this concept is largely restricted to ionic compounds, and the parameters evaluated to date are not necessarily effective when applied to oxides, since the bonding state of many oxides represents a

mixture of ionic and covalent characteristics. As an example, MgO represents a typical ionic compound while SiO_2 is covalent. The degree of ionicity of the oxide is important and dominates its ionization behavior in the melt, as will be discussed in Section 4.4.3.

The defect structure is uncertain for many oxide compounds, due to an inadequate understanding of issues related to the crystal structure, including the coordination sites for metal cations, ionic-covalent bonding states, oxygen-to-oxygen interactions, degree of oxidation, effect of the size ratio of metal to oxygen on crystal deformation, defect interactions, population of point defects and their configurations associated with ordering and disordering, solid solution ranges, and so on. For instance, the balance between ionic and covalent bonding depends on the individual oxide, which leads to variable energy states associated with defect generation and makes it difficult to predict bond length, lattice size, and the mode of deformation associated with defects and defect interactions. The oxide object of a defect study can vary from simple cubic to more complex perovskite and other crystal types. Many oxides show significant deviations from stoichiometry; depending on the oxygen pressure, the transition metals form numerous nonstoichiometric oxide phases with ions of differing valences, resulting in varying metal ion-to-oxygen ratios. Among these are the Magnéli phases, such as V_nO_{2n-1} and Ti_nO_{2n-1} [13,14]. The general form of two-element cubic compounds can be addressed as $M_{a/b}O$ crystals, where a/b is the ratio of metal (M) to oxygen (O) per mole of O atoms. The transition metal monoxides, such as NiO and CoO, are representative examples of $M_{a/b}O$. For these materials, the ratio of metal to oxygen varies; NbO and VO have ratios both greater and less than one ($NbO_{1.9975}$ to $NbO_{2.003}$ and $VO_{0.8}$ to $VO_{1.27}$), while CoO and CdO have ratios less than or equal to one ($Co_{0.99}O$ to $Co_{1.00}O$) and equal to or greater than one ($Cd_{1.00}O$ to $Cd_{1.0005}O$) [15]. The a/b ratio affects the physical properties of oxides, which are dominated by conventional factors, such as the ionic size, electronegativity, ionic-covalent character, attractive and repulsive interactions, and various oxidation states of the metal. The molar volume is one characteristic of oxide structures that is indicative of their crystalline properties and changes with the a/b ratio [16].

$LiNbO_3$ is a ferroelectric material with a more complex ilmenite structure belonging to the perovskite family, and it has a wide compositional range as a solid solution [17]. However, its composition deviates from stoichiometric only toward the Nb-rich side, since Li does not replace Nb at the Nb site. This behavior cannot be simply explained in terms of the ionic radius or electronegativity of the elements but is rather related to the inhibition of oxygen deficiency formation as a means of charge compensation due to the replacement of Nb^{5+} by Li^{+}. In extreme cases, the stoichiometric composition of some semiconducting materials, such as SnTe, does not belong to its solid solution range [18]. The oxidation of such materials is thought to be governed by the growth atmosphere, although the ease or difficulty of oxidation may also depend on the structure and the composition of the oxides. Coloring due to the presence of impurities or oxygen vacancies that act as color centers is also often observed in oxide crystals when they are grown or annealed in an insufficient oxygen atmosphere.

A simple consideration of the presence of interstitial impurities and/or oxygen vacancies may not be sufficient to confirm defect structures in oxides, since point defects in nonstoichiometric materials are not necessarily randomly distributed but rather may be ordered at high concentrations in cases where this ordering reduces the free energy of the structure. These point defects may also interact with one another [19], and one consequence of this interaction may be the formation of clusters. As an example, vacancy clustering has been proposed for Fe vacancies in a $Fe_{1-x}O$ phase [20] (Figure 4.1). Interactions between defects can also form a superlattice; a 13:4 Koch–Cohn cluster is an example of one such superlattice [21]. On the other hand, vacancies, interstitials, and clusters, including their origins and formation mechanisms, have been well studied during Si crystal growth on both a theoretical and experimental basis [22–26].

Regarding analytical studies of point defects, thermodynamics and statistical thermodynamics are useful for understanding the generation and distribution of point defects from the energetic point of view, not only with regard to the random distribution of defects at low defect densities but also as a means of understanding the ordering and clustering of highly populous defects by taking into account the interactions between defects [27,28].

It is thus important to be aware of and to consider the quantity of impurities or vacancies. In this chapter, point defects will be discussed in association with the solid solution. Impurities or vacancies replacing the host elements at lattice points are often 10^{-4} to 10^{-2} mol or more in quantity if vacancies are required for charge compensation due to valence differences between the host element and the impurity. A thermally

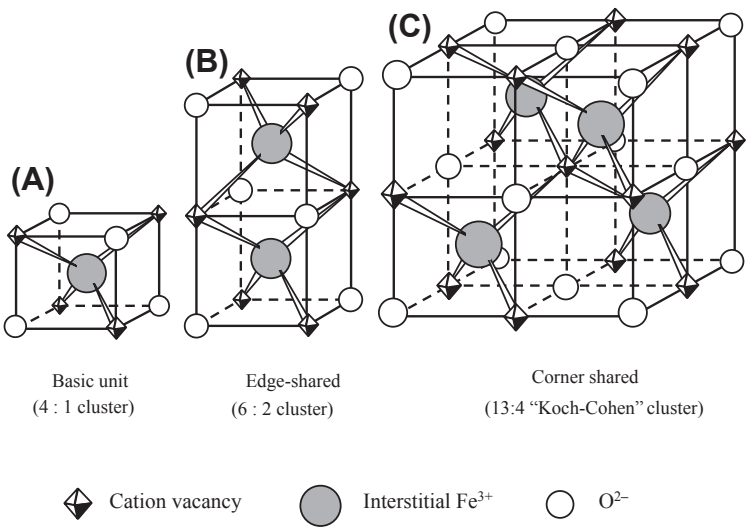


FIGURE 4.1 Vacancy clustering and cluster association of iron-deficient $Fe_{1-x}O$. (A) The basic unit: a so-called 4:1 cluster consisting of four vacancies and one interstitial Fe^{3+} ion. (B) Two basic units combined to form an edge-shared 6:2 cluster complex. (C) Four basic units combined to form a corner-shared 13:4 cluster complex representing a superlattice known as a Koch–Cohen cluster [21].

activated vacancy is one of the intrinsic vacancies and can become enriched, depending on the band gap of the material, near the melting temperature of the material. However, the amount of vacancies of this type or of color-causing impurity centers is much smaller than the quantity of defects replacing the host element, by at least two orders of magnitude at high temperatures. For this reason, intrinsic point defects are disregarded in the following discussion.

4.1.2 Crystal Sites in Oxide Crystals

An oxide crystal consists of an oxygen framework that provides space for a metal cation based on its coordination number. For descriptive purposes, the space is often referred to as a crystal site, as in *tetrahedral* (4-coordination), *octahedral* (6-coordination), and *decahedral* (8-coordination) sites. Each of the constituent cations of the crystal is located in its specific site depending on its radius, valence, and electronegativity, such that attractive and repulsive forces between oxygen atoms and cations are in equilibrium. Each site in a conventional stoichiometric compound is completely filled with a specific element without any foreign elements or vacancies, so that a simple atomic ratio may be used to describe the relative quantities of the constituent elements; this is the law of multiple proportions [1]. Stoichiometric compositions of this type are readily described and understood without any uncertainties. The conventional stoichiometric compound $LiNbO_3$ (s-LN), for example, contains three crystal sites: the Li, Nb, and O sites, each of which is completely filled with Li, O, and Nb, without any foreign elements or vacancies. In contrast, conventional congruent-melting $LiNbO_3$ (c-LN) has completely filled Nb and O sites, while the Li site is occupied not only by Li but also by Nb (antisite Nb) as well as by vacancies necessary for charge compensation. Therefore, the components of this material consist of Li, Nb, and vacancies.

There are always issues concerning the extent to which an oxide crystal changes its oxygen content when it is grown or annealed in a different oxygen atmosphere. In cubic crystals, such as FeO, NiO, ZnO, and VO_5, and tetragonal crystals, such as TiO_2, the metal-to-oxygen ratio varies readily, concomitant with a change in the metal ion oxidation states. The oxygen quantity varies significantly, and this variation is explicitly reflected in the chemical formula of the substance. Although many questions remain unanswered regarding the oxygen deficiency mechanism associated with the defect structures, the formation energies of metal oxides available in the Ellingham diagram provide a good indication of whether these oxides easily lose oxygen upon exposure to a reduction atmosphere. The oxidation states of iron, nickel, and copper change readily according to the oxygen atmosphere because the transition energies between different oxidation states are small for these metal oxides. In contrast, once oxidized, lithium, magnesium, aluminum, and titanium cannot be reduced because they have elevated oxidation energies. $LiNbO_3$ alone barely loses 0.01–0.1 mol of oxygen when annealed under reducing atmosphere. However, when doped with Fe^{3+} ions and annealed under reducing atmosphere, $LiNbO_3$ easily loses oxygen in an amount equivalent to the oxygen

loss obtained for the reduction of Fe_2O_3 to FeO contained in $LiNbO_3$. The reverse process (oxidation) is also the case. Metal oxides containing elements that evaporate readily exhibit more complex defect structures because mass conservation rules do not hold. These materials include $PbTiO_3$, PZNT ($Pb(Zn_{1/3}Nb_{2/3})O_3$–$PbTiO_3$), PMNT ($Pb(Mg_{1/3}Nb_{2/3})O_3$–$PbTiO_3$) [29–31], and $La_3Ga_5SiO_{14}$ [32,33].

It is also well known that some of oxides exhibit a color change from colorless to yellow, orange, or even black when they are grown or annealed in an oxygen-poor atmosphere. However, the amount of oxygen that must be lost to yield a color center is often on the order of ppm or less, whereas the quantity of point defects associated with forming a solid solution will range from 0.01 to 0.1 mol. This chapter focuses on oxides with relatively complex structures, such as garnet, langasite, and perovskites involving ilmenite, such as $LiNbO_3$ or $LiTaO_3$, in which oxygen is assumed to form a rather stable framework and the oxygen sites are fully saturated with oxygen, although they may still contain a small quantity of oxygen deficiencies that act as color centers.

4.1.3 Partitioning Behavior of Ionic Solutes in an Oxide Melt

Typically, defect structures have been reviewed and discussed in association only with the solid state. Many excellent articles and books have been published concerning point defects in oxides [15,34–38], however, almost all of these regard defects only as a solid state issue; few publications have discussed defects in relation to crystallization from the melt, with the exception of some works regarding Si [22–26]. Since impurities or even vacancies are partitioned from the melt into the crystal during crystallization, it is imperative to discuss defects by relating their formation, characteristics, and thermodynamic stability to the coexistence of solid and liquid during crystallization.

Solute partitioning behavior is generally investigated using an equilibrium phase diagram. However, most oxide phase diagrams are pseudo-phase diagrams where the stoichiometry of oxygen to metal always holds and the oxygen is not considered an independent component. For this reason, conventional oxide phase diagrams do not take into account the ionization of neutral species that actually occurs in the oxide melt. Ionic species, which represent one of the major sources of point defects, are present in the melt [39] and are partitioned into the solid during crystallization [40–43]. Point defects, therefore, should be discussed in association with crystallization, and the oxygen in the melt should be considered as an independent component in the phase diagram. As will be discussed in detail in Section 4.3.1, the oxygen ions and metal ionic species behave independently, which has been experimentally demonstrated by the observation of crystallization electromotive force (c-EMF) generated by the segregation of ionic species at the solid–liquid interface, even in the case of c-$LiNbO_3$. Therefore, although the $LiNbO_3$ congruent point in the pseudo-binary diagram of Li_2O–Nb_2O_5 is invariant, it is not invariant in the ternary diagram of Li–Nb–O, and thus a variety of population ratios among the ionic species are allowed.

If we simply assume that $LiNbO_3$ is dissociated into Li_2O and Nb_2O_5 and then ionized into further species, such as Li^+, OLi^-, Nb^{5+}, $Nb_2O_4^{2+}$, and O^{2-} [39], in the ternary diagram of Li-Nb-O, all of these species can be regarded as components rather than just Li_2O and Nb_2O_5 that are valid in the pseudo-binary diagram of Li_2O–Nb_2O_5. The number of components therefore increases from two to five. However, even this is not entirely correct since we also have ongoing chemical reactions such as those shown in Eqns (4.1) and (4.2) following:

$$OL^- \rightleftarrows O^{2-} + Li^+ \tag{4.1}$$

$$Nb_2O_4^{2+} \rightleftarrows 2Nb^{5+} + 4O^{2-} \tag{4.2}$$

Accordingly, we can write the chemical potential relationships

$$\mu^{OL^-} = \mu^{O^{2-}} + \mu^{Li^+} \tag{4.3}$$

and

$$\mu^{Nb_2O_4^{2+}} = 2\mu^{Nb^{5+}} + 4\mu^{O^{2-}}. \tag{4.4}$$

Equations (4.3) and (4.4) represent constraints that reduce the number of independent components to three species: Li^+, Nb^{5+}, and O^{2-}. The above chemical species are hypothetical so that the general forms of the Li-bearing species and Nb-bearing species may be written as

$$Li(1),\ Li(2),\ \ldots,\ Li(k) \tag{4.5}$$

and

$$Nb(1), Nb(2),\ \ldots, Nb(l). \tag{4.6}$$

The oxygen ion, O^{2-}, is also a constituent chemical species. It therefore appears that the number of components is $k + l + 1$, however, the independent components are only Li^+, Nb^{5+}, and O^{2-} since other components are complexes of Li and O or Nb and O so that constraints are in effect similar to those presented in Eqns (4.3) and (4.4), meaning that

$$\mu^{Li(i)} = m\mu^{Li^+} + n\mu^{O^{2-}} \tag{4.7}$$

and

$$\mu^{Nb(j)} = s\mu^{Nb^{5+}} + t\mu^{O^{2-}}. \tag{4.8}$$

Since there are three components instead of two, one more degree of freedom is available in the melt, which enables the population ratio to vary among the ionic species even for a congruent-melting composition. This population ratio will in turn depend on the growth conditions, such as growth velocity and temperature gradient. Although it initially seems strange that the ratio between ionic species changes with growth velocity while the congruent melt composition does not, this occurs because the congruent composition is defined on the basis of the Li_2O–Nb_2O_5 pseudo-binary associated with $LiNbO_3$, and this is invariant regardless of the growth velocity. In contrast, the presence of ionic species requires the oxygen ion to be considered as an

independent component, which adds one more degree of freedom to the ternary system and enables population change among the ionic species with growth velocity.

However, when these ionic species are partitioned into the solid, this one degree of freedom in the melt phase must be consumed in such a way that the partition coefficient for each ionic species is determined. Otherwise, the solid composition cannot be fixed in the pseudo-phase diagram. It should be noted that the physical phenomenon with a higher degree of freedom does not surpass the one governed by the lower degree of freedom. Thus, the same number of constraints as that for the melt species is required to solve the partition coefficient for each species. The constraint that consumes the one degree of freedom must then be determined. Because the composition of the solid phase is determined in the pseudo-phase diagram, a stoichiometric ratio between the metal and oxygen atoms in the solid is required, which is the last constraint necessary to achieve zero degrees of freedom.

4.2 Extended Concept of Stoichiometry in Oxide Crystals

Nonstoichiometric compounds often demonstrate interesting physical properties that may be induced by tuning their chemical compositions. Examples include variation of the Curie temperature and dielectric constant of ferroelectric materials used in ceramic capacitors such as (Ba,Sr)TiO_3 with the addition of small amounts of Y, Nd, and Sm [44,45] and the large piezoelectric constant associated with the Curie temperature of PZNT or PMNT [29–31] at compositions near the polymorph phase boundary (MPB), which enable applications in ultrasonic sensors and other piezoelectric applications.

Due to its ordered configuration of constituent elements on the lattice points, the stoichiometric compound will exhibit certain superior physical properties to the nonstoichiometric compound. Stoichiometric $LiNbO_3$ (s-LN) or $LiTaO_3$ (s-LT) crystals [2,3], for instance, show a higher conversion efficiency for secondary harmonic generation (SHG) in nonlinear optical applications and much lower coercive electric fields associated with poling compared to c-LN and congruent $LiTaO_3$ (c-LT) crystals. However, these stoichiometric compositions are rigid, and only a small number of stoichiometric crystals are available that correspond to the law of multiple proportions [1]. Thus, it would be beneficial from both scientific and technological perspectives if stoichiometric compounds could be obtained over a wider compositional range. In this chapter, by considering the essential concept of stoichiometry, we redefined *stoichiometry* in such a way that a material in which the activities of all the constituent elements can be unity is stoichiometric [46,47]. The constituent elements in this definition include not only the primary constituent atoms but also the impurities and vacancies that replace these primary atoms. This revised concept of stoichiometry expands the current definition of stoichiometric composition from a single point to a range described by a line.

4.2.1 The Thermodynamic Definition of Stoichiometry Associated with the Activity of Constituent Elements

The application of thermodynamic principles throughout a broad spectrum of scientific and engineering disciplines may be considered to fall into two categories. One approach is related to solving practical problems regarding the conversion efficiency between different kinds of energy and is concerned with the reaction constant, the activity of elements during the chemical reaction process and so on. The other involves analytical considerations and is more conceptual in nature. Our development of an extended stoichiometry will use the latter approach.

One of the thermodynamic principles used for this purpose is *activity*, a term that is associated with the standard-state chemical potential. The chemical potential of component j (termed μ^j) in a nonstoichiometric material is given by Eqn (4.9):

$$\mu^j = \mu_0^j + RT \ln a^j, \tag{4.9}$$

where μ_0^j is the standard-state chemical potential and a^j is the activity for j that indicates mixing with another solute. It should be noted that the chemical potential, μ^j, has a definite value, while the activity, a^j, is variable. Once μ_0^j is chosen, a^j is determined accordingly using the relationship

$$\ln a^j = \frac{\mu^j - \mu_0^j}{RT}. \tag{4.10}$$

The value of the standard-state chemical potential, μ_0^j, is essentially chosen based on one's preference, although in most cases it represents the chemical potential of a pure substance at 1 atm and 298.15 K. This is because the energy of formation of pure substances is typically readily available, which is helpful for many thermodynamic calculations associated with practical applications.

The essence of "stoichiometry" may be elucidated by comparing it with "nonstoichiometry" from a thermodynamic perspective. A nonstoichiometric composition is derivative from a stoichiometric composition by the addition of impurities, which implies that the activity of the chemical species or element j is not unity,

$$a^j \neq 1. \tag{4.11}$$

In contrast, the chemical potential of the stoichiometric composition has no mixing term, and thus

$$\mu^j = \mu_0^j, \tag{4.12}$$

which is equivalent to saying that the activity of species j is unity,

$$a^j = 1. \tag{4.13}$$

We should note here that the statement $a^j = 1$ also implies that the mole fraction is unity ($x^j = 1$), which means that the standard-state composition about j reflects the bulk composition, j of the material itself regardless of the quantity of j in the material. It should be noted again that the standard-state composition does not have to be purely j.

In order for the material to be stoichiometric, Eqn (4.13) must be valid for any component, *j*. That is, the activity of all constituent components in the stoichiometric material is unity. Conversely, if the activity of all constituent components is unity, the material can be considered to be stoichiometric [46,47]. Following, we will demonstrate the validity of an extended stoichiometry concept by comparing this concept with the conventional understanding of stoichiometry.

4.2.2 Conventional Stoichiometric Compositions and the Law of Multiple Proportions

The composition of a conventional stoichiometric crystal is represented by a small integer that gives the ratio of constituent elements, representing the so-called law of multiple proportions [1]. Here we use two kinds of stoichiometric crystals to demonstrate different types of compositions: (1) $Li_2B_4O_7$, which is a line-compound, and (2) s-LN, which belongs to a solid solution with a wide compositional range.

$Li_2B_4O_7$ has a very limited solid solution range, therefore it can be considered as a line compound (Figure 4.2(A)). This solid solution range is determined by investigating the limits of the compositional shift from the stoichiometric composition [48], using XRD or DTA to determine whether or not the secondary phase is present in the residual melt after normal freezing under vigorous mixing of the melt. The solid solution was found to vary around the stoichiometric composition by a margin of ±0.3 mol% Li_2O [48]. However, the observation of zero c-EMF, which will be explained in Section 4.3.1, revealed that $Li_2B_4O_7$ is a line compound with almost zero solid solution range. It should be noted here that there may still be some limited compositional variation in this

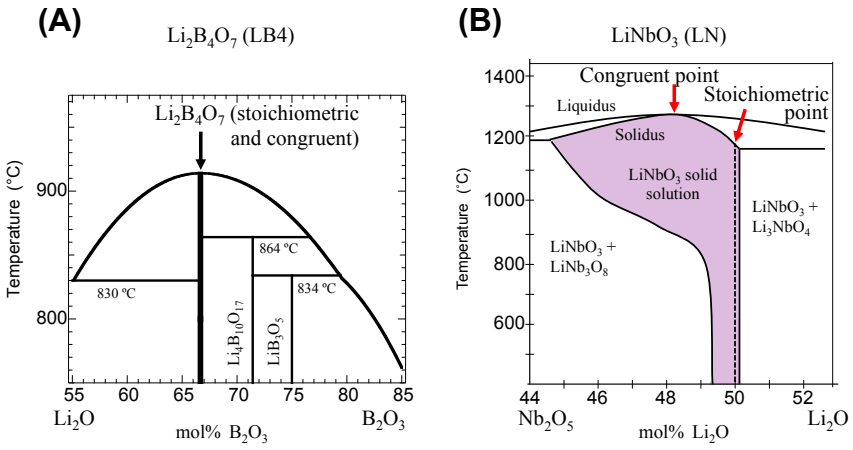


FIGURE 4.2 Equilibrium phase diagrams of pseudo-binary systems. (A) Li_2O–B_2O_3 around $Li_2B_4O_7$ *(after Ref. [48])* and (B) Li_2O–Nb_2O_5 around $LiNbO_3$ *(after Ref. [17])*. Note that the stoichiometric composition and the congruent point coincide in $Li_2B_4O_7$ but not in $LiNbO_3$.

material due to the presence of thermally activated vacancies. However, this small amount of compositional variation is neglected in the present discussion. The most important feature of this line-compound stoichiometric crystal is that it is coincident with the congruent point.

The phase diagram of the pseudo-binary system of Li_2O–Nb_2O_5 is presented [17] in Figure 4.2(B), in which conventional stoichiometric $LiNbO_3$ (s-LN) and conventional congruent-melting $LiNbO_3$ (c-LN) are indicated. In contrast to the stoichiometric line-compound $Li_2B_4O_7$, s-LN is separate from c-LN and is located at the far end of Li_2O composition in the $LiNbO_3$ solid solution. Although both $Li_2B_4O_7$ and s-LN are stoichiometric, $Li_2B_4O_7$ is coincident with the congruent point while s-LN is not, which reflects the different equilibrium partitioning of their constituent elements. This will be discussed in detail in Section 4.3.2.

4.2.2.1 *Activities of the Constituent Elements in a Conventional Stoichiometric Crystal*

In the following discussion, we will demonstrate that the activities of all constituent elements in a conventional stoichiometric compound are unity [46]. The composition of a conventional stoichiometric compound is represented by a small integer ratio between constituent elements, and as such the line-compound $Li_2B_4O_7$ may be employed as a conventional stoichiometric material.

In both the $Li_2B_4O_7$ melt and the solid at its crystallization temperature, the most likely possibility would be for $Li_2B_4O_7$ to dissociate into Li_2O and B_2O_3 [46], as shown by Eqn (4.14), which is analogous to the dissociation of $LiNbO_3$ [39], since the $Li_2B_4O_7$ crystal consists of firm B-O frameworks with Li in the pore spaces.

$$Li_2B_4O_7 \rightleftarrows Li_2O + 2B_2O_3 \tag{4.14}$$

Subsequently, each species would be expected to ionize. For instance, Li_2O may ionize at high temperature, as shown in Eqn (4.15):

$$Li_2O \rightleftarrows Li^+ + OLi^- \tag{4.15}$$

In the following discussion, the chemical species j are differentiated in two ways: $\bar{j}$ is the net chemical species that normally appears as a component in the pseudo-binary phase diagram, while j is the actual chemical species existing in the solid or melt and that forms the net chemical species $\bar{j}$. $\overline{Li_2O}$ and $\overline{B_2O_3}$ are thus the net species while Li_2O, Li^+, OLi^-, and others represent the proposed real chemical species, j, that form $\overline{Li_2O}$. Thus, $Li_2B_4O_7$ may be decomposed into several chemical species as indicated by Eqn (4.16):

$$n\,Li_2B_4O_7 \rightleftarrows a\,Li_2O + b\,Li^+ + c\,OLi^- + d\,B_2O_3 \tag{4.16}$$

At this time, we do not take into account the further decomposition of B_2O_3. The chemical potential of $Li_2B_4O_7$ is written as Eqn (4.17):

$$n\,\mu^{Li_2B_4O_7} = a\,\mu^{Li_2O} + b\,\mu^{Li^+} + c\,\mu^{OLi^-} + d\,\mu^{B_2O_3}, \tag{4.17}$$

and

$$n\,\mu^{\mathrm{Li_2B_4O_7}} = n\,\mu^{\overline{\mathrm{Li_2O}}} + 2n\,\mu^{\overline{\mathrm{B_2O_3}}}. \tag{4.18}$$

Equation (4.19) then describes B_2O_3:

$$2n\,\mu^{\overline{\mathrm{B_2O_3}}} = d\,\mu^{\mathrm{B_2O_3}} \tag{4.19}$$

The relationship between the net chemical species and the real chemical species for the Li_2O component can then be written for both the solid and liquid assuming that all chemical species in the melt are partitioned into the solid so that the discussion can be made based on the equilibrium partitioning of each chemical species. Equations (4.17), (4.18), and (4.19) yield Eqn (4.20):

$$n\,\mu_\beta^{\overline{\mathrm{Li_2O}}} = a\,\mu_\beta^{\mathrm{Li_2O}} + b\,\mu_\beta^{\mathrm{Li^+}} + c\,\mu_\beta^{\mathrm{OLi^-}}, \tag{4.20}$$

where β is either the solid (S) or liquid (L) phase. This describes the relationship in chemical potential between the net $\overline{\mathrm{Li_2O}}$ and Li-bearing species, and a similar relationship can be made between the net $\overline{\mathrm{B_2O_3}}$ and B-bearing species. Thus, hereon we will deal exclusively with Li_2O.

The above chemical species are hypothetical, and it is not clear that they actually exist. Thus, Eqn (4.21) is hereafter used for the general discussion instead of Eqn (4.20):

$$n\,\mu_\beta^{\overline{\mathrm{Li_2O}}} = m_1\,\mu_\beta^{\mathrm{Li}(1)} + m_2\,\mu_\beta^{\mathrm{Li}(2)} + m_3\,\mu_\beta^{\mathrm{Li}(3)} + \cdots + m_k\,\mu_\beta^{\mathrm{Li}(k)}, \tag{4.21}$$

where Li(i) are the chemical species containing Li and for which $m_i \geq 0$. It is important to note that the Li-bearing species in Eqn (4.21) are assumed to be common to both the solid and liquid states since all the chemical species in the melt are partitioned into the solid during growth. Specifically, the chemical potential of the net $\overline{\mathrm{Li_2O}}$ in the solid at a given temperature T may be written as

$$n\,\mu_{S(T)}^{\overline{\mathrm{Li_2O}}} = m_1\,\mu_{S(T)}^{\mathrm{Li}(1)} + m_2\,\mu_{S(T)}^{\mathrm{Li}(2)} + m_3\,\mu_{S(T)}^{\mathrm{Li}(3)} + \cdots + m_k\,\mu_{S(T)}^{\mathrm{Li}(k)}, \tag{4.22}$$

and this equation may be rewritten as Eqn (4.23) by breaking down the chemical potential into the standard-state chemical potential and the mixing term as

$$\begin{aligned} &n\,\mu_{S(T,0)}^{\overline{\mathrm{Li_2O}}} + n\,RT\ln a_{S(T)}^{\overline{\mathrm{Li_2O}}} - \left(m_1\,\mu_{S(T,0)}^{\mathrm{Li}(1)} + m_2\,\mu_{S(T,0)}^{\mathrm{Li}(2)} + m_3\,\mu_{S(T,0)}^{\mathrm{Li}(3)} + \cdots + m_k\,\mu_{S(T,0)}^{\mathrm{Li}(k)}\right) \\ &= \Delta\mu_{S(T,0)}^{\mathrm{Li}^*} + n\,RT\ln a_{S(T)}^{\overline{\mathrm{Li_2O}}} = RT\left(m_1\ln a_{S(T)}^{\mathrm{Li}(1)} + m_2\ln a_{S(T)}^{\mathrm{Li}(2)} + m_3\ln a_{S(T)}^{\mathrm{Li}(3)} + \cdots + m_k\ln a_{S(T)}^{\mathrm{Li}(k)}\right) \end{aligned} \tag{4.23}$$

where $\mu_{S(T,0)}^{j}$ is the standard-state chemical potential of the solid about j at a temperature T, and $\Delta\mu_{S(T,0)}^{\mathrm{Li}^*}$ is the difference in standard-state chemical potentials between $\overline{\mathrm{Li_2O}}$ and the dissolved species, Li(i), which is written as

$$\Delta\mu_{S(T,0)}^{\mathrm{Li}^*} = n\,\mu_{S(T,0)}^{\overline{\mathrm{Li_2O}}} - \left(m_1\,\mu_{S(T,0)}^{\mathrm{Li}(1)} + m_2\,\mu_{S(T,0)}^{\mathrm{Li}(2)} + m_3\,\mu_{S(T,0)}^{\mathrm{Li}(3)} + \cdots + m_k\,\mu_{S(T,0)}^{\mathrm{Li}(k)}\right). \tag{4.24}$$

Here, the standard-state chemical potential is defined for $\mu_{S(T,0)}^{\overline{\mathrm{Li_2O}}}$ and $\mu_{S(T,0)}^{\mathrm{Li}(i)} (1 \leq i \leq k)$, individually: $\mu_{S(T,0)}^{\overline{\mathrm{Li_2O}}}$ is defined as the chemical potential of the solid for the exact Li_2O

concentration in the stoichiometric composition of $Li_2B_4O_7$ at a temperature T. The activity, $a_{S(T)}^{\overline{Li_2O}}$, is then written as

$$\ln a_{S(T)}^{\overline{Li_2O}} = \frac{\left(\mu_{S(T)}^{\overline{Li_2O}} - \mu_{S(T,0)}^{\overline{Li_2O}}\right)}{RT}. \tag{4.25}$$

The term $\mu_{S(T,0)}^{Li(i)}$ can be defined at a given temperature T in such a way that the $\Delta\mu_{S(T,0)}^{Li^*}$ term in Eqn (4.24) becomes zero. This is allowed because the Li, B, and O sites are fully occupied by their own specific elements in stoichiometric $Li_2B_4O_7$, which provides one degree of freedom to define $\mu_{S(T,0)}^{Li(i)}$. This freedom associated with the crystal sites and its usage is discussed in more detail in Section 4.2.3.1. Thus $\mu_{S(T,0)}^{Li(i)}$ is defined as in Eqn (4.26):

$$n\,\mu_{S(T,0)}^{\overline{Li_2O}} = m_1\,\mu_{S(T,0)}^{Li(1)} + m_2\,\mu_{S(T,0)}^{Li(2)} + m_3\,\mu_{S(T,0)}^{Li(3)} + \cdots + m_k\,\mu_{S(T,0)}^{Li(k)}, \tag{4.26}$$

where the appropriate value can be assigned to each $\mu_{S(T,0)}^{Li(i)}$ term such that the equation is satisfied at temperature T, even though $\mu_{S(T,0)}^{Li(i)}$ is not individually specified. Combining Eqn (4.26) with Eqn (4.23) gives

$$n \ln a_{S(T)}^{\overline{Li_2O}} = m_1 \ln a_{S(T)}^{Li(1)} + m_2 \ln a_{S(T)}^{Li(2)} + m_3 \ln a_{S(T)}^{Li(3)} + \cdots + m_k \ln a_{S(T)}^{Li(k)}, \tag{4.27}$$

and thus

$$\left(a_{S(T)}^{\overline{Li_2O}}\right)^n = \left(a_{S(T)}^{Li(1)}\right)^{m_1}\left(a_{S(T)}^{Li(2)}\right)^{m_2}\left(a_{S(T)}^{Li(3)}\right)^{m_3}\cdots\left(a_{S(T)}^{Li(k)}\right)^{m_k}. \tag{4.28}$$

$Li_2B_4O_7$ is peculiar in that the solid solution range of $Li_2B_4O_7$ is sufficiently small such that the composition of the solid $Li_2B_4O_7$ coexisting with the liquid can be assumed to be nearly constant and equal to the stoichiometric composition at any given temperature, T, which is coincident with the congruent composition. Thus, Eqns (4.29) and (4.30) hold true at any temperature. As such, the chemical potential for the bulk Li_2O, $\mu_{S(T)}^{\overline{Li_2O}}$, is assigned to the standard-state chemical potential, $\mu_{S(T)}^{\overline{Li_2O}}$, as in Eqn (4.29):

$$\mu_{S(T)}^{\overline{Li_2O}} = \mu_{S(T,0)}^{\overline{Li_2O}} \tag{4.29}$$

Combining this equation with Eqn (4.25), Eqn (4.30) is obtained:

$$a_{S(T)}^{\overline{Li_2O}} = X_{S(T)}^{\overline{Li_2O}} = 1, \tag{4.30}$$

where $X_{S(T)}^{\overline{Li_2O}}$ is the net concentration of Li_2O in the solid $Li_2B_4O_7$ at a temperature, T. Combining with Eqn (4.28), the product of $a_{S(T)}^{Li(i)}$ is unity as shown by Eqn (4.31):

$$\left(a_{S(T)}^{Li(1)}\right)^{m_1}\left(a_{S(T)}^{Li(2)}\right)^{m_2}\left(a_{S(T)}^{Li(3)}\right)^{m_3}\cdots\left(a_{S(T)}^{Li(k)}\right)^{m_k} = 1. \tag{4.31}$$

Equation (4.31) holds at any temperature, T, and thus we can write Eqns (4.32) and (4.33):

$$a_{S(T)}^{Li(1)} = a_{S(T)}^{Li(2)} = a_{S(T)}^{Li(3)} = \cdots = a_{S(T)}^{Li(k)} = 1 \tag{4.32}$$

$$X_{S(T)}^{\mathrm{Li}(1)} = X_{S(T)}^{\mathrm{Li}(2)} = X_{S(T)}^{\mathrm{Li}(3)} = \cdots = X_{S(T)}^{\mathrm{Li}(k)} = 1 \tag{4.33}$$

As in Eqn (4.25), $a_{S(T)}^{\mathrm{Li}(i)}$ is defined as

$$\ln a_{S(T)}^{\mathrm{Li}(i)} = \frac{\left(\mu_{S(T)}^{\mathrm{Li}(i)} - \mu_{S(T,0)}^{\mathrm{Li}(i)}\right)}{RT}. \tag{4.34}$$

The value of $\mu_{S(T,0)}^{\mathrm{Li}(i)}$ in Eqn (4.34) is determined in order to satisfy Eqn (4.26), and $a_{S(T)}^{\mathrm{Li}(i)} = 1$ leads to $\mu_{S(T)}^{\mathrm{Li}(i)} = \mu_{S(T,0)}^{\mathrm{Li}(i)}$ for any temperature. This same process also applies to the B-bearing species. Overall, then, the activity of all constituent elements in the stoichiometric compound $Li_2B_4O_7$ is unity [46],

$$a_{\mathrm{Li\ site}}^{\mathrm{Li}} = a_{\mathrm{B\ site}}^{\mathrm{B}} = a_{\mathrm{O\ site}}^{\mathrm{O}} = 1. \tag{4.35}$$

4.2.3 Extended Stoichiometric Compositions Including Impurities and Vacancies

In the previous section, we demonstrated that a material in which the activity of each constituent element can be unity represents a stoichiometric material, by assigning each chemical potential, μ^j, to the standard-state chemical potential, μ_0^j. In other words, if we are allowed to assign μ^j to μ_0^j for any component, j, such that $\mu_0^j = \mu^j$, the material can be considered stoichiometric. Stoichiometric materials by this definition may include impurities and vacancies and so may be more stable in terms of entropy than conventional stoichiometric substances. This raises an important objection—if an assignment such as this is possible in any material, then every material could potentially be considered stoichiometric, which is certainly not the case. The question is how we can know whether or not μ^j can be assigned to μ_0^j. To make such an assignment, we require a degree of freedom, since the standard-state chemical potential, μ_0^j, can have any value if one degree of freedom is present in the crystal site where the element j is present.

4.2.3.1 Degrees of Freedom in a Crystal Site

It is important to determine the element occupancy of each crystal site. These elements include constituent cations, impurity ions, antisite defects, and vacancies. The possible element occupancy at a site is examined by considering the associated degrees of freedom, and we can explain the degrees of freedom of a crystal site by employing $LiNbO_3$ as an example. Here the vacancy is a defect that forms in order to compensate for the charge imbalance due to the difference between the valences of the impurity ions and that of the host ion present in a site. The quantity of vacancies will be on the order of 10^{-4} to 10^{-2} mol or more depending on the population of impurity ions or antisite defects. In the following discussion, we assume that the oxygen sites are saturated with oxygen and that no oxygen vacancies are present even when $LiNbO_3$ is exposed to an

oxygen-reduced atmosphere during the growth or annealing processes. Such a reduced atmosphere is known to generate a color center in the crystal, with a concurrent change from colorless to yellow or orange, but the accompanying extent of oxygen deficiency is much less than the degree of oxygen vacancy required for the charge compensation, by at least two orders of magnitude. Thus, the possible presence of elements at each cation site will be discussed, assuming that oxygen saturation is maintained.

The degrees of freedom are obtained by subtracting the number of constraints from the number of parameters. In this case, the number of parameters at a site is the number of elements, based on the following three constraints:

1. Mass conservation holds at each site. That is, the sum of the mole fractions of each constituent element, j (where $j = 1$ to C), is unity, as in Eqn (4.36):

$$X^1 + X^2 + \cdots + X^C = 1 \tag{4.36}$$

2. If an element is present at multiple sites in a crystal, its chemical potentials at those sites are equal, thus

$$\mu^j_{\text{site 1}} = \mu^j_{\text{site 2}} = \cdots. \tag{4.37}$$

3. The vacancy population is calculated in such a way that overall charge neutrality is maintained in the bulk crystal.

These three constraints are necessary conditions, although additional restrictions may be added to decrease the degree of freedom at a given site.

4.2.3.2 Activity of the Constituent Elements in an Extended Stoichiometric Crystal

Based on our newly defined "stoichiometry," referring to a material in which the activity of all the constituent elements can be unity, we extend the stoichiometric composition from a single point to a linear series. Introducing impurities and vacancies is necessary in this case, which is not allowed in a conventional stoichiometric compound. As an example, we may consider MgO-doped $LiNbO_3$ in the pseudo-ternary system of Li_2O–Nb_2O_5–MgO, shown in Figure 4.3. The isoconcentration line at 50 mol% Nb_2O_5 is termed line *A*, and the MgO-doped $LiNbO_3$ on line *A* is termed 50Nb-LN. The conventional stoichiometric compound $LiNbO_3$ (s-LN), the conventional congruent-melting $LiNbO_3$ (c-LN) on the pseudo-binary line of Li_2O–Nb_2O_5, and c-LN doped with 5 mol% MgO (5MgO:LN) are also indicated.

The site structures of these crystals are presented in Figure 4.3(B). All the sites of conventional s-LN are filled with Li, Nb, and O, without any excess atoms or deficiencies. Thus, it is readily understood that the activity of each site as well as the activity of the element at each site is unity. The site structures of c-LN and 50Nb-LN are similar in that both have fully occupied Nb and O sites, and their Li sites are occupied with vacancies or cations other than Li, such as antisite Nb in c-LN and Mg in 50Nb-LN. Although the elements present in their Li sites are much the same, c-LN is not

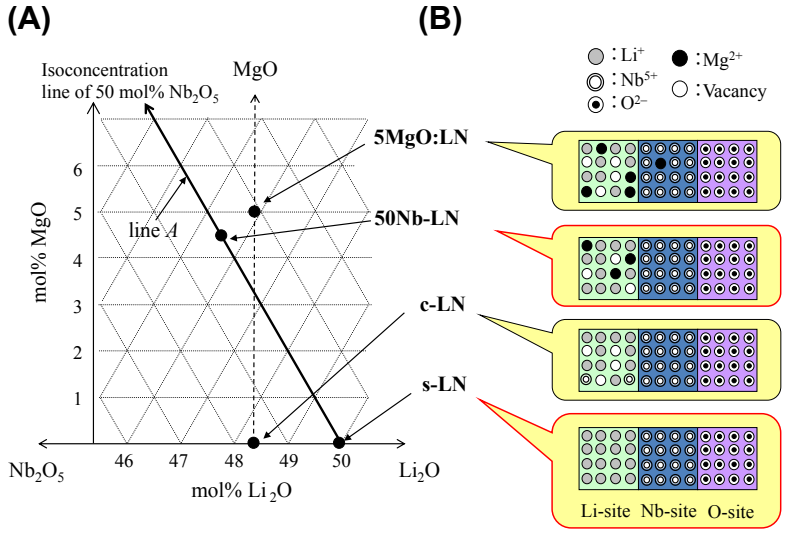


FIGURE 4.3 (A) Composition diagram of various $LiNbO_3$ compositions in the pseudo-ternary system of Li_2O–Nb_2O_5-MgO. The isoconcentration line of 50 mol% Nb_2O_5 line is termed line *A* and represents an O-to-Nb ratio of exactly 3.0. Legend: s-LN = stoichiometric $LiNbO_3$ on the Li_2O–Nb_2O_5 line, c-LN = congruent $LiNbO_3$ on the Li_2O–Nb_2O_5 line, 50Nb-LN = MgO-doped $LiNbO_3$ on line *A*, 5MgO:LN = c-LN doped with 5 mol% MgO. (B) Crystal site structures for each of the various $LiNbO_3$ crystals.

stoichiometric while 50Nb-LN is [47], a phenomenon that will be discussed subsequently. In 5MgO:LN, the Mg is located at both Li and Nb sites, but no antisite Nb is present at the Li sites.

Whether a crystal is stoichiometric or not is determined by examining the possibility of each of the constituent elements in the crystal having an activity of unity. This question directly applies to the activities of the elements at Li sites in c-LN and 50Nb-LN.

Firstly, we can examine the degrees of freedom associated with the Li sites in c-LN. These sites have three parameters: Li, Nb, and vacancies. The constraints on the Li site include conservation of mass, the exchange of equilibrium Nb between Li and Nb sites, and the vacancy population required for charge compensation. Since the overall number of constraints is three, zero degrees of freedom are available for assigning μ^j to the standard-state chemical potential, μ_0^j. Therefore, the activity of the elements at the Li sites cannot be unity, and it has been proven on a thermodynamic basis that c-LN is not stoichiometric.

The degrees of freedom of the Li sites in 50Nb-LN can subsequently be examined using a similar process. Here the number of parameters is three while the number of constraints is two, since we are only concerned with conservation of mass and the vacancy population. As a consequence, the Li sites have one degree of freedom, which allows unrestricted choice of the standard-state chemical potential, μ_0. It should be noted that, in the 50Nb-LN crystal, each element, including Mg and vacancies, belongs

only to one specific site. However, $\mu_0^j = \mu^j$ must be valid for all three of the elements that may be present at the Li sites ($j =$ Li, Mg, and vacancy) in order for the activities of these elements to be unity. Thus, it appears that three degrees of freedom are required, and hence the challenge is to use one degree of freedom to assign μ to μ_0 for each of the three elements. To address this, one can consider the chemical potentials and activities of the crystal sites as well as the constituent elements in 50Nb-LN. The Nb site is fully occupied by Nb and Nb does not enter into any other site, which is also the case for the O sites and O atoms. Hence, the activity of the Nb and O sites is unity, and simultaneously the activity of Nb at the Nb sites and O at the O sites is unity. Since the Li sites preserve the overall charge neutrality with regard to the Nb and O sites, the Li site must also be neutral and the activity of the Li site must be unity. Thus,

$$a_{\text{Li site}} = a_{\text{Nb site}} = a_{\text{O site}} = 1, \tag{4.38}$$

and

$$a^{\text{Nb}}_{\text{Nb site}} = a^{\text{O}}_{\text{O site}} = 1. \tag{4.39}$$

The chemical potential of the Li sites in 50Nb-LN is calculated by summing that of each element, as follows:

$$\mu^{\text{Li site}} = p\mu^{\text{Li}}_{\text{Li site}} + q\mu^{\text{Mg}}_{\text{Li site}} + r\mu^{\text{Va}}_{\text{Li site}} \tag{4.40}$$

where Va denotes a vacancy and p, q, and r are real numbers. By differentiating the chemical potential into the standard-state chemical potential and the mixing term, Eqn (4.40) may be rewritten as:

$$\mu_0^{\text{Li site}} + RT \ln a_{\text{Li site}} = p\left(\mu_0^{\text{Li}} + \ln a^{\text{Li}}_{\text{Li site}}\right) + q\left(\mu_0^{\text{Mg}} + \ln a^{\text{Mg}}_{\text{Li site}}\right) + r\left(\mu_0^{\text{Va}} + \ln a^{\text{Va}}_{\text{Li site}}\right) \tag{4.41}$$

One degree of freedom allows the standard-state chemical relationship between the Li site and its constituent elements, which is shown in Eqn (4.42):

$$\mu_0^{\text{Li site}} = p\mu_0^{\text{Li}} + q\mu_0^{\text{Mg}} + r\mu_0^{\text{Va}} \tag{4.42}$$

Subsequently, the activity relationship between the Li site and its constituent elements is obtained by combining with Eqn (4.41) to give Eqn (4.43):

$$a_{\text{Li site}} = \left(a^{\text{Li}}_{\text{Li site}}\right)^p \left(a^{\text{Mg}}_{\text{Li site}}\right)^q \left(a^{\text{Va}}_{\text{Li site}}\right)^r = 1 \tag{4.43}$$

Equation (4.43) is valid for any combination of p, q, and r that lies on the iso-concentration line of 50 mol% Nb_2O_5 (line A in Figure 4.3). Therefore, the activity of all three elements is unity [47], expressed as

$$a^{\text{Li}}_{\text{Li site}} = a^{\text{Mg}}_{\text{Li site}} = a^{\text{Va}}_{\text{Li site}} = 1. \tag{4.44}$$

Combining this equation with Eqn (4.39), the activity of all elements in 50Nb-LN can be shown to equal unity, and so it is proven on a thermodynamic basis that 50Nb-LN, a Mg-doped LN on the isoconcentration line of 50 mol% Nb_2O_5, is stoichiometric.

4.3 Growth Characteristics of Stoichiometric Crystals

A line-compound stoichiometric crystal is always congruent, but the conventional stoichiometric oxide crystal accompanying the solid solution is often not congruent. Although both types of crystals have an activity equal to unity for each of their constituent elements, the partitioning of constituent elements from the liquid to the solid during growth differs between the two. This indicates that there is a close relationship between the partitioning behavior of an ionic species, stoichiometry, and the congruency of the growing crystal.

Most oxide melts are electrically conductive [39,49,50], therefore they are considered to contain ionic constituent species in addition to neutral species, as illustrated in Figure 4.4. The existence of ionic species in the oxide melt is easily demonstrated when one considers that molten salts are used as electrolyte solutions in electrochemical studies. The populations of these ionic species are dependent on the composition and temperature of the oxide melt. We have seen that pseudo-oxide phase diagrams, in which the stoichiometry between the metal and oxygen holds, only address the oxide components of the material and do not provide any information concerning ionic species. It is generally believed that the partition coefficients are unity at the congruent composition since the solidus line coincides with the liquidus line at that point, meaning that the solid and liquid compositions are the same. However, this is true only for the oxide components appearing in the pseudo-phase diagram; it is not true when one considers that oxygen ions represent one of the ionic species present in the melt. As an example, the pseudo-binary diagram of Li_2O–Nb_2O_5 is not sufficient to explain the partitioning of ionic species in the $LiNbO_3$ melt. In this case, oxygen should be treated as an independent component, and the ternary Li-Nb-O diagram should be used instead of the pseudo-binary Li_2O–Nb_2O_5 diagram. Each of the constituent species, including ionic species, has its own equilibrium partition coefficient, and it can be shown that the equilibrium partition coefficient of any element in the exact line-compound melt is unity, meaning that there is no segregation of ionic species near the interface during growth. This arises from the fact that the line compound is both stoichiometric and

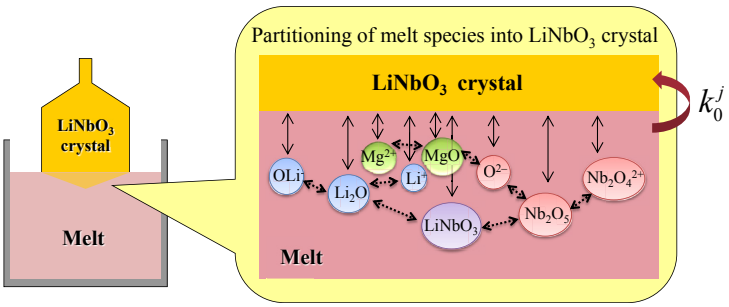


FIGURE 4.4 Neutral and ionic chemical species possibly present in the $LiNbO_3$ melt. Each species, *j*, is partitioned into the solid based on its own unique equilibrium partition coefficient, k_0^j.

congruent. In contrast, even a congruent crystal that is not consistent with the stoichiometric point grows in such a manner that the ionic solutes have nonunity partition coefficients. Consequently, the ionic species in the solute boundary layer undergo segregation, resulting in the accumulation or depletion of these species and generating a c-EMF at the interface, which in turn produces an interface electric field that complicates the growth process.

In this section, c-LN is employed as an example of a congruent crystal that generates a nonzero c-EMF during growth, which demonstrates that the partition coefficients of the ionic species present in the congruent melt are nonunity [51–53], even though the partition coefficients of the net $\overline{Li_2O}$ or $\overline{Nb_2O_5}$ are unity. However, an extended stoichiometric crystal including both impurities and vacancies could coincide with the congruent point at a certain composition, in which case no segregation will occur even if the material is not a line compound. This process will be discussed further in Section 4.3.2.

4.3.1 Crystallization Electromotive Force

Each ionic species, j, is partitioned based on its own equilibrium partitioning coefficient, k_0^j, and is accumulated or depleted at the growth interface due to solute segregation depending on the value of k_0^j, the diffusivity, D^j, and the growth rate, which results in the differential segregation of ions of opposing valences to generate a net charge in the liquid boundary layer as well as a charge of the opposite sign in the crystal, thus producing the c-EMF [40,47,51–53], $\Delta\phi_{EMF}$. This process is illustrated in Figure 4.5. D'yakov et al. [54] first measured the c-EMF generated by c-LN, and

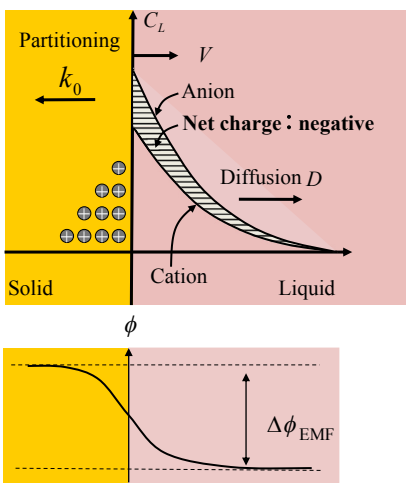


FIGURE 4.5 Schematic illustration of the segregation of ionic species in the liquid boundary layer (above) and the generation of crystallization electromotive force (c-EMF), $\Delta\phi_{EMF}$ (below) during the growth of the $LiNbO_3$ crystal. The horizontal axis shows the distance from the growth interface; the vertical axis shows the net concentrations of anions and cations (above) and electric potential (below) [47].

Aleksandrovskii et al. [55] reported similar measurements using the Czochralski method but did not investigate its origin. The c-EMF is useful not only in terms of growth equilibrium, as discussed herein, but also since it allows the study of growth dynamics. $\Delta\phi_{\mathrm{EMF}}$ is easily measured using a micro-pulling down (μ-PD) system, as shown in Figure 4.6 [47]. The furnace in this apparatus consists of a small platinum (Pt) crucible with a capillary nozzle attached to its base. A sintered material is charged into the Pt crucible and is melted by sending an electric current through the crucible, such that the crucible serves as a resistance heater. Two thermocouples (TC_1, TC_2) are set in the liquid and solid phases; TC_1 is fixed in the liquid, while TC_2 is mobile. During the pulling down (crystal growth stages: 1, 2, and 3), halt and pulling up (crystal melting stages: 4, 5, and 6) of TC_2, T_1 and T_2 (the temperatures of TC_1 and TC_2, respectively) and the potential difference, $\Delta\phi$, between TC_1 and TC_2 are measured. During pulling down and pulling up, the rate of movement of TC_2 is constant. When a *c*-oriented crystal is used to seed the crystal nucleation, the resulting crystal has the same (*c*-) orientation as the seed crystal. During all stages, T_1 is kept nearly constant and only T_2 and $\Delta\phi$ vary. Figure 4.7 depicts an example of a $\Delta\phi - T_2$ curve acquired during the crystal growth process, showing the halt and melting of c-LN [47]. The $\Delta\phi$ term may be represented by Eqn (4.45) [47,51,54]:

$$\Delta\phi = \alpha_L(T_1 - T_i) + \alpha_S(T_i - T_2) + \Delta\phi_{\mathrm{EMF}}, \tag{4.45}$$

where α_L and α_S are the Seebeck coefficients of the liquid and solid phase, and correspond to the slopes of the $\Delta\phi - T_2$ curve in the liquid (stages 1–2 and 5–6 in Figure 4.6) and solid regions (stage 2–3 and 4–5 in Figure 4.6), respectively. The value of α_L is 0.27 mV/K, while α_S is -0.85 mV/K [47], both of which are nearly constant for all compositions. T_i is the temperature at the solid/liquid interface, and $\Delta\phi_{\mathrm{EMF}}$ is the c-EMF that is obtained from the extent of hysteresis generated in the solid region during the

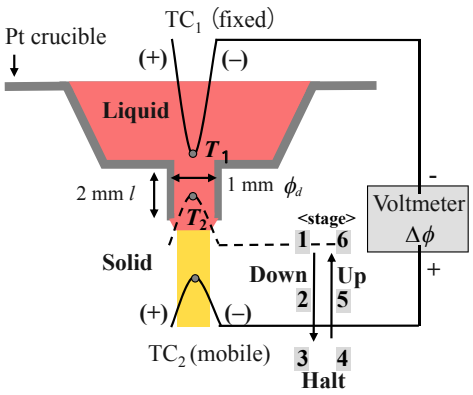


FIGURE 4.6 The micro-pulling-down (μ-PD) system for the measurement of crystallization electromotive force (c-EMF). TC_1 is fixed, while TC_2 is mobile. $\Delta\phi$ is measured by the difference in electric potentials measured at T_1 and T_2 during the pulling down (growth stage 1–3), halt and pulling up (melting stage 4–6) processes. *After Ref. [47].*

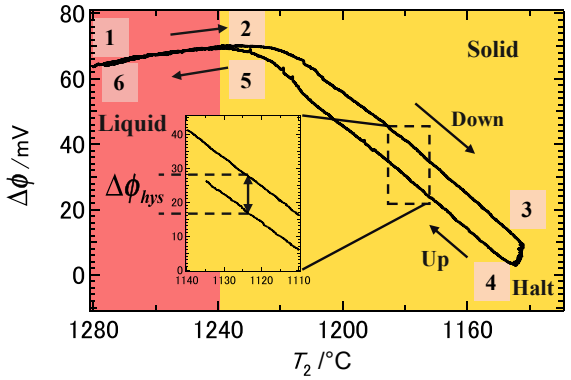


FIGURE 4.7 The $\Delta\phi - T_2$ curve obtained during the growth, halt and melting processes of c-LN [47]. The slope of the $\Delta\phi - T_2$ curve may be used to obtain the Seebeck coefficient of the liquid (α_L) and the solid (α_S), while $\Delta\phi_{EMF}$ is obtained from the hysteresis potential ($\Delta\phi_{hys}$) in the solid region.

solidification and melting process. It should be noted that, in Eqn (4.45), both α_L and α_S are fixed values specific to the material being examined, whereas $\Delta\phi_{EMF}$ is completely growth-system dependent. The value of this term is normally in the range of several mV and can be measured with reasonable accuracy.

The generation of a c-EMF during the growth of $LiNbO_3$ from melts with various compositions demonstrates that the EMF has a nonzero value at the conventional stoichiometric point and even at the congruent point (Figure 4.8) [52]. This EMF arises during growth as the result of the segregation of ionic solutes in the liquid boundary layer in a congruent melt in which the equilibrium partition coefficients are not unity, even though the coefficients of the net $\overline{Li_2O}$ and $\overline{Nb_2O_5}$ are unity, based on interpretation of the conventional pseudo-phase diagram. In contrast, Figure 4.9 shows that the c-EMF observed during the growth of $Li_2B_4O_7$ is zero at the stoichiometric-congruent point but nonzero during growth from the off-stoichiometric melt [52]. The

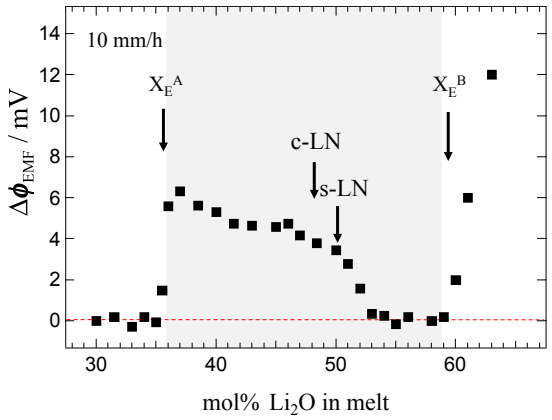


FIGURE 4.8 The c-EMF ($\Delta\phi_{EMF}$) values obtained during the growth of $LiNbO_3$ via a μ-PD measurement system plotted as a function of the Li_2O composition of the melt in the pseudo-binary Li_2O–Nb_2O_5 system at a growth rate of 10 mm/h. X_E^A and X_E^B are the eutectic points in the Nb-rich and Li-rich regions, respectively, while c-LN is the congruent point and s-LN is the stoichiometric composition. Note that $\Delta\phi_{EMF} \neq 0$ even for the c-LN composition. *After Ref. [52].*

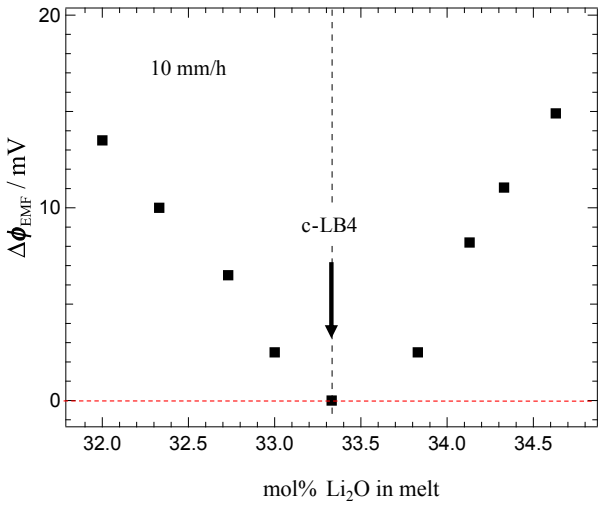


FIGURE 4.9 The c-EMF ($\Delta\phi_{EMF}$) values obtained during the growth of $Li_2B_4O_7$ via a μ-PD measurement system plotted as a function of the Li_2O composition of the melt in the pseudo-binary Li_2O-B_2O_3 system at a growth rate of 10 mm/h. The congruent point of $Li_2B_4O_7$ is c-LB4, which is coincident with its stoichiometric composition. Note that $\Delta\phi_{EMF} = 0$ at the composition for which the congruent point coincides with the stoichiometric composition. *After Ref. [52].*

zero-EMF exhibited by the stoichiometric-congruent composition will always occur during growth at various growth rates, although different growth rates normally result in varying degrees of segregation of the ionic species at the interface if their partition coefficients are not unity. Therefore, no segregation of ionic species occurs at the stoichiometric-congruent point, leading to the conclusion that every ionic species in the $Li_2B_4O_7$ melt has an equilibrium partition coefficient of unity [46]. On this basis, we may conclude that the value of c-EMF will be zero during growth when the crystal is simultaneously stoichiometric and congruent. The opposite is equally true; the crystal is stoichiometric and congruent if no segregation of any constituent species occurs (c-EMF = 0) or if, in other words, the equilibrium partition coefficient of any constituent species is unity. These statements will be proven on a thermodynamic basis in the following sections.

4.3.2 Activities of Constituent Elements in the Melt and Solid States for an Oxide Crystal that is Both Stoichiometric and Congruent

In the previous section, it was demonstrated on an experimental basis that a crystal presents zero c-EMF when it is simultaneously stoichiometric and congruent. The thermodynamic requirement for the coincidence of stoichiometry and congruency in the solid can be analyzed only when one sees that the activities of the constituent species in both the melt and the solid are unity.

The activity of a chemical species, i, in liquid $Li_2B_4O_7$, $a_{L(T)}^{\mathrm{Li}(i)}$, coexisting with the stoichiometric solid has been discussed [46]. It should be noted again that all the species in the liquid are assumed to be identical to those in the solid since they are partitioned from the liquid into the solid during growth. The chemical

potentials of both phases in equilibrium about the net $\overline{Li_2O}$ are therefore equal and we may write

$$\mu_{L(T)}^{\overline{Li_2O}} = \mu_{S(T)}^{\overline{Li_2O}}. \tag{4.46}$$

Since $a_{S(T)}^{\overline{Li_2O}} = 1$ at any temperature, T, it is also true that

$$a_{L(T)}^{\overline{Li_2O}} = \exp\left[\frac{\Delta\mu_{(T,0)}^{\overline{Li_2O}}}{RT}\right], \tag{4.47}$$

where $\Delta\mu_{(T,0)}^{\overline{Li_2O}} = \mu_{S(T,0)}^{\overline{Li_2O}} - \mu_{L(T,0)}^{\overline{Li_2O}}$ and $\mu_{L(T,0)}^{\overline{Li_2O}}$ is the standard-state chemical potential of the liquid phase associated with $\overline{Li_2O}$. Similar to the $\mu_{S(T,0)}^{\overline{Li_2O}}$ term, $\mu_{L(T,0)}^{\overline{Li_2O}}$ is defined as the chemical potential of the liquid associated with the exact Li_2O concentration in the stoichiometric composition of $Li_2B_4O_7$ at a temperature, T. Concerning the equilibrium at the congruent point, $a_{L(T)}^{\overline{Li_2O}} = 1$ in Eqn (4.47), and thus $\Delta\mu_{(T,0)}^{\overline{Li_2O}} = 0$ and $\mu_{S(T,0)}^{\overline{Li_2O}} = \mu_{L(T,0)}^{\overline{Li_2O}}$. Furthermore, $\mu_{L(T)}^{\overline{Li_2O}}$ is transitioned into a linear coupling of the chemical potential of the real liquid chemical species, $\mu_{L(T)}^{\mathrm{Li}(i)}$, in the same manner as Eqn (4.22) regarding $\mu_{S(T)}^{\overline{Li_2O}}$ and $\mu_{S(T)}^{\mathrm{Li}(i)}$, such that the relationship shown in Eqn (4.48) holds.

$$p\,\mu_{L(T)}^{\overline{Li_2O}} = q_1\,\mu_{L(T)}^{\mathrm{Li}(1)} + q_2\,\mu_{L(T)}^{\mathrm{Li}(2)} + q_3\,\mu_{L(T)}^{\mathrm{Li}(3)} + \cdots + q_k\,\mu_{L(T)}^{\mathrm{Li}(k)} \tag{4.48}$$

Based on the above relationship between the chemical potentials, the appropriate standard-state chemical potential, $\mu_{L(T,0)}^{\mathrm{Li}(i)}$, may be chosen to satisfy Eqn (4.49):

$$n\,\mu_{L(T,0)}^{\overline{Li_2O}} = m_1\,\mu_{L(T,0)}^{\mathrm{Li}(1)} + m_2\,\mu_{L(T,0)}^{\mathrm{Li}(2)} + m_3\,\mu_{L(T,0)}^{\mathrm{Li}(3)} + \cdots + m_k\,\mu_{L(T,0)}^{\mathrm{Li}(k)}. \tag{4.49}$$

It should be noted that $\mu_{L(T,0)}^{\mathrm{Li(i)}}$ is represented in the molar free energy diagram by the intersection of the molar free energy curve of the Li(i) species in the liquid with the principal axis ($X_L^{\mathrm{Li}(i)} = 1$). The axis in this diagram is the same as that associated with the Li(i) species in the solid, where $X_S^{\mathrm{Li}(i)} = 1$, and $\mu_{L(T,0)}^{\mathrm{Li}(i)}$ does not necessarily have to equal to $\mu_{S(T,0)}^{\mathrm{Li}(i)}$. Combining Eqn (4.49) with Eqn (4.26) and taking into account the equivalency $\mu_{S(T,0)}^{\overline{Li_2O}} = \mu_{L(T,0)}^{\overline{Li_2O}}$ at the congruent point, Eqn (4.50) is obtained:

$$\begin{aligned} &m_1\left(\mu_{S(T,0)}^{\mathrm{Li}(1)} - \mu_{L(T,0)}^{\mathrm{Li}(1)}\right) + m_2\left(\mu_{S(T,0)}^{\mathrm{Li}(2)} - \mu_{L(T,0)}^{\mathrm{Li}(2)}\right) + m_3\left(\mu_{S(T,0)}^{\mathrm{Li}(3)} - \mu_{L(T,0)}^{\mathrm{Li}(3)}\right) + \cdots + m_k\left(\mu_{S(T,0)}^{\mathrm{Li}(k)} - \mu_{L(T,0)}^{\mathrm{Li}(k)}\right) \\ &= m_1\,\Delta\mu_{(T,0)}^{\mathrm{Li}(1)} + m_2\,\Delta\mu_{(T,0)}^{\mathrm{Li}(2)} + m_3\,\Delta\mu_{(T,0)}^{\mathrm{Li}(3)} + \cdots + m_k\,\Delta\mu_{(T,0)}^{\mathrm{Li}(k)} = 0, \end{aligned} \tag{4.50}$$

where $\Delta\mu_{(T,0)}^{\mathrm{Li}(i)} = \mu_{S(T,0)}^{\mathrm{Li}(i)} - \mu_{L(T,0)}^{\mathrm{Li}(i)}$. At this point, we can discuss the sign of $\Delta\mu_{(T,0)}^{\mathrm{Li}(i)}$, which determines whether or not the liquid can coexist with the solid. The molar free energy of Li(i) in the solid is actually represented by a point on the principal axis (Figure 4.10) [46]

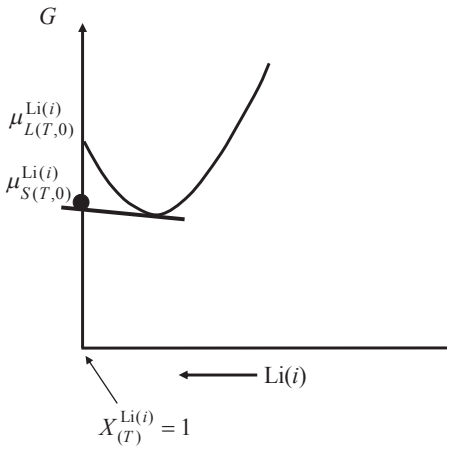


FIGURE 4.10 Relationship between the standard-state chemical potentials of the liquid, $\mu^{\mathrm{Li}(i)}_{L(T,0)}$, and solid, $\mu^{\mathrm{Li}(i)}_{S(T,0)}$, about Li-bearing chemical species, Li(i) [46]. Due to the limited solid solution range, the molar free energy of the solid Li(i) is represented by a dot on the principal axis corresponding to $\mu^{\mathrm{Li}(i)}_{S(T,0)}$. $\mu^{\mathrm{Li}(i)}_{L(T,0)}$ should be greater than or equal to $\mu^{\mathrm{Li}(i)}_{S(T,0)}$ in order to draw a common tangent between the molar free energies, G, of the liquid and solid. Note that both free energy curves have a common principal axis at $X^{\mathrm{Li}(i)}_{(T)} = 1$.

at which $\mu^{\mathrm{Li}(i)}_{S(T)} = \mu^{\mathrm{Li}(i)}_{S(T,0)}$, since $Li_2B_4O_7$ is a line compound, and, as discussed previously, the activity of any species in the solid, $a^{\mathrm{Li}(i)}_{S(T)}$, is unity at any temperature. When drawing a common tangent to the molar free energy curve between the solid and the liquid, Eqn (4.51) should always hold since the tangent point on the solid molar free energy curve always lies on the principal axis corresponding to $\mu^{\mathrm{Li}(i)}_{S(T,0)}$, while the tangent point on the liquid molar free energy curve moves away from the axis, such that

$$\mu^{\mathrm{Li}(i)}_{S(T,0)} \leq \mu^{\mathrm{Li}(i)}_{L(T,0)}. \tag{4.51}$$

This is illustrated in Figure 4.10 and leads to Eqn (4.52):

$$\Delta\mu^{\mathrm{Li}(i)}_{(T,0)} \leq 0. \tag{4.52}$$

Combining Eqn (4.50) with $m_i \geq 0$ and $\Delta\mu^{\mathrm{Li}(i)}_{(T,0)} \leq 0$, the value of the $\Delta\mu^{\mathrm{Li}(i)}_{(T,0)}$ term at the congruent point is zero, as in Eqn (4.53):

$$\Delta\mu^{\mathrm{Li}(i)}_{(T,0)} = 0. \tag{4.53}$$

At the liquidus temperature, T, the chemical potentials of all real chemical species, Li(i), will be identical in either the solid or the liquid, such that

$$\mu^{\mathrm{Li}(i)}_{S(T)} = \mu^{\mathrm{Li}(i)}_{L(T)}. \tag{4.54}$$

Then, since $a^{\mathrm{Li}(i)}_{S(T)} = 1$,

$$\ln a_{L(T)}^{\mathrm{Li}(i)} = \frac{\Delta\mu_{(T,0)}^{\mathrm{Li}(i)}}{RT}. \tag{4.55}$$

At the congruent point, $\Delta\mu_{(T,0)}^{\mathrm{Li}(i)} = 0$ (from Eqn (4.53)), and so

$$a_{L(T)}^{\mathrm{Li}(i)} = X_L^{\mathrm{Li}(i)} = 1. \tag{4.56}$$

Finally, the equilibrium partition coefficient of Li(i), $k_0^{\mathrm{Li}(i)}$, is given by

$$k_0^{\mathrm{Li}(i)} = \frac{X_S^{\mathrm{Li}(i)}}{X_L^{\mathrm{Li}(i)}} = \frac{1}{1} = 1. \tag{4.57}$$

The equilibrium partition coefficient, $k_0^{\mathrm{Li}(i)}$, for each of the Li-bearing species is found to be unity in the composition associated with a material that is simultaneously stoichiometric and congruent, such as $Li_2B_4O_7$ [46]. The partition coefficients of the B-bearing chemical species can be discussed in a similar manner, and all of these species can be shown to also have coefficients of unity at the congruent point. Therefore, no segregation of any constituent species, including ionic species, takes place and a c-EMF value of zero will be apparent.

In contrast, in the case of a material for which the stoichiometric composition is located somewhere other than the congruent point, such as holds true for $LiNbO_3$, the constituent species do not necessarily have partition coefficients equal to unity. Although the activity of the bulk liquid in equilibrium with the stoichiometric solid could be unity, i.e., $a_{L(T)}^{\overline{Li_2O}} = 1$, the principal axis ($X = 1$) cannot be common to the solid and the liquid since the equilibrium compositions are different between these two phases, such that $a_{S(T)}^{\overline{Li_2O}} = 1$ and $a_{L(T)}^{\overline{Li_2O}} = 1$ do not both hold true in the same coordination system. As a result, Eqns (4.48) to (4.57) cannot be used to analyze a stoichiometric crystal with a composition that differs from the congruent point. Thus, Eqn (4.53) is not valid, meaning that

$$\Delta\mu_{(T,0)}^{\mathrm{Li}(i)} \neq 0, \tag{4.58}$$

and

$$\ln\left(\frac{a_{L(T)}^{\mathrm{Li}(i)}}{a_{S(T)}^{\mathrm{Li}(i)}}\right) = \frac{\Delta\mu_{(T,0)}^{\mathrm{Li}(i)}}{RT} \neq 0.$$

Therefore, Eqn (4.59), below, can be written

$$\frac{a_{L(T)}^{\mathrm{Li}(i)}}{a_{S(T)}^{\mathrm{Li}(i)}} = a_{L(T)}^{\mathrm{Li}(i)} \neq 1. \tag{4.59}$$

The partition coefficients, $k_0^{\mathrm{Li}(i)}$, for the constituent melt species of a stoichiometric crystal that is not congruent will not equal unity and so segregation will occur, which in turn leads to a nonzero value of c-EMF. This is demonstrated on an experimental basis in Figure 4.8.

4.4 Oxide Crystals Having a Stoichiometric Composition Coincident with the Congruent Point

As discussed in the previous section, when a compound has a stoichiometric structure at the congruent point, the equilibrium partitioning coefficients of all chemical species become unity and segregation does not occur. With regard to conventional c-LN or c-LT, the segregation of ionic species is observed during growth because these materials are not stoichiometric. A unique relationship between the unity of the partitioning coefficients of ionic species and the coincidence of stoichiometry and congruency is quite beneficial in terms of developing LN or LT crystals that are superior to conventional LN and LT. The congruent compositions are found on the stoichiometric line of 50 mol% Nb_2O_5 (line *A* in Figure 4.3) in the Li_2O–Nb_2O_5–MgO system and on the stoichiometric line of 50 mol% Ta_2O_5 (line *B* in Figure 4.11) in the Li_2O-Ta_2O_5-MgO system. Using Kröger–Vink notation, these materials may be written as $(Li_{Li}^{\times})_{0.906}(Mg_{Li}^{\cdot})_{0.047}(V_{Li}^{\prime})_{0.047}(Nb_{Nb}^{\times})(O_o^{\times})_3$, termed cs-MgO:LN [47], and $(Li_{Li}^{\times})_{0.816}(Mg_{Li}^{\cdot})_{0.092}(V_{Li}^{\prime})_{0.092}(Ta_{Ta}^{\times})(O_o^{\times})_3$, termed cs-MgO:LT [56], respectively. These compounds show no segregation of ionic species during growth, which demonstrates that the equilibrium partition coefficients, k_0, are unity for all constituent chemical species in the melt, and thus cs-MgO:LN and cs-MgO:LT are congruent and stoichiometric simultaneously. Therefore, the conventional congruent materials $LiNbO_3$ and $LiTaO_3$, both of which are nonstoichiometric, are converted to congruent,

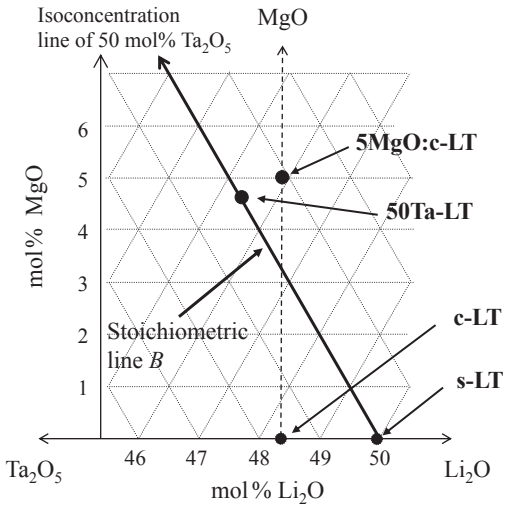


FIGURE 4.11 Composition diagram of various $LiTaO_3$ compositions in the pseudo-ternary Li_2O-Ta_2O_5-MgO system. The isoconcentration line of 50 mol% Ta_2O_5 is termed line *B* and represents a ratio of O to Ta of exactly 3.0. Legend: s-LT = stoichiometric $LiTaO_3$ on the Li_2O-Ta_2O_5 line; c-LT = congruent $LiTaO_3$ on the Li_2O-Ta_2O_5 line; 50Ta-LT = MgO-doped $LiTaO_3$ on line *B*; 5MgO:LT = c-LT doped with 5 mol% MgO.

stoichiometric crystals based on our extended concept of stoichiometry. Both materials, however, contain impurities and vacancies that are not allowed under the conventional definition of stoichiometry.

4.4.1 MgO-Doped $LiNbO_3$

In a conventional stoichiometric crystal, each element occupies only its own unique site in the crystal lattice, and no defects arising from nonstoichiometry are allowed. Thus, a conventional stoichiometric crystal usually shows good optical properties. However, the stoichiometric composition often does not coincide with the congruent composition. In such a case, growth of the stoichiometric compound is difficult due to the compositional deviation caused by the segregation of constituent species in the melt during growth. $LiNbO_3$ (LN) is one such compound. As shown in Figure 4.2(B) [17], the stoichiometric composition (s-LN; $Li_2O = 50$ mol%) does not coincide with the congruent composition (c-LN; $Li_2O = 48.38$ mol% [57]). Although it can easily be grown from the melt, the optical properties of c-LN, such as conversion efficiency of SHG, photoconductivity, and transparency in the short-wavelength region, are inferior to those of s-LN [2], and so impurity doping of c-LN is performed to improve its optical properties. MgO is considered to be an effective impurity for LN, because MgO-doped s-LN exhibits high photoconductivity and has a high threshold for optical damage [2]. However, these crystals are no longer congruent nor are they stoichiometric. In the following section, the growth and characterization of cs-MgO:LN, a crystal that is simultaneously congruent and stoichiometric, is described.

First, the distribution of the melting temperatures of sintered materials with various compositions was determined for the Li_2O–Nb_2O_5–MgO ternary system via DTA. The composition possessing the highest melting temperature is the congruent point of the MgO-doped LN. Next, this composition was confirmed to be the exact ternary congruent composition by confirming that c-EMF was not generated. Such a composition is found on the 50 mol% Nb_2O_5 stoichiometric line, where the congruent point meets the stoichiometric composition. The compound cs-MgO:LN (Li_2O:Nb_2O_5:MgO = 45.3:50:4.7) is such a compound and is expected to be easily grown and to exhibit superior optical properties compared to those of conventional c-LN, MgO-doped c-LN, and s-LN.

4.4.1.1 Distribution of the Melting Temperatures of MgO-Doped $LiNbO_3$

The ternary phase diagram for the MgO-Li_2O–Nb_2O_5 system is illustrated in Figure 4.12. The isoconcentration line for 50 mol% Nb_2O_5 is also drawn as a stoichiometric line on this diagram (line *A*). Every LN crystal on this line (50Nb-LN) could potentially be stoichiometric. The melting temperatures of sintered versions of this material with various compositions were measured by high-temperature DTA around line *A*. The distribution of melting points is plotted in Figure 4.12, which represents the solidus surface of the Mg-doped LN in this region. The congruent composition, which

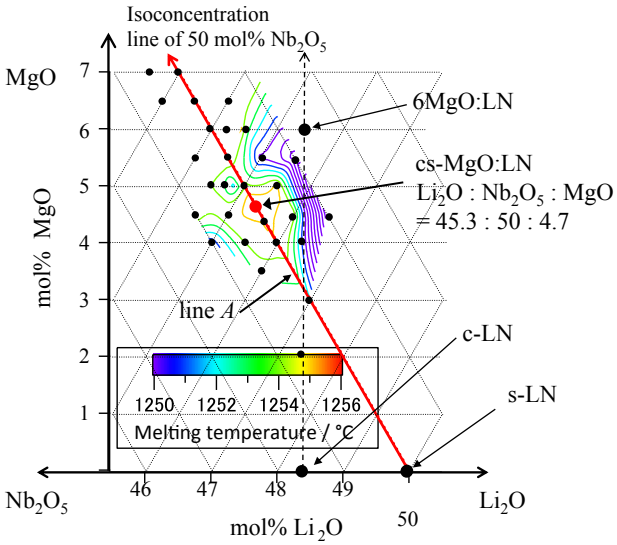


FIGURE 4.12 Distribution of the melting temperatures of sintered compounds with various compositions around the isoconcentration line of 50 mol% Nb_2O_5 (line *A*) in the pseudo-ternary system of Li_2O–Nb_2O_5–MgO. Isothermal contours showing the solidus plane around cs-MgO:LN are also included. The highest melting temperature corresponding to the congruent point is found on line *A*, and this represents the cs-MgO:LN composition at which the congruent point is coincident with the stoichiometric composition: Li_2O:Nb_2O_5:MgO = 45.3:50:4.7. *After Ref. [58].*

corresponds to the highest melting point, was found at the composition Li_2O:Nb_2O_5:MgO = 45.3:50:4.7 on line *A* [58]. This represents a new LN composition, termed cs-MgO:LN, equal to the congruent point coincident with the stoichiometric composition. Although several studies concerning Mg-doped LN have been reported [59–61], none of these has discussed this unique composition. In the pseudo-binary system of Li_2O–Nb_2O_5, c-LN does not coincide with s-LN, and the ionic species segregate during growth. Although 5MgO:LN exhibits improved optical properties [62], it is neither congruent nor stoichiometric, and thus segregation of bulk components as well as ionic species occurs during growth, yielding an inhomogeneous compositional distribution in the crystal.

4.4.1.2 Crystallization Electromotive Force

The c-EMF measurement methods are described in detail in Section 4.3.1. The c-EMF is represented as a hysteresis gap on the potential curve in Figure 4.7 during the melting and solidification processes and has a magnitude of several mV. The value of c-EMF is also growth-rate dependent and not specific to the material itself. Furthermore, because a high temperature gradient is present in the melt near the interface in a μ-PD measurement system, an intrinsic interface electric field is generated due to a thermoelectric process (via the Seebeck effect), and this in turn influences the distribution of ionic species in the solute boundary layer. Thus, an extrapolation of c-EMF toward the zero temperature gradient is drawn in Figure 4.13 for the compositions c-LN, s-LN, c-LN doped with 6 mol% MgO (6MgO:LN), and cs-MgO:LN. In the case of each of these compositions, the amplitude of c-EMF ($\Delta\phi_{EMF}$) decreases with decreasing temperature gradient. The exact value of $\Delta\phi_{EMF}$ can be obtained at 0 °C/cm where no electric field

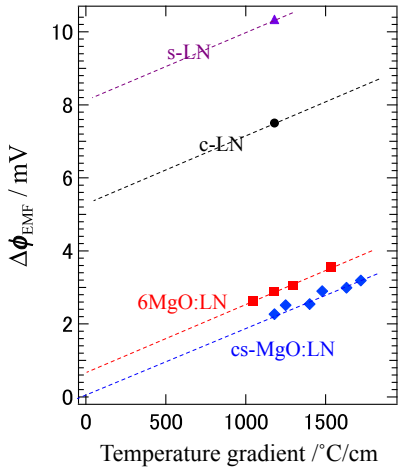


FIGURE 4.13 Temperature gradient dependence of $\Delta\phi_{EMF}$ for s-LN, c-LN, 6MgO:LN, and cs-MgO:LN. For each composition, $\Delta\phi_{EMF}$ decreases as the temperature gradient decreases. The $\Delta\phi_{EMF}$ value of cs-MgO:LN becomes 0 mV at 0 °C/cm, while the values of the other materials become nonzero. *Modified from Ref. [47].*

exists, and cs-MgO:LN is the only composition that attains zero c-EMF at 0 °C/cm [47]. These data also confirm that no segregation occurs during growth at any growth rate, demonstrating that the partition coefficients of all the solute components, including the ionic species, are unity and thus assuring the true congruency of cs-MgO:LN [47]. In the previous section, it was shown on a thermodynamic basis that partition coefficients equal to unity of the solute components leads to activity values equal to unity in both the solid and liquid states, and this in turn results in cs-MgO:LN simultaneously having a stoichiometric structure. The converse is also true—based on thermodynamic arguments, in a compound that has a stoichiometric structure at the congruent composition, the equilibrium partition coefficient of all chemical species, including ionic species, becomes unity. In other words, the activities of all constituent elements both in the melt and the solid will be unity in the case of a compound that exhibits the simultaneous occurrence of stoichiometry and congruency, as described in Section 4.3.2.

In contrast, the conventional congruent material $LiNbO_3$ (c-LN) requires the partition coefficient to equal unity only for the bulk $\overline{Li_2O}$ and $\overline{Nb_2O_5}$ components, and its crystallization EMF has a nonzero value, suggesting that the ionic species in the c-LN melt have nonunity partition coefficients. The values obtained for activity, a^j ($j = O$, Nb, Li, Mg, and vacancy) and equilibrium partition coefficient, k_0^j, are summarized in Table 4.1 for the growth of s-LN (stoichiometric), c-LN (congruent), and cs-MgO:LN (stoichiometric and congruent) crystals.

4.4.1.3 Bulk Crystal Growth of cs-MgO:$LiNbO_3$ and Its Nonlinear Optical Characterization

Only cs-MgO:LN has no segregation of ionic species [47] and therefore is truly congruent, easy to grow, and is expected to have a higher compositional homogeneity

Table 4.1 Activity, a^j, and Equilibrium Partition Coefficient, k_0^j, Crystallization Electromotive Force (c-EMF), $\Delta\phi_{EMF}$, Values Associated with the Growth of s-LN (Stoichiometric), c-LN (Congruent) and cs-MgO:LN (Stoichiometric and Congruent) Crystals

	Activity, a^j (j = Li, Nb, Mg, O, Vacancy)	**Equilibrium Partition Coefficient, k_0^j (j = Li, Nb, Mg, O, Vacancy)**	**c-EMF, $\Delta\phi_{EMF}$**
s-LN	$a_S^j = 1,\ a_L^j \neq 1$	$k_0^j \neq 1$	$\Delta\phi_{EMF} \neq 0$
c-LN	$a_S^j \neq 1,\ a_L^j \neq 1$	$k_0^j \neq 1$*	$\Delta\phi_{EMF} \neq 0$
cs-MgO:LN	$a_S^j = 1,\ a_L^j = 1$	$k_0^j = 1$	$\Delta\phi_{EMF} = 0$

*k_0^j is unity for the bulk component $\overline{Li_2O}$ and $\overline{Nb_2O_5}$ in the pseudo-binary system of Li_2O–Nb_2O_5.

than the conventional crystals of c-LN and MgO-doped c-LN. It is also expected to exhibit a SHG conversion efficiency as high as that of s-LN because of its stoichiometric structure. Bulk crystals of cs-MgO:LN were therefore grown, and its superiority in both compositional homogeneity and nonlinear optical properties were demonstrated by comparing its SHG properties with those of c-LN, s-LN, and 5MgO:LN.

A bulk single crystal of cs-MgO:LN was grown at a rate of 2 mm/h via the Czochralski method along the Z-axis in an air atmosphere. The resulting crystal was colorless and inclusion-free, as shown in Figure 4.14 [58]. The crystal was about 22 mm in diameter and 40 mm long, and its solidified melt fraction, g, was 0.25 after the completion of growth. The Czochralski technique was also used to grow c-LN, 5MgO:LN, and s-LN,

FIGURE 4.14 A cs-MgO:LN single bulk crystal grown along the Z-axis by the Czochralski method [58].

although the s-LN was produced using the double-crucible method [63]. The SHG properties of these bulk single crystals were measured. In preparation, all crystals were poled and subsequently sliced into $5 \times 5 \times 5$ mm blocks (the solidified melt fraction value was $g = 0.13$ in the case of cs-MgO:LN) along the $x[1\bar{2}10]$, $y[10\bar{1}0]$, $z[0001]$ directions and the x-planes were polished to a mirror finish to form the incident plane. An optical parametric oscillator (OPO) pumped by the third harmonic wave of a Q-switched Nd:YAG laser (1064 nm) was used to evaluate the SHG properties associated with noncritical phase matching, using d_{31} of the LN crystal in the infrared region from 800 to 1200 nm.

Since the phase-matching wavelength is sensitive to the crystal composition [64], its distribution in a crystal test plate was used to evaluate the compositional homogeneity of the crystal. Figure 4.15 shows the in-plane distribution of the noncritical phase-matching wavelength for a 4×4 mm area on each crystal. It should be noted that the vertical direction in each test plate in Figure 4.15 was parallel to the growth axis, while the horizontal direction was parallel to the radial direction. The phase-matching wavelength of the cs-MgO:LN was almost constant over the whole test plate ($\Delta g = 0.024 \sim 0.025$) [58]. This homogeneity was surprisingly sustained during growth even with possible variations in the growth rate due to the changing crystal diameter,

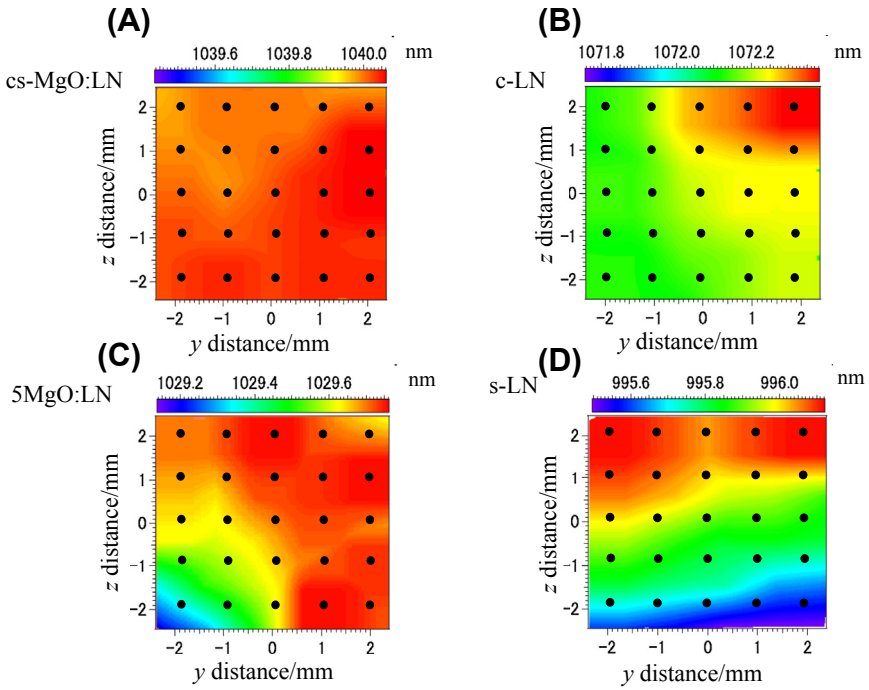


FIGURE 4.15 The in-plane distribution (4×4 mm) of the noncritical phase-matching wavelength [58]. Solid circles indicate the measured points: (A) cs-MgO:LN, (B) c-LN, (C) 5MgO:LN, and (D) s-LN.

which was attributed to the complete unity of the partition coefficients of all the melt species. To confirm the compositional homogeneity over the entire length of the cs-MgO:LN crystal, its Curie temperature was measured via differential scanning calorimetry (DSC) at the top and the base. The Curie temperature difference between the top and bottom portions was 0.8 °C, which corresponds to a 0.02 mol% Li_2O variation [65] when neglecting the effect of MgO incorporation. This variation was sufficiently small as to demonstrate the congruency of cs-MgO:LN. The c-LN crystal was slightly less homogeneous than the cs-MgO:LN, but was much more homogeneous than both the s-LN and 5MgO:LN materials. These results are consistent with the partitioning behavior of the solute components, since both c-LN and cs-MgO:LN have congruent-melting crystals and bulk components, such as Li_2O, Nb_2O_5, and/or MgO, which are not segregated but rather partitioned into the crystal with partition coefficients of unity. In the case of cs-MgO:LN, neither the bulk components nor the ionic species are segregated, so its composition would be expected to be more homogeneous than that of c-LN. In contrast, the phase-matching wavelengths of s-LN and 5MgO:LN varied drastically within the test plate since these compounds were not congruent. The maximum compositional deviations of s-LN and c-LN within the measured areas were estimated to be 0.0092 mol% Li_2O and 0.0042 mol% Li_2O, respectively. The observed variation in the phase-matching wavelength in s-LN might reflect the compositional variation along the growth axis during growth. This indicates the difficultly of growing homogeneous s-LN using the double-crucible method [63].

The SHG conversion efficiency was measured at the center of each test plate (*z*: 0 mm, *y*: 0 mm in Figure 4.15) at a constant value of the fundamental beam power. As shown in Figure 4.16, the conversion efficiency of cs-MgO:LN is slightly lower than those obtained with 5MgO:LN and s-LN [58], however, these values are almost equal in magnitude, while the value measured for the c-LN sample is considerably lower. These experiments demonstrated that cs-MgO:LN has been generated, based on our extended concept of stoichiometry, and that this material, due to its congruent state, is superior to

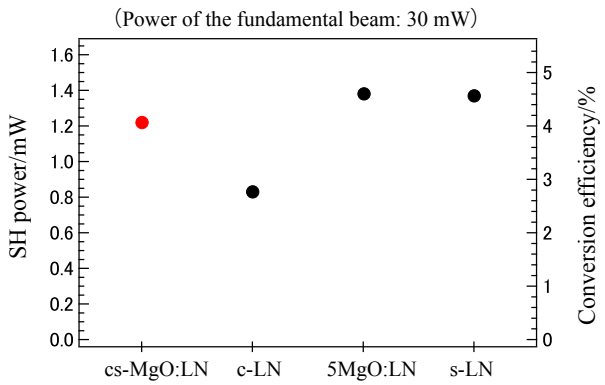


FIGURE 4.16 Plots of the secondary harmonic generation conversion efficiency obtained from cs-MgO:LN, c-LN, 5MgO:LN, and s-LN.

conventional $LiNbO_3$ crystals having a highly homogeneous composition. This substance also exhibits a high conversion efficiency of secondary harmonic generation due to its stoichiometric structure.

4.4.2 MgO-Doped $LiTaO_3$

Lithium tantalate (LT), like lithium niobate (LN), is an oxide material whose congruent composition differs from its stoichiometric composition in the binary Li_2O–Ta_2O_5 system. By employing our extended concept of stoichiometry, a new lithium tantalate incorporating MgO doping, cs-MgO:LT (Li_2O:Ta_2O_5:MgO = 40.8:50.0:9.2), has been developed [56]. As in the case of cs-MgO:LN, this material is stoichiometric and congruent. Because of this coincidence, cs-MgO:LT does not exhibit any c-EMF and does not show any segregation during growth, even in the case of ionic species. A bulk crystal was grown from a cs-MgO:LT melt via the Czochralski method and showed excellent compositional homogeneity, as demonstrated by the constant distribution of the Curie temperature throughout the crystal.

4.4.2.1 Distribution of Melting Temperatures of MgO-Doped $LiTaO_3$

The ternary phase diagram for the system of MgO–Li_2O–Ta_2O_5 is illustrated in Figure 4.17. The isoconcentration line of 50 mol% Ta_2O_5 is also drawn as a stoichiometric line (line *B*). Every LT crystal on this line (50Ta-LT) could potentially be stoichiometric. The melting temperatures of sintered materials with various compositions were measured by high-temperature DTA around line *B*. The distribution of melting points is

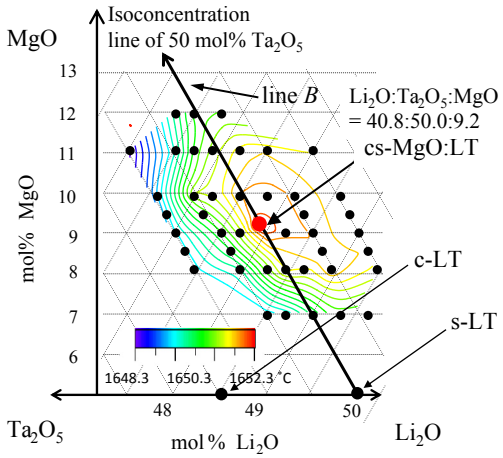


FIGURE 4.17 Distribution of the melting temperatures of sintered compounds with various compositions around the isoconcentration line of 50 mol% Ta_2O_5 (line *B*) in the pseudo-ternary system of Li_2O-Ta_2O_5-MgO. Isothermal contours showing the solidus plane around cs-MgO:LT are also drawn. The highest melting temperature corresponding to the congruent point is found on line *B*. This is the cs-MgO:LT composition at which the congruent point is coincident with the stoichiometric composition: Li_2O:Ta_2O_5:MgO = 40.8:50.0:9.2. *After Ref. [56].*

plotted in Figure 4.17, which represents the solidus surface of the Mg-doped LT in this region. The congruent composition, which corresponds to the highest melting point, was found at the composition corresponding to Li_2O:Ta_2O_5:MgO = 40.8:50.0:9.2 on line *B* [56]. This represents a new LT composition, cs-MgO:LT, that exists at the congruent point coincident with the stoichiometric composition.

Since cs-MgO:LT has a composition that lies well within the solid solution range in the Li_2O-Ta_2O_5-MgO ternary system, it can be grown without any compositional constraints, similar to the case of cs-MgO:LN. In contrast, conventional s-LT has a composition located near the solubility limit of Li_2O so that the practical upper limit for the Li_2O content in this material is 49.8 mol% [66,67]. Thus cs-MgO:LT is more tolerant than s-LT from a compositional perspective, which offers a significant advantage with regard to crystal growth.

4.4.2.2 Crystallization Electromotive Force

Measurement of c-EMF was carried out for cs-MgO:LT and for conventional c-LT using the same μ-PD technique as was applied in the case of LN. The furnace apparatus used during these measurements was composed of Pt containing 20% Rh rather than pure Pt because of the high melting temperature of LT. Growth via the μ-PD technique is accompanied by a large temperature gradient near the interface that leads to an intrinsic electric field due to the Seebeck effect, as is observed during LN growth. An intrinsic electric field such as this significantly affects the segregation of ionic species at the interface [40,68], and thus the $\Delta\phi_{EMF}$ values obtained for cs-MgO:LT and c-LT were corrected by extrapolation to the temperature gradient at 0 °C/cm. Following this correction, it was found that the $\Delta\phi_{EMF}$ of c-LT was nonzero, while that of cs-MgO:LT was zero [56]. These results demonstrate experimental verification that the partition coefficient, k_0, of the ionic species in the cs-MgO:LT melt is unity. Consequently, the activity of all constituent elements both in the melt and in the solid state of cs-MgO:LT is unity, meaning that cs-MgO:LT is both stoichiometric and congruent.

4.4.2.3 Bulk Crystal Growth and Curie Temperature of cs-MgO:$LiTaO_3$

A bulk single crystal was grown along the Z-axis from the cs-MgO:LT melt under an atmosphere composed of 99 vol% Ar/1 vol% O_2 via the CZ method. The resulting crystal is shown in Figure 4.18 [56]. Since the cs-MgO:LT composition is congruent, a homogeneous compositional distribution was expected in this material. The Curie temperatures of this crystal were measured via DSC at the top, middle, and bottom of the sample along its central axis. Figure 4.19 summarizes the distribution of the Curie temperature over the cs-MgO:LT crystal as well as the temperature of the residual melt. The maximum difference in Curie temperature over the cs-MgO:LT crystal was 0.4 °C [56], corresponding to a 0.03 mol% Li_2O difference [69] when neglecting the effect of MgO doping. This result demonstrates the superior homogeneity of the cs-MgO:LT crystal and shows that the cs-MgO:LT is congruent. It should be noted that growth from the melt of

FIGURE 4.18 A cs-MgO:LT single bulk crystal grown along the Z-axis by the Czochralski method [56].

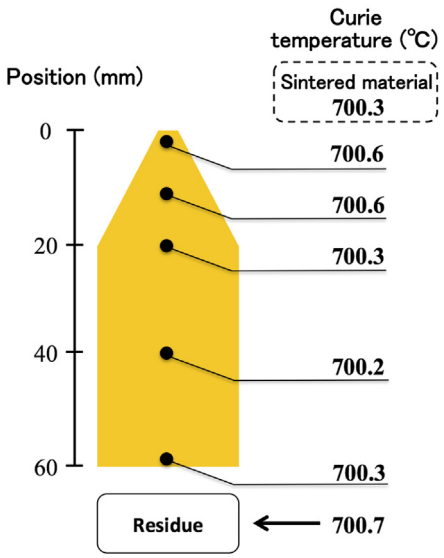


FIGURE 4.19 Distribution of the Curie temperatures of as-grown cs-MgO:LT along the growth axis, the seed crystal, and the residual melt [56].

the cs-MgO:LT composition does not accompany segregation of any ionic melt species. This material thus represents a true congruent state, whereas conventional c-LT in the Li_2O–Ta_2O_5 binary system does not since its $\Delta\phi_{EMF}$value is not zero, meaning that segregation of ionic species takes place. Therefore, cs-MgO:LT contains essentially no uncoupled ionic species in the crystal and is expected to have superior nonlinear optical properties as compared to conventional LT crystals.

4.4.3 Thermodynamic Requirements for Impurity Doping

MgO is a suitable doping oxide for the preparation of high quality LN and LT crystals exhibiting the concurrent occurrence of stoichiometry and congruency in the ternary system. It is natural to inquire as to whether any other oxides produce a similar effect when added to LN or LT such that the resulting crystals become simultaneously stoichiometric and congruent and whether there are any specific requirements for an oxide to behave as an effective dopant. Figure 4.20 illustrates the ternary Li_2O–Nb_2O_5–ZnO diagram, in which the distribution of melting points is drawn around the stoichiometric line of 50 mol% Nb_2O_5 (line *C*). The highest melting point that corresponds to the congruent point does not lie on the stoichiometric line *C*, and thus it is not possible to make LN simultaneously congruent and stoichiometric by adding ZnO. We may then ask ourselves, what is the difference between MgO and ZnO that renders MgO effective but ZnO ineffective?

In order to develop an LN or LT composition, the crystals must have a zero c-EMF value. In other words, the activity of every constituent element in both the solid and

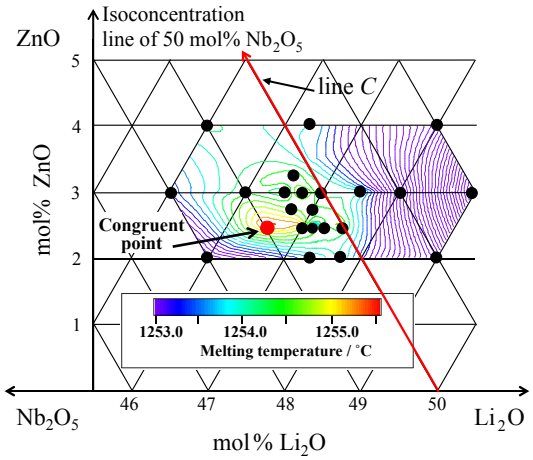


FIGURE 4.20 Distribution of the melting temperatures of sintered compounds with various compositions around the isoconcentration line of 50 mol% Nb_2O_5 (line *C*) in the pseudo-ternary Li_2O–Nb_2O_5–ZnO system. Isothermal contours showing the solidus plane around line *C* are also drawn. The highest melting temperature corresponding to the congruent point is located apart from line *C*.

liquid states is required to be unity. Suppose that a divalent oxide, AO, is doped into LN and subsequently ionized in the melt according to the following equation:

$$\mathrm{AO} \rightleftarrows \mathrm{A}^{2+} + \mathrm{O}^{2-} \tag{4.60}$$

The congruent point is assumed to form at a certain composition in the ternary diagram of Li_2O–Nb_2O_5–AO. At the congruent melt composition, the partition coefficients of the three bulk components will be unity, such that

$$k_0^{\overline{\mathrm{AO}}} = k_0^{\overline{\mathrm{Li_2O}}} = k_0^{\overline{\mathrm{Nb_2O_5}}} = 1, \tag{4.61}$$

where the bar over a species indicates a bulk component. Accordingly, the activities of these three bulk components in both the solid and liquid are unity, meaning that

$$a_\beta^{\overline{\mathrm{AO}}} = a_\beta^{\overline{\mathrm{Li_2O}}} = a_\beta^{\overline{\mathrm{Nb_2O_5}}} = 1, \tag{4.62}$$

where β indicates solid (S) or liquid (L) phases. Because the value of c-EMF is zero, one degree of freedom is available to establish the appropriate relationship between the standard-state chemical potentials of AO, A^{2+}, and O^{2-} in both the solid and liquid, as follows:

$$a_\beta^{\overline{\mathrm{AO}}} = a_\beta^{\mathrm{A}^{2+}} a_\beta^{\mathrm{O}^{2-}} = 1 \tag{4.63}$$

Likewise,

$$a_\beta^{\overline{\mathrm{Li_2O}}} = \left(a_\beta^{\mathrm{Li}^+}\right)^2 a_\beta^{\mathrm{O}^{2-}} = 1, \tag{4.64}$$

and

$$a_\beta^{\overline{\mathrm{Nb_2O_5}}} = \left(a_\beta^{\mathrm{Nb}^{5+}}\right)^2 \left(a_\beta^{\mathrm{O}^{2-}}\right)^5 = 1. \tag{4.65}$$

At this point, we also wish to determine the thermodynamic criteria that produce activities of unity for the species A^{2+}, O^{2-}, Li^+, and Nb^{5+}. If AO is completely ionized, the partitioned AO in the solid is attributed only to A^{2+} in the melt. Thus,

$$k_0^{\mathrm{A}^{2+}} = 1, \tag{4.66}$$

and so

$$a_\beta^{\mathrm{A}^{2+}} = 1. \tag{4.67}$$

Inserting Eqn (4.67) into Eqn (4.63) leads to $a_\beta^{\mathrm{O}^{2-}} = 1$, and combining this with Eqns (4.64) and (4.65) gives

$$a_\beta^{\mathrm{Li}+} = a_\beta^{\mathrm{Nb}^{5+}} = 1. \tag{4.68}$$

Thus, the activity of every constituent element in the AO-doped LN is unity in both the liquid and solid states. This composition is found on the stoichiometric line of 50 mol% Nb_2O_5 and can be termed cs-AO:LN, since it is both congruent and stoichiometric. This scenario occurs in the case of MgO doping, which is understandable when one considers that MgO is an almost completely ionic oxide. In contrast, if AO is only partially ionized, both AO and A^{2+} will be partitioned into the crystal. As a result, even

if $a_\beta^{\overline{AO}} = 1$ (Eqn (4.63)) is true, neither $a_\beta^{A^{2+}}$ nor $a_\beta^{O^{2-}}$ will necessarily equal unity, and so the congruent point will not lie on the stoichiometric line for this material. In general, then, a dopant that is only partially ionized in the melt will not generate a compound LN in which stoichiometry and congruency are coincident. In contrast, when the oxide dopant AO is completely ionized, the congruent point can appear on the stoichiometric line.

4.5 Summary

An extended concept of a *stoichiometric* oxide has been developed. The chemical potential of this new stoichiometric compound has no mixing term, and thus the activities of all its constituent elements will equal unity. The opposite is equally true—when the activities of all the constituent elements can be unity, the material is stoichiometric. In this definition, the term *constituent element* is expanded to potentially include both impurities and vacancies, neither of which is allowed in the conventional definition of stoichiometry. An activity of unity for a species is possible only when one degree of freedom is available at the crystal site, so as to assign the chemical potential of each species, j, to the standard-state chemical potential, such that $\mu_0^j = \mu^j$. Based on this extended understanding of stoichiometry, new $LiNbO_3$ and $LiTaO_3$ crystals have been developed by doping with MgO. These materials are simultaneously stoichiometric and congruent and are therefore easy to grow and exhibit excellent nonlinear optical properties. Thermodynamic arguments demonstrate that the concurrent occurrence of stoichiometry and congruency requires activity values of unity for all the constituent elements of the substance, not only in the solid but also in the liquid state at equilibrium. This means that the partition coefficients of all the chemical species in the melt must also be unity such that no segregation occurs; this has been demonstrated experimentally by the observation of zero c-EMF. It is very important that the stoichiometry of an oxide crystal and its relevant point defects are discussed in association with solute partitioning during crystal growth. This is because the point defects are partitioned with other elements from the liquid to the solid state as the crystal is formed, and the associated partitioning is related to the activity of the constituent elements in terms of their respective equilibrium partition coefficients, which in turn determines the stoichiometry or nonstoichiometry of the material.

References

[1] Dalton J. A new system of chemical philosophy part 1. Manchester: S. Russell; 1808.

[2] Furukawa Y, Kitamura K, Takekawa S. Opt Lett 1998;23:1892.

[3] Nakamura M, Takekawa S, Terabe K, Kitamura K, Usami T, Nakamura K, et al. Ferroelectrics 2002; 273:199.

[4] Bube R. Electrons in solids: an introductory survey. Academic Press; 1988.

[5] Sze SM. Semiconductor devices: physics and technology. 3rd ed. John Wiley & Sons; 2012.

[6] Oda O. Compound semiconductor bulk materials and characterizations. World Scientific Pub Co Inc; 2003.

[7] Rudolph P, Jurisch M. J Cryst Growth 1999;198–199:325.

[8] Dhar A, Mansingh A. J Appl Phys 1990;68:5804.

[9] Pauling L. J Am Chem Soc 1929;51:1010.

[10] Brown ID, Shannon RD. Acta Cryst A 1972;29:266.

[11] Brown ID. Acta Cryst B 1992;48:553.

[12] Brown ID. The chemical bond in inorganic chemistry. The bond valence model. New York: Oxford Univ. Press; 2002.

[13] Andersson J, Collen B, Kuylenstierna U, Magnéli A. Acta Chem Scand 1957;11:1641.

[14] Magnéli A. Pure Appl Chem 1978;50:1261.

[15] Tilley RJD. Defect crystal chemistry and its applications. Blackie & Son Limited, Glasgow and London. Published in USA by Chapman and Hall, New York; 1987.

[16] Stokołsa A, Laskowska B. J Chem Crystallogr 2008;38:913.

[17] Svaasand LO, Eriksrud M, Nakken G, Grande AP. J Cryst Growth 1974;22:179.

[18] Rogacheva EL, Gorne GV, Zhigareva NK, Ivanova AB. Inorg Mater 1991;27:194.

[19] Collongues R. Jpn J Appl Phys 1993;32:442.

[20] Roth WL. Acta Cryst 1960;13:146.

[21] Koch F, Cohen JB. Acta Cryst B 1969;25:275.

[22] Nakamura K, Saishoji T, Kubota T, Iida T, Shimanuki Y, Kotooka T, et al. J Cryst Growth 1997;180:61.

[23] Voronkov VV, Falster R. J Cryst Growth 1998;194:76.

[24] Falster R, Voronkov VV. Mater Sci Eng B 2000;73:87.

[25] Abe T, Takahashi T. J Cryst Growth 2011;334:16.

[26] Voronkov VV, Falster R. J Cryst Growth 2012;351:115.

[27] Catlow CRA, Fender BEF. J Phys C 1975;8:3267.

[28] Catlow CRA, Fender BEF, Muxworthy DG. J Phys (Paris) C 1977;7:67.

[29] Ye ZG, Dong M. J Appl Phys 2000;87:2312.

[30] Xu GS, Luo HS, Xu HQ, Yin ZW. Phys Rev B 2001;64:020102.

[31] Yamashita Y, Hosono Y, Harada K, Yasuda N. IEEE Trans Ultrason Ferroelectr Freq Control 2002;49:184.

[32] Gotalskaya N, Drezin DI, Bezdelkin VV, Stassevich VN. Proc IEEE Freq Symp 1993:339.

[33] Uda S, Buzanov OA. J Cryst Growth 2000;211:318.

[34] Kröger FA. The chemistry of imperfect crystals. 2nd rev. Amsterdam: North-Holland Publishing Company; 1974.

[35] ToftSørensen O, editor. Nonstoichiometric oxides. New York: Academic Press Inc.; 1981.

[36] Agullo-Lopez F, Catlow CRA, Townsend PD. Point defects in materials. London: Academic Press Limited; 1988.

[37] Gellings PJ, Bouwmeester HJM, editors. The CRC handbook of solid state chemistry. London: CRC Press; 1997.

[38] Innocenti A, editor. Stoichiometry and materials science – when numbers matter. In Tech; 2012.

[39] Uda S, Tiller WA. J Cryst Growth 1992;121:155.

[40] Uda S, Tiller WA. J Cryst Growth 1992;121:93.
[41] Uda S, Kon J, Ichikawa J, Inaba K, Shimamura K, Fukuda T. J Cryst Growth 1997;179:567.
[42] Uda S, Kon J, Shimamura K, Ichikawa J, Inaba K, Fukuda T. J Cryst Growth 1997;182:403.
[43] Uda S, Tsubota T. J Cryst Growth 2010;312:3650.
[44] Boikov YA, Goltsman BM, Yarmarkin VK, Lemanov VV. Appl Phys Lett 2001;78:3866.
[45] Surowiak Z, Margolin AM, Zakharchenko IN, Biryukov SV. Thin Solid Films 1989;176:227.
[46] Uda S. J Cryst Growth 2008;310:3864.
[47] Kimura H, Uda S. J Cryst Growth 2009;311:4094.
[48] Uda S, Komatsu R, Takayama K. J Cryst Growth 1997;171:458.
[49] Uda S, Koh S, Huang X. J Cryst Growth 2006;292:1.
[50] Azuma Y, Uda S. J Cryst Growth 2007;306:217.
[51] Koh S, Uda S, Nishida M, Huang X. J Cryst Growth 2006;297:247.
[52] Koh S, Uda S, Huang X. J Cryst Growth 2007;306:406.
[53] Kimura H, Koizumi H, Uchida T, Uda S. J Cryst Growth 2009;311:1553.
[54] D'yakov VA, Shumov DP, Rashkovich LN, Aleksandrovskii AL. Bull Acad Sci USSR, Phys Ser 1985;49:117.
[55] Aleksandrovskii AL, Shumov DP. Cryst Res Technol 1990;25:1239.
[56] Fujii S, Uda S, Maeda K, Nozawa J, Koizumi H, Fujiwara K, et al. J Cryst Growth 2013;383:63.
[57] Bourdi PF, Norwood RG, Bird CD, Calvert GD. J Cryst Growth 1991;113:61.
[58] Kimura H, Taniuchi T, Iida S, Uda S. J Cryst Growth 2010;312:3425.
[59] Grabmaier BG, Otto F. J Cryst Growth 1986;79:682.
[60] Zhou YF, Wang JC, Wang PL, Tang LA, Zhu QB, Wu YA, et al. J Cryst Growth 1991;114:87.
[61] Dakki A, Ferriol M, Cohen-Adad MT. Eur J Solid State Inorg Chem 1996;33:19.
[62] Bryan DA, Gerson R, Tomaschke HE. Appl Phys Lett 1984;44:847.
[63] Kitamura K, Yamamoto JK, Iyi N, Kimura S. J Cryst Growth 1992;116:327.
[64] Jundt DH, Fejer MM, Byer RL. IEEE J Quant Electron 1990;26:135.
[65] O'Bryan HM, Gallagher PK, Brandle CD. J Am Ceram Soc 1985;68:493.
[66] Kitamura K, Furukawa Y, Niwa K, Gopalan V, Mitchell TE. Appl Phys Lett 1998;73:3073.
[67] Tian L, Gopalan V, Galambos L. Appl Phys Lett 2004;85:4445.
[68] Uda S, Tiller WA. J Cryst Growth 1993;126:396.
[69] Kushibiki J, Ohashi Y. IEEE Trans Ultrason Ferroelectr Freq Control 2006;53:385.

5 Equilibrium Shape of Crystals

T.L. Einstein

DEPARTMENT OF PHYSICS AND CONDENSED MATTER THEORY CENTER, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD, USA

CHAPTER OUTLINE

5.1 Introduction **216**
5.2 From Surface Free Energies to Equilibrium Crystal Shape **217**
5.2.1 General Considerations 217
5.2.2 More Formal Treatment 220
5.3 Applications of Formal Results **224**
5.3.1 Cusps and Facets 224
5.3.2 Sharp Edges and Forbidden Regions 225
5.3.3 Experiments on Lead Going beyond Wulff Plots 227
5.4 Some Physical Implications of Wulff Constructions **229**
5.4.1 Thermal Faceting and Reconstruction 229
5.4.2 Types A and B 231
5.4.3 2D Studies 233
5.5 Vicinal Surfaces–Entrée to Rough Regions Near Facets **234**
5.6 Critical Behavior of Rough Regions Near Facets **239**
5.6.1 Theory 239
5.6.2 Experiment on Leads 241
5.6.3 Summary of Highlights of Novel Approach to Behavior Near Smooth Edges 243
5.7 Sharp Edges and First-Order Transitions—Examples and Issues **245**
5.7.1 Sharp Edges Induced by Facet Reconstruction 245
5.8 Gold–Prototype or Anomaly of Attractive Step–Step Interaction? **250**
5.9 Well-Established Attractive Step–Step Interactions Other Than ℓ^{-2} **253**
5.9.1 Atomic-Range Attractions 253
5.9.2 Attractions at Periodic Ranges of Separation via Oscillatory Friedel-Type Interactions 255
5.10 Conclusions **257**
Acknowledgments **257**
References **258**

Handbook of Crystal Growth. http://dx.doi.org/10.1016/B978-0-444-56369-9.00005-1

5.1 Introduction

The notion of equilibrium crystal shape (ECS) is arguably the platonic ideal of crystal growth and underpins much of our thinking about crystals. Accordingly, it has been the subject of several special reviews and tutorials [1–4,215] and is a prominent part of most volumes and extended review articles and texts about crystals and their growth [5–9]. In actual situations, there are many complications that thwart observation of such behavior, including kinetic barriers, impurities, and other bulk defects like dislocations. Furthermore, the notion of a well-defined equilibrium shape requires that there is no contact of the crystal with a wall or surface, since that would alter its shape. By the same token, the crystal cannot then be supported, so gravity is neglected. For discussions of the effect of gravity or contact with walls, see, e.g., Ref. [7].

Gibbs [10] is generally credited with being the first to recognize that the equilibrium shape of a substance is that which, for a fixed volume, minimizes the (orientation-dependent) surface free energy integrated over the entire surface; the bulk free energy is irrelevant since the volume is conserved, while edge or corner energies are ignored as being higher order effects that play no role in the thermodynamic limit. Herring [11,12] Surveys the early history in detail: The formulation of the problem was also carried out independently by Ref. [13]. The solution of this ECS problem, the celebrated Wulff construction, was stated by Ref. [14]; but his proof was incorrect. Correct proofs were subsequently given by Ref. [15–17], who presented a critical review. However, these proofs, while convincing of the theorem, were not general (and evidently applied only to $T = 0$, since they assumed the ECS to be a polyhedron and, compared the sum over the facets of the surface free energy of each facet times its area with a similar sum over a similar polyhedron with the same facet planes but slightly different areas (and the same volume)). Dinghas [18] showed that the Brunn–Minkowski inequality could be used to prove directly that any shape differing from that resulting from the Wulff construction has a higher surface free energy. Although Dinghas again considered only a special class of polyhedral shapes, Herring [11,12] completed the proof by noting that Dinghas' method is easily extended to arbitrary shapes, since the inequality is true for convex bodies in general. In their seminal paper on crystal growth, Burton, Cabrera, and Frank [19] present a novel proof of the theorem in two dimensions (2D).

Since equilibrium implies minimum Helmholtz free energy for a given volume and number, and since the bulk free energy is ipso facto independent of shape, the goal is to determine the shape that minimizes the integrated surface free energy of the crystal. The prescription takes the following form: One begins by creating a polar plot of the surface free energy as a function of orientation angle (of the surface normal) and draws a perpendicular plane (or line in 2D) through the tip of each ray. (There are many fine reviews of this subject in Refs [20–23].) Since the surface free energy in three dimensions (3D) is frequently denoted by γ, this is often called a γ plot. The shape is then the formed by the interior envelope of these planes or lines, often referred to as a pedal. At zero

temperature, when the free energy is just the energy, this shape is a polyhedron in 3D and a polygon in 2D, each reflecting the symmetry of the underlying lattice. At finite temperature, the shapes become more complex. In 2D, the sharp corners are rounded. In 3D, the behavior is richer, with two possible modes of evolution with rising temperature. For what Wortis terms type-A crystals, all sharp boundaries smooth together, while in type-B, first the corners smooth, then above a temperature denoted T_0, the edges also smooth. The smooth regions correspond to thermodynamic rough phases, with height–height correlation functions that diverge for large lateral separation l—like l^{α}, with α (typically $0 < \alpha < 1$) called the roughening exponent—in contrast to facets, where they attain some finite value as $l \to \infty$ [5]. The faceted regions in turn correspond to "frozen" regions. Pursuing the correspondence, sharp and smooth edges correspond to first-order and second-order phase transitions, respectively.

The aim of this chapter is primarily to explore physical ideas regarding ECS and the underlying Wulff constructions. This topic has also attracted considerable interest in the mathematics community. Readers interested in more formal and sophisticated approaches are referred to two books, Refs [24,25] and to many articles, including [26–36]. Particular attention is devoted to the origin of sharp edges on the ECS, the impact of reconstructed or adsorbed surface phases coexisting with unadorned phases, and the role and nature of possible attractive stepestep interactions.

5.2 From Surface Free Energies to Equilibrium Crystal Shape

5.2.1 General Considerations

To examine this process more closely, we examine the free energy expansion for a vicinal surface, that is, a surface misoriented by some angle θ from a facet direction. Cf. Figure 5.1. Unfortunately, this polar angle is denoted by ϕ in much of the literature on vicinal surfaces, with θ used for in-plane misorientation; most reviews of ECS use θ for the polar angle, as we shall here. The term *vicinal* implies that the surface is in the vicinity of the orientation. It is generally assumed that the surface orientation itself is rough (while the facet direction is below its roughening temperature and so is smooth). We consider the projected surface free energy $f_p(\theta,T)$ [37] (with the projection being onto the low-index reference, facet direction of terraces):

$$f_p(\theta,T) = f_0(T) + \beta(T)\frac{|\tan\theta|}{h} + g(T)|\tan\theta|^3 + c(\tan\theta)^4. \tag{5.1}$$

The first term is the surface free energy per area of the terrace (facet) orientation; it is often denoted σ. The average density of steps (the inverse of their mean separation $\langle\ell\rangle$) is $\tan\theta/h$, where h is the step height. In the second term, $\beta(T)$ is the line tension or free energy per length of step formation. (Since 2D is a dimension smaller than 3D, one uses β rather than γ. Skirting over the difference in units resulting from the dimensional

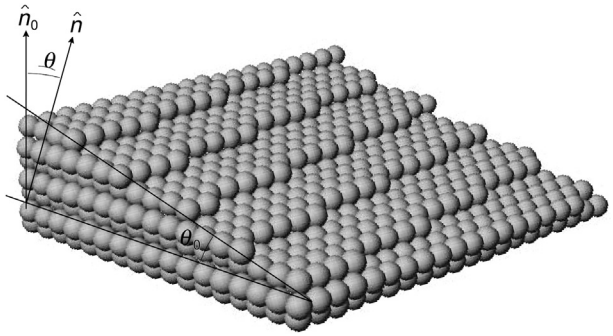


FIGURE 5.1 Portion of a (3,$\bar{2}$,16) surface, vicinal to an fcc (001), to illustrate a misoriented, vicinal surface. The vicinal surface and terrace normals are $\hat{n} = (3, -2, 16)/\sqrt{269}$ and $\hat{n} = (0, 0, 1)$, respectively. The polar angle θ [with respect to the (001) direction], denoted ϕ in the original figure (consistent with most of the literature on vicinal surfaces), is across $(16/\sqrt{269})$, while azimuthal angle φ (denoted θ in most of the literature on vicinal surfaces), indicating how much $\hat{n}$ is rotated around $\hat{n}_0$ away from the vertical border on which θ_0 is marked, is clearly arctan(1/5); $\tan\theta_0 = \tan\theta_0 \cos\phi$. Since h is $a_1/\sqrt{2}$, where a_1 is the nearest-neighbor spacing, the mean distance ℓ (in a terrace plane) between steps is $a_1/(\sqrt{2}\tan\theta) = 8\sqrt{2/13}a_1 = 3.138a_1$. While the average distance from one step to the next along a principal, (110) direction looks like $3.5a_1$, it is in fact $a_1/(\sqrt{2}\tan\theta_0) = 3.2a_1$. The "projected area" of this surface segment, used to compute the surface free energy f_p, is the size of a (001) layer: $20a_1 \times 17a_1 = 340a_1^2$; the width is $20a_1$. In "Maryland notation" (see text), z is in the $\hat{n}_0$ direction, while the normal to the vicinal, $\hat{n}$, lies in the x–z plane and y runs along the mean direction of the edges of the steps. In most discussions, $\phi = 0$, so that this direction would be that of the upper and lower edges of the depicted surface. *Adapted from Ref. [38].*

difference, many use γ in both cases.) While step free energy per length and line tension are equivalent for these systems, where the surface is at constant (zero) charge, they are inequivalent in electrochemical systems, where it is the electrode potential conjugate to the surface charge that is held fixed [39]. The third term is associated with interactions between steps, in this case assumed to be proportional to ℓ^{-2} (so that this term, which also includes the ℓ^{-1} density of steps, goes like ℓ^{-3}). The final term is the leading correction.

The ℓ^{-2} interaction is due to a combination (not a simple sum) of two repulsive potential energies: the entropic repulsion due to the forbidden crossing of steps and the elastic repulsions due to dipolar strains near each step. An explicit form for $g(T)$ is given in Eqn (5.27) below. The ℓ^{-2} of the entropic interaction can be understood from viewing the step as performing a random walk in the direction between steps (the x direction in "Maryland notation"[1] as a function of the y direction (which is timelike in the fermion transcription to be discussed later), cf. Figure 5.1, so the distance (y) it must go until it touches a neighboring step satisfies $\ell^2 \propto y$. To get a crude understanding of the origin of the elastic repulsion, one can imagine that since a step is unstable relative to a flat surface, it will try to relax toward a flatter shape, pushing atoms away from the location of the step by a distance decaying with distance from the step. When two steps are close

[1]This term was coined by a speaker at a workshop in Traverse City in August 1996—see Ref. [43] for the proceedings—and then used by several other speakers.

to each other, such relaxation will be frustrated because atoms on the terrace this pair of steps are pushed in opposite directions, so they relax less than if the steps are widely separated, leading to a repulsive interaction. Analyzed in more detail [7,40,41], this repulsion is dipolar and so proportional to ℓ^{-2}. However, attempts to reconcile the prefactor with the elastic constants of the surface have met with limited success. The quartic term in Eqn (5.1) is due to the leading (ℓ^{-3}) correction to the elastic repulsion [42], a dipole-quadrupole repulsion. It generally has no significant consequences but is included to show the leading correction to the critical behavior near a smooth edge on the ECS, to be discussed below.

The absence of a quadratic term in Eqn (5.1) reflects that there is no ℓ^{-1} interaction between steps. In fact, there are some rare geometries, notably vicinals to (110) surfaces of fcc crystals (Au in particular) that exhibit what amounts to ℓ^{-1} repulsions, which lead to more subtle behavior [44]. Details about this fascinating idiosyncratic surface are beyond the scope of this chapter; readers should see the thorough, readable discussion by van Albada et al. [45].

As temperature increases, $\beta(T)$ decreases due to increasing entropy associated with step-edge excitations (via the formation of kinks). Eventually, $\beta(T)$ vanishes at a temperature T_R associated with the roughening transition. At and above this T_R of the facet orientation, there is a profusion of steps, and the idea of a vicinal surface becomes meaningless. For rough surfaces, the projected surface free energy $f_p(\theta,T)$ is quadratic in $\tan\theta$. To avoid the singularity at $\theta = 0$ in the free energy expansion that thwarts attempts to proceed analytically, some treatments, notably Bonzel and Preuss [46], approximate $f_p(\theta,T)$ as quadratic in a small region near $\theta = 0$. It is important to recognize that the vicinal orientation is thermodynamically rough, even though the underlying facet orientation is smooth. The two regions correspond to incommensurate and commensurate phases, respectively. Thus, in a rough region, the mean spacing $\langle\ell\rangle$ between steps is not in general simply related to (i.e., an integer multiple plus some simple fraction) the atomic spacing.

Details of the roughening process have been reviewed by Weeks [216] and by van Beijeren and Nolden [9]; the chapter by Akutsu in this Handbook provides an up-to-date account. However, for use later, we note that much of our understanding of this process is rooted in the mapping between the restricted body-centered (cubic) solid-on-solid (BCSOS) model and the exactly solvable [47,48] symmetric 6-vertex model [49], which has a transition in the same universality class as roughening. This BCSOS model is based on the BCC crystal structure, involving square net layers with ABAB stacking, so that sites in each layer are lateral displaced to lie over the centers of squares in the preceding (or following) layer. Being an SOS model means that for each column of sites along the vertical direction, there is a unique upper occupied site, with no vacancies below it or floating atoms above it. Viewed from above, the surface is a square network with one pair of diagonally opposed corners on A layers and the other pair on B layers. The restriction is that neighboring sites must be on adjacent layers (so that their separation is the distance from a corner to the center of the BCC lattice). There are then six possible

configurations: two in which the two B corners are both either above or below the A corners and four in which one pair of catercorners are on the same layer and the other pair are on different layers (one above and one below the first pair). In the symmetric model, there are three energies, $-\epsilon$ for the first pair, and $\pm\delta/2$ for the others, the sign depending on whether the catercorner pair on the same lattice is on A or B [50]. The case $\delta = 0$ corresponds to the *F*-model, which has an infinite-order phase transition and an essential singularity at the critical point, in the class of the Kosterlitz-Thouless [51] transition [52]. (In the "ice" model, ϵ also is 0.) For the asymmetric 6-vertex model, each of the six configurations can have a different energy; this model can also be solved exactly [53,54].

5.2.2 More Formal Treatment

To proceed more formally, we largely follow [1]. The shape of a crystal is given by the length $R(\widehat{\mathbf{h}})$ of a radial vector to the crystal surface for any direction $\widehat{\mathbf{h}}$. The shape of the crystal is defined as the thermodynamic limit of this crystal for increasing volume V, specifically.

$$r\left(\widehat{\mathbf{h}}, T\right) \equiv \lim_{V\to\infty}\left[R\left(\widehat{\mathbf{h}}\right) \Big/ \alpha V^{1/3}\right], \tag{5.2}$$

where α is an arbitrary dimensionless variable. This function $r(\widehat{\mathbf{h}}, T)$ corresponds to a free energy. In particular, since both independent variables are fieldlike (and so intrinsically intensive), this is a Gibbs-like free energy. Like the Gibbs free energy, $r(\widehat{\mathbf{h}}, T)$ is continuous and convex in $\widehat{\mathbf{h}}$.

The Wulff construction then amounts to a Legendre transformation[2] to $r(\widehat{\mathbf{h}}, T)$ from the orientation $\widehat{\mathbf{m}}$-dependent interfacial free energy $f_i(\widehat{\mathbf{m}}, T)$ (or in perhaps the more common but less explicit notation, $\gamma(\widehat{\mathbf{m}}, T)$, which is $f_p(\theta, T)/\cos(\theta)$. For liquids, of course,$f_i(\widehat{\mathbf{m}}, T)$ is spherically symmetric, as is the equilibrium shape. Herring [12] mentions rigorous proofs of this problem by Schwarz in 1884 and by Minkowski in 1901. For crystals, $f_i(\widehat{\mathbf{m}}, T)$ is not spherically symmetric but does have the symmetry of the crystal lattice. For a system with cubic symmetry, one can write

$$f_i\left(\widehat{\mathbf{m}}, T\right) = \gamma_0(T)\left[1 + a(T)\left(m_x^4 + m_y^4 + m_z^4\right)\right], \tag{5.3}$$

where $\gamma_0(T)$ and $a(T)$ are constants. As illustrated in Figure 5.2, for $a = 1/4$ the asymmetry leads to minor distortions, which are rather inconsequential. However, for

[2]As exposited clearly in Ref. [55], one considers a [convex] function $y = y(x)$ and denotes its derivative as $p = \partial y/\partial x$. If one then tries to consider p instead of x as the independent variable, there is information lost: one cannot reconstruct $y(x)$ uniquely from $y(p)$. Indeed, $y = y(p)$ is a first-order differential equation, whose integration gives $y = y(x)$ only to within an undetermined integration constant. Thus, $y = y(p)$ corresponds to a family of displaced curves, only one of which is the original $y = y(x)$. The key concept is that the locus of points satisfying $y = y(x)$ can be equally well represented by a family of lines tangent to $y(x)$ at all x, each with a y-intercept ψ determined by the slope p at $(x,y(x))$. That is, $\psi = \psi(p)$ contains all the information of $y = y(x)$. Recognizing that $p = (y - \psi)/(x - 0)$, one finds the transform $\psi = y - px$. Readers should recall that this is the form of the relationship between thermodynamic functions, particularly the Helmholtz and the Gibbs free energies.

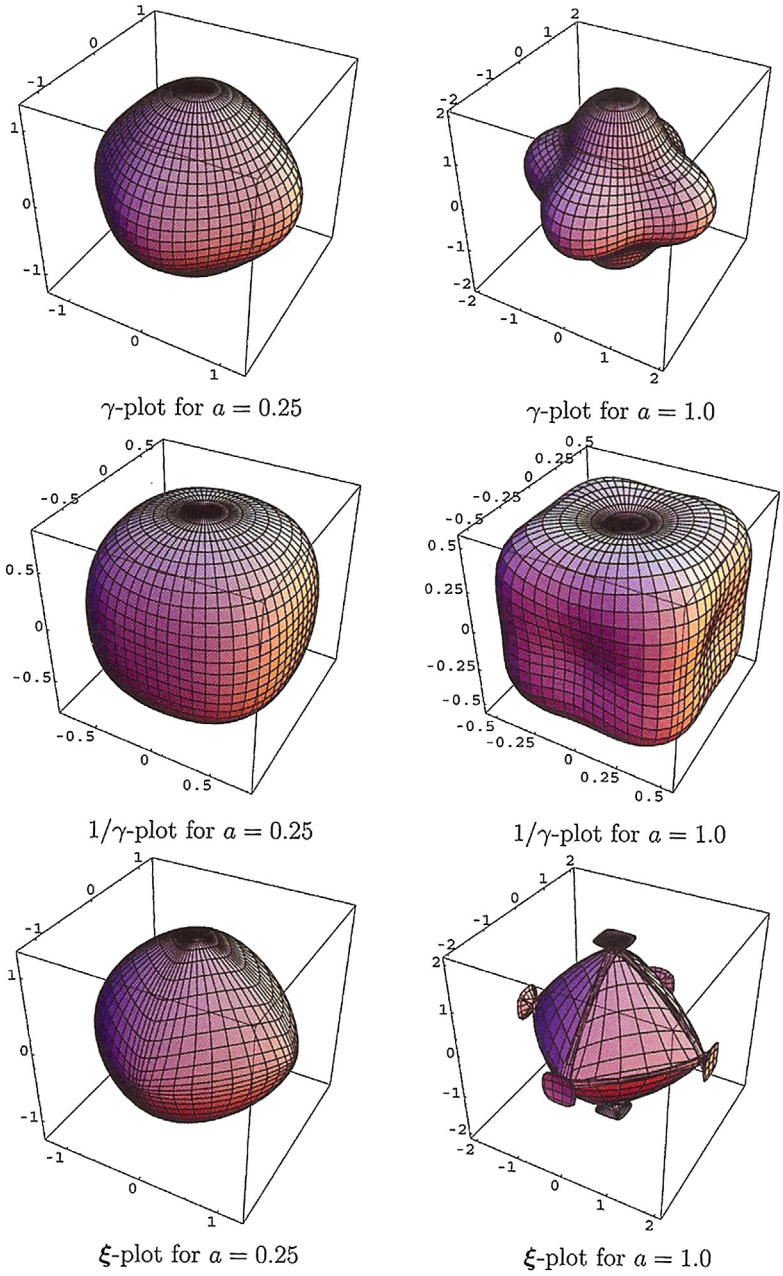


FIGURE 5.2 γ-plots (plots of $f_i(\widehat{\mathbf{m}})$, $1/\gamma$-plots and ξ-plots for Eqn (5.3) for positive values of *a*). For $a = 1/4$, all orientations appear on the ECS. For $a = 1.0$, the $1/\gamma$-plot has concave regions, and the ξ-plot has ears and flaps that must be truncated to give the ECS essentially an octahedron with curved faces. *From Ref. [8]*, which shows in a subsequent figure that the γ- and $1/\gamma$-plots for $a = -0.2$ and -0.5 resemble the $1/\gamma$- and γ-plots, respectively, for $a = 1/4$ and 1.

$a = 1$, the enclosed region is no longer convex, leading to an instability to be discussed shortly.

One considers the change in the interfacial free energy associated with changes in shape. The constraint of constant volume is incorporated by subtracting from the change in the integral of $f_i(\widehat{\mathbf{m}}, T)$ the corresponding change in volume, multiplied by a Lagrange multiplier λ. Herring [11,12] showed that this constrained minimization problem has a unique and rather simple solution that is physically meaningful in the limit that it is satisfactory to neglect edge, corner, and kink energies in $f_i(\widehat{\mathbf{m}}, T)$, that is, in the limit of large volume. In this case $\lambda \propto V^{-1/3}$; by choosing the proportionality constant as essentially the inverse of α, we can write the result as

$$r\left(\widehat{\mathbf{h}}, T\right) = \min_{\widehat{\mathbf{m}}}\left(\frac{f_i\left(\widehat{\mathbf{m}}, T\right)}{\widehat{\mathbf{m}} \cdot \widehat{\mathbf{h}}}\right) \tag{5.4}$$

The Wulff construction is illustrated in Figure 5.3. The interfacial free energy $f_i(\widehat{\mathbf{m}})$, at some assumed T is displayed as a polar plot. The crystal shape is then the interior envelope of the family of perpendicular planes (lines in 2D) passing through the ends of the radial vectors $\widehat{\mathbf{m}} f_i(\widehat{\mathbf{m}})$. Based on Eqn (5.4) one can, at least in principle, determine $\widehat{\mathbf{m}}(\widehat{\mathbf{h}})$ or $\widehat{\mathbf{h}}(\widehat{\mathbf{m}})$, which thus amounts to the equation of state of the equilibrium crystal shape. One can also write the inverse of Eqn (5.4):

$$\frac{1}{f_i\left(\widehat{\mathbf{m}}, T\right)} = \min_{\widehat{\mathbf{m}}}\left(\frac{1/f_i\left(\widehat{\mathbf{h}}, T\right)}{\widehat{\mathbf{m}} \cdot \widehat{\mathbf{h}}}\right) \tag{5.5}$$

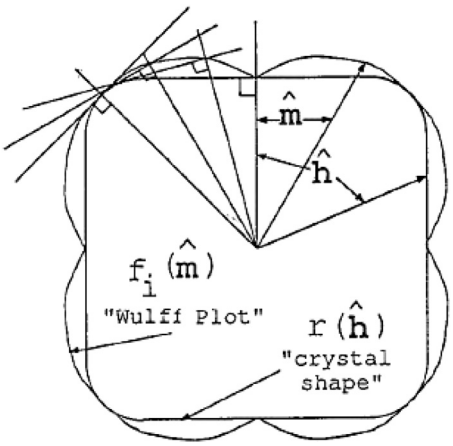


FIGURE 5.3 Schematic of the Wulff construction. The interfacial free energy per unit area $f_i\widehat{\mathbf{m}}$ is plotted in polar form (the "Wulff plot" or "γ-plot"). One draws a radius vector in each direction $\widehat{\mathbf{m}}$ and constructs a perpendicular plane where this vector hits the Wulff plot. The interior envelope of the family of "Wulff planes" thus formed, expressed algebraically in Eqn (5.4), is the crystal shape, up to an arbitrary overall scale factor that may be chosen as unity. *From Ref. [1].*

Thus, a Wulff construction using the inverse of the crystal shape function yields the inverse free energy.

To be more explicit, consider the ECS in Cartesian coordinates $z(x,y)$, i.e., $\widehat{\mathbf{h}} \propto (x, y, z(x,y))$, assuming (without dire consequences [1]) that $z(x,y)$ is single-valued. Then, for any displacement to be tangent to $\widehat{\mathbf{h}}$, $dz - p_x dx - p_y dy = 0$

$$\widehat{\mathbf{h}} = \frac{1}{\sqrt{1 + p_x^2 + p_y^2}} \Big(- p_x z, -p_y z, 1 \Big), \tag{5.6}$$

where p_x is shorthand for $\partial z/\partial x$.

Then the total free energy and volume are

$$\begin{aligned} F_i(T) &= \iint f_p(p_x, p_y)\, dx\, dy \\ V &= \iint z(x,y) dx\, dy \end{aligned} \tag{5.7}$$

where f_p, which incorporates the line-segment length, is $f_p \equiv [1 + p_x^2 + p_y^2]^{1/2} f_i$. Minimizing F_i subject to the constraint of fixed V leads to the Euler–Lagrange equation

$$\frac{\partial}{\partial x} \frac{f_p(\partial_x z, p_y)}{p_x} + \frac{\partial}{\partial y} \frac{f_p(p_x, p_y)}{p_y} = -2\lambda \tag{5.8}$$

(Actually, one should work with macroscopic lengths, then divide by the $V^{1/3}$ times the proportionality constant. Note that this leaves p_x and p_y unchanged [1].) On the right-hand side, 2λ can be identified as the chemical potential μ, so that the constancy of the left-hand side is a reflection of equilibrium. Equation (5.8) is strictly valid only if the derivatives of f_p exist, so one must be careful near high-symmetry orientations below their roughening temperature, for which facets occur. To show that this highly nonlinear second-order partial differential equation with unspecified boundary conditions is equivalent to Eqn (5.4), we first note that the first integral of Eqn (5.8) is simply

$$z - x p_x - y p_y = f_p(p_x, p_y) \tag{5.9}$$

The right-hand side is just a function of derivatives, consistent with this being a Legendre transformation. Then, differentiating yields

$$x = -\partial f_p \Big/ \partial(p_x), \quad y = -\partial f_p \Big/ \partial(p_y) \tag{5.10}$$

Hence, one can show that

$$z(x,y) = \min_{p_x, p_y} \big(f_p(p_x, p_y) + x p_x + y p_y \big) \tag{5.11}$$

5.3 Applications of Formal Results

5.3.1 Cusps and Facets

The distinguishing feature of Wulff plots of faceted crystals compared to liquids is the existence of (pointed) cusps in $f_i(\widehat{\mathbf{m}}, T)$, which underpin these facets. The simplest way to see why the cusp arises is to examine a square lattice with nearest-neighbor bonds having bond energy ϵ_1, often called a 2D Kossel [56,57] crystal; note also [210]. In this model, the energy to cleave the crystal is the Manhattan distance between the ends of the cut; i.e., as illustrated in Figure 5.4, the energy of severing the bonds between (0,0) and (*X*,*Y*) is just $+\epsilon_1\,(|X| + |Y|)$. The interfacial area, i.e., length, is $2(X^2 + Y^2)$ since the cleavage creates *two* surfaces. At $T = 0$, entropy plays no role, so that

$$f_i(\theta) = \frac{\epsilon_1}{2}(|\sin\theta| + |\cos\theta|) \sim \frac{\epsilon_1}{2}(1 + |\theta| + \ldots) \tag{5.12}$$

At finite T fluctuations and attendant entropy do contribute, and the argument needs more care. Recalling Eqn (5.1), we see that if there is a linear cusp at $\theta = 0$, then

$$f_i(\theta, T) = f_i(0, T) + B(T)|\theta|, \tag{5.13}$$

where $B = \beta(T)/h$, since the difference between $f_i(\theta)$ and $f_p(\theta)$ only appears at order θ^2. Comparing Eqns (5.12) and (5.13), we see that for the Kossel square $f_i(0,0) = \epsilon_1/2$ and $B(0) = \epsilon_1/2$. Further discussion of the 2D $f_i(\theta)$ is deferred to Section 5.4.3 below.

To see how a cusp in $f_i(\widehat{\mathbf{m}}, T)$ leads to a facet in the ECS, consider Figure 5.5: the Wulff plane for $\theta \gtrsim 0$ intersects the horizontal $\theta = 0$ plane at a distance $f_i(0) + d(\theta)$ from the vertical axis. The crystal will have a horizontal axis if and only if $d(\theta)$ does not

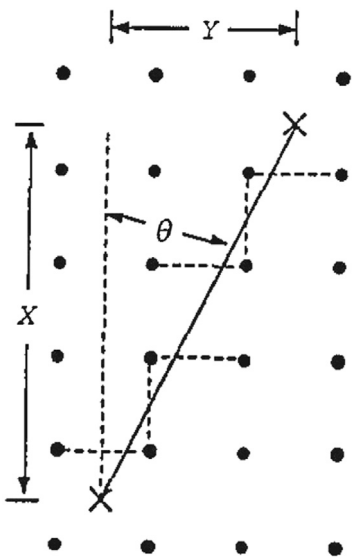


FIGURE 5.4 Kossel crystal at $T = 0$. The energy to cleave the crystal along the depicted slanted. Interface ($\tan\theta = Y/Z$) is $\epsilon_1\,(|X| + |Y|)$. *From Ref. [1].*

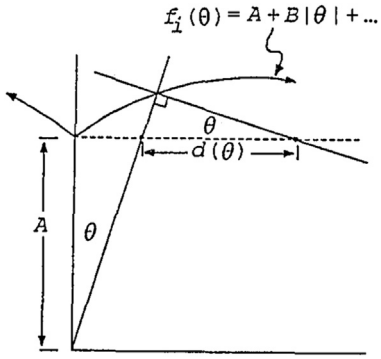


FIGURE 5.5 Wulff plot with a linear cusp at $\theta = 0$. If $d(\theta) \to 0$ as $\theta \to 0$, then there is no facet corresponding to $\theta = 0$, and the $\theta = 0$ Wulff plane (dotted line) is tangent to the crystal shape at just a single point. Since $d(\theta) = B$, a cusp in the Wulff plot leads to a facet of the corresponding orientation on the equilibrium crystal shape. *From Ref. [1].*

vanish as $\theta \to 0$. From Figure 5.5, it is clear that $\theta \approx \sin\theta \approx B\theta/d(\theta)$ for θ near 0, so that $d(0) = B > 0$. For a weaker dependence on θ, e.g., $B|\theta|^{\zeta}$ with $\zeta > 1$, $d(0) = 0$, and there is no facet. Likewise, at the roughening temperature, β vanishes and the facet disappears.

5.3.2 Sharp Edges and Forbidden Regions

When there is a sharp edge (or corner) on the ECS $r(\widehat{\mathbf{h}}, T)$, Wulff planes with a range of orientations $\widehat{\mathbf{m}}$ will not be part of the inner envelope determining this ECS; they will lie completely outside it. There is no portion of the ECS whose surface tangent has these orientations. As in the analogous problems with forbidden values of the "density" variable, the free energy $f_i(\widehat{\mathbf{m}}, T)$ is actually not properly defined for forbidden values of $\widehat{\mathbf{m}}$; those unphysical values should actually be removed from the Wulff plot. Figure 5.6 depicts several possible ECSs and their associated Wulff plots. It is worth emphasizing that, in the extreme case of the fully faceted ECS at $T = 0$, the Wulff plot is simply a set of discrete points in the facet directions.

Now if we denote by $\widehat{\mathbf{m}}_+$ and $\widehat{\mathbf{m}}_-$, the limiting orientations of the tangent planes approaching the edge from either side, then all intermediate values do not occur as stable orientations. These missing, not stable, "forbidden" orientations are just like the forbidden densities at liquid–gas transitions, forbidden magnetizations in ferromagnets at $T < T_c$ [58], and miscibility gaps in binary alloys. Herring [11,12] first presented an elegant way to determine these missing orientations using a spherical construction. For any orientation $\widehat{\mathbf{m}}$, this tangent sphere (often called a Herring sphere) passes through the origin and is tangent to the Wulff plot at $f_i(\widehat{\mathbf{m}})$. From geometry, he invoked the theorem that an angle inscribed in a semicircle is a right angle. Thence, if the orientation $\widehat{\mathbf{m}}$ appears on the ECS, it appears at an orientation that points outward along the radius of that sphere. Herring then observes that only if such a sphere lies completely inside the plot of $f_i(\widehat{\mathbf{m}})$ does that orientation appear on the ECS. If some part were inside, its Wulff

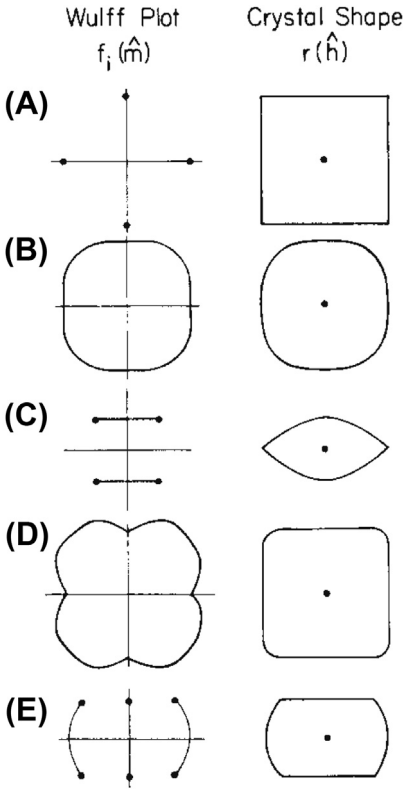


FIGURE 5.6 Some possible Wulff plots and corresponding equilibrium crystal shapes. Faceted and curved surfaces may appear, joined at sharp or smooth edges in a variety of combinations. *From Ref. [4]; the equilibrium crystal shape are also in Ref. [12].*

plane would clip off the orientation of the point of tangency, so that orientation would be forbidden.

The origin of a hill-and-valley structure from the constituent free energies [59,221,222] is illustrated schematically in Figure 5.7. It arises when they satisfy the inequality

$$f_i\left(\widehat{\mathbf{m}} = \mathbf{n}_1\right)A_1 + f_i(\mathbf{n}_2)A_2 < f_i(\mathbf{n})A, \tag{5.14}$$

where A_1 and A_2 are the areas of strips of orientation $\mathbf{n}_1$ and $\mathbf{n}_2$, respectively, while A is the area of the sum of these areas projected onto the plane bounded by the dashed lines in the figure. This behavior, again, is consistent with the identification of the misorientation as a density (or magnetizationlike) variable rather than a fieldlike one.

The details of the lever rule for coexistence regimes were elucidated by Wortis [1]: As depicted in Figure 5.8, which denotes as P and Q the two orientations bounding the region that is not stable, the lever rule interpolations lie on segments of a spherical surface. Let the edge on the ECS be at R. Then an interface created at a forbidden $\widehat{\mathbf{m}}$ will

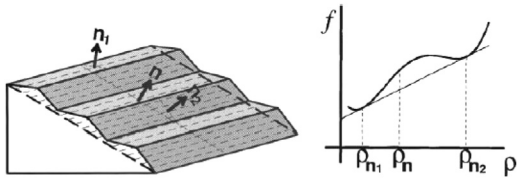


FIGURE 5.7 Illustration of how orientational phase separation occurs when a "hill-and-valley" structure has a lower total surface free energy per area than a flat surface as in Eqn (5.14). The sketch of the free energy versus $\rho \equiv \tan\theta$ shows that this situation reflects a region with negative convexity which is accordingly not stable. The dashed line is the tie bar of a Maxwell or double-tangent construction. The misorientations are the coexisting slatlike planes, with orientations $\mathbf{n}_1$ and $\mathbf{n}_2$, in the hill-and-valley structure. *From Ref. [59].*

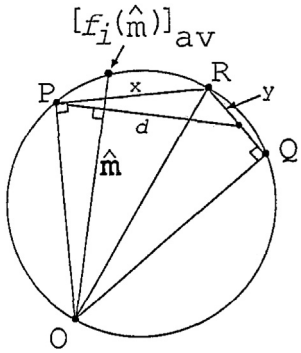


FIGURE 5.8 Equilibrium crystal shape (ECS) analogue of the Maxwell double-tangent construction. O is the center of the crystal. Points P and Q are on the (stable) Wulff plot, but the region between them is unstable; hence, the ECS follows PRQ and has an edge at R. An interface at the intermediate orientation $\widehat{\mathbf{m}}$ breaks up into the orientations P and Q with relative proportions x:y; thus, the average free energy per unit area is given by Eqn (5.15), which in turn shows that $f_i(\widehat{\mathbf{m}})_{\text{avr}}$ lies on the circle. *From Ref. [1].*

evolve toward a hill-and-valley structure with orientations P and Q with a free energy per area of

$$\left[f_i\left(\widehat{\mathbf{m}}\right)\right]_{\text{avr}} = \frac{\mathrm{x}f_i(\mathrm{P}) + \mathrm{y}f_i(\mathrm{Q})}{d}. \tag{5.15}$$

It can then be shown that $\widehat{\mathbf{m}}[f_i(\widehat{\mathbf{m}})]_{\text{avr}}$ lies on the depicted circle, so that the Wulff plane passes through the edge at R.

5.3.3 Experiments on Lead Going beyond Wulff Plots

To determine the limits of forbidden regions, it is more direct and straightforward to carry out a polar plot of $1/f_i(\widehat{\mathbf{m}})$ [20] rather than $f_i(\widehat{\mathbf{m}})$, as discussed in Sekerka's review chapter [8]. Then a sphere passing through the origin becomes a corresponding plane; in particular, a Herring sphere for some point becomes a plane tangent to the plot of $1/f_i(\widehat{\mathbf{m}})$. If the Herring sphere is inside the Wulff plot, then its associated plane lies outside the plot of $1/f_i(\widehat{\mathbf{m}})$. If, on the other hand, if some part of the Wulff plot is inside a

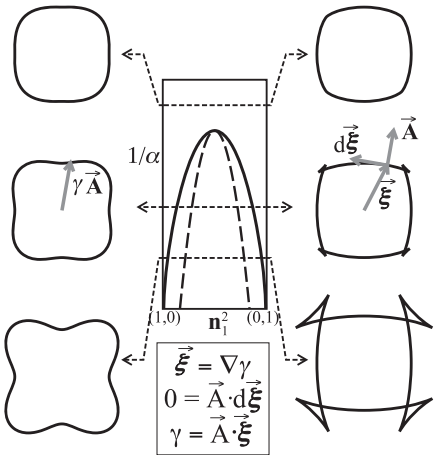


FIGURE 5.9 Graphical constructions for an anisotropic $f_i(\widehat{\mathbf{m}})$ for various values of an anisotropy parameter α, where $f_i \propto 1 + \alpha\cos^2\theta\sin^2\theta$. In the left column $f_i(\theta)$ is plotted from top to bottom for $\alpha = 1/2,1,2$. Anisotropy increases with positive α, so $1/\alpha$ corresponds in some sense to a temperature in conventional plots. In the center panel, $\mathbf{n}_1^2$ is $\cos^2\theta$. The shape resulting from the gradient construction with the ears removed is the Wulff equilibrium crystal shape. *From Ref. [60].*

Herring sphere, the corresponding part of the $1/f_i(\widehat{\mathbf{m}})$ plot will be outside the plane. Thus, if the plot of $1/f_i(\widehat{\mathbf{m}})$ is convex, all its tangent planes will lie outside, and all orientations will appear on the ECS. If it is not convex, it can be made so by adding tangent planes. The orientations associated with such tangent planes are forbidden, so their contact curve with the $1/f_i(\widehat{\mathbf{m}})$ plot gives the bounding stable orientations into which forbidden orientations phase separate.

Summarizing the discussion in Ref. [8], the convexity of $1/f_i(\widehat{\mathbf{m}})$ can indeed be determined analytically since the curvature $1/f_i(\widehat{\mathbf{m}})$ is proportional (with a positive-definite proportionality constant) to the stiffness, i.e., in 2D, $\gamma + \partial^2\gamma/\partial\theta^2 = \tilde{\gamma}$, or preferably $\beta + \partial^2\beta/\partial\theta^2 = \tilde{\beta}$ as in Eqn (5.1) to emphasize that the stiffness and (step) free energy per length have different units in 2D from 3D. Hence, $1/f_i(\widehat{\mathbf{m}})$ is not convex where the stiffness is negative. The very complicated generalization of this criterion to 3D is made tractable via the $\boldsymbol{\xi}$-vector formalism of Refs [30,61], where $\boldsymbol{\xi} = \nabla(r\, f_i(\widehat{\mathbf{m}}))$, where r is the distance from the origin of the γ plot. Thus.

$$f_i(\widehat{\mathbf{m}}) = \boldsymbol{\xi}\cdot\widehat{\mathbf{m}}, \quad \widehat{\mathbf{m}}\cdot d\boldsymbol{\xi} = 0, \tag{5.16}$$

which is discussed well by Refs [8,62]. To elucidate the process, we consider just the 2D case [60].

The solid curve in Figure 5.10 is the $\boldsymbol{\xi}$ plot and the dashed curve is the $1/\gamma$-plot for $f_i(\widehat{\mathbf{m}}) \equiv \gamma \propto 1 + 0.2\cos 4\theta$. For this case, the $1/\gamma$-plot is not convex and the $\boldsymbol{\xi}$ plot forms "ears." The equilibrium shape is given by the interior envelope of the $\boldsymbol{\xi}$ plot; in this case it exhibits four corners.

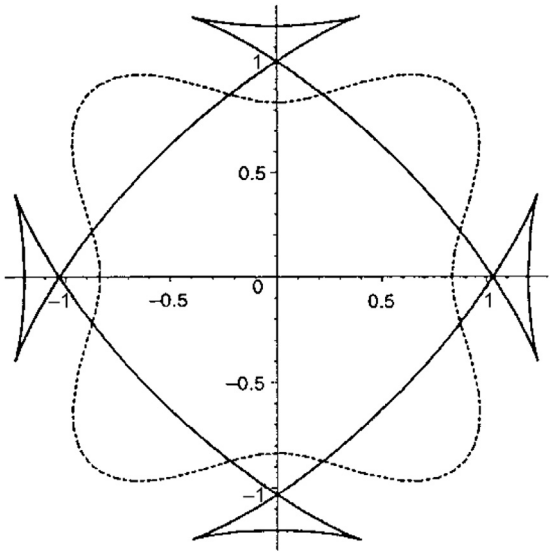


FIGURE 5.10 The solid curve is the ξ plot, while the dashed curve is the $1/\gamma$-plot for $f_i(\hat{m}) \equiv \gamma \propto 1 + 0.2\cos 4\,\theta$. For this case (but not for small values of a), the $1/\gamma$-plot is not convex, and the ξ plot forms "ears." These ears are then removed, so that the equilibrium shape is given by the interior envelope of the ξ plot, in this case having four corners. *From Ref. [62].*

Pursuing this analogy, we see that if one cleaves a crystal at some orientation $\widehat{\mathbf{m}}$ that is not on the ECS, i.e., between $\widehat{\mathbf{m}}_+$ and $\widehat{\mathbf{m}}_-$, then this orientation will break up into segments with orientations $\widehat{\mathbf{m}}_+$ and $\widehat{\mathbf{m}}_-$ such that the net orientation is still $\widehat{\mathbf{m}}$, providing another example of the lever rule associated with Maxwell double-tangent constructions for the analogous problems. The time to evolve to this equilibrium state depends strongly on the size of energy barriers to mass transport in the crystalline material; it could be exceedingly long. To achieve rapid equilibration, many nice experiments were performed on solid hcp ^{4}He bathed in superfluid ^{4}He, for which equilibration occurs in seconds [63–66], and many more (see Ref. [67] for a comprehensive recent review). Longer but manageable equilibration times are found for Si and for Au, Pb, and other soft transition metals.

5.4 Some Physical Implications of Wulff Constructions

5.4.1 Thermal Faceting and Reconstruction

A particularly dramatic example is the case of surfaces vicinal to Si (111) by a few degrees. In one misorientation direction, the vicinal surface is stable above the reconstruction temperature of the (111) facet, but below that temperature, $f_i(111)$ decreases significantly so that the original orientation is no longer stable and phase separates into reconstructed (111) terraces and more highly misoriented segments [68,69]. The

correspondence to other systems with phase separation at first-order transitions is even more robust. Within the coexistence regime, one can in mean field determine a spinodal curve. Between it and the coexistence boundary, one observes phase separation by nucleation and growth, as for metastable systems; inside the spinodal, one observes much more rapid separation with a characteristic most-unstable length [70]. This system is discussed further below.

Wortis [1] describes "thermal faceting" experiments in which metal crystals, typically late-transition or noble metal elements like Cu, Ag, and Fe, are cut at a high Miller index direction and polished. They are then annealed at high temperatures. If the initial plane is in a forbidden direction, optical striations, due to hill-and-valley formation, appear once these structures have reached optical wavelengths. While the characteristic size of this pattern continues to grow as in spinodal decomposition, the coarsening process is eventually slowed and halted by kinetic limitations.

There are more recent examples of such phenomena. After sputtering and annealing above 800 K, Au(4,5,5) at 300 K forms a hill-and-valley structure of two Au(111) vicinal surfaces, one that is reconstructed and the other not, as seen in Figure 5.11. This seems to be an equilibrium phenomenon: It is reversible and independent of cooling rate [77]. Furthermore, while it has been long known that adsorbed gases can induce faceting on bcc (111) metals [72], ultrathin metal films have also been found to produce faceting of W(111), W(211), and Mo(111) [73,74].

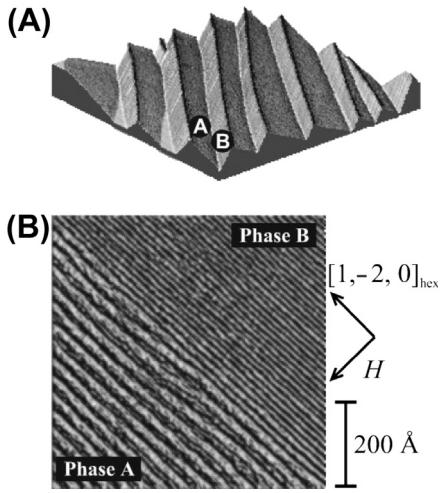


FIGURE 5.11 Morphology of the faceted Au(4,5,5) surface measured at room temperature. (A) 3D plot of a large-scale (scan area: 1.4 × 1.4 mm) scanning tunneling microscopy (STM) image. Phases A and B form the hill-and-valley morphology. (B) STM image zoomed in on a boundary between the two phases. All steps single-height, i.e., 2.35Å high. Phase B has smaller terraces, 13Å wide, while phase A terraces are about 30Å wide. This particular surface has (2,3,3) orientation. *From Ref. [71].*

5.4.2 Types A and B

The above analysis indicates that at $T = 0$, the ECS of a crystal is a polyhedron having the point symmetry of the crystal lattice, a result believed to be general for finite-range interactions [75]. All boundaries between facets are sharp edges, with associated forbidden nonfacet orientation; indeed, the Wulff plot is just a set of discrete points in the symmetry directions. At finite temperature, two possibilities have been delineated (with cautions [1], labeled nonmnemonically) A and B. In type A, there are smooth curves between facet planes rather than edges and corners. Smooth here means, of course, that not only is the ECS continuous, but so is its slope, so that there are no forbidden orientations anywhere. This situation corresponds to continuous phase transitions. In type B, in contrast, corners round at finite T but edges stay sharp until some temperature T_0. For $T_0 < T < T_1$, there are some rounded edges and some sharp edges, while above T_1 all edges are rounded.

Rottman and Wortis [4] present a comprehensive catalog of the orientation phase diagrams, Wulff plots, and ECSs for the cases of nonexistent, weakly attractive, and weakly repulsive next-nearest-neighbor (NNN) bonds in 3D. Figures 5.12 and 5.13 show the orientation phase diagrams and the Wulff plots with associated ECSs, respectively, for weakly attractive NNN bonds. As indicated in the caption, it is easy to describe what then happens when $\epsilon_2 = 0$ and only {100} facets occur. Likewise, Figures 5.14 and 5.15 show the orientation phase diagrams and the Wulff plots with associated ECSs, respectively, for weakly repulsive NNN bonds.

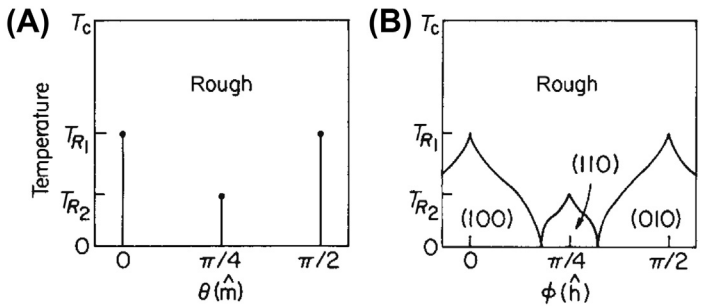


FIGURE 5.12 Interfacial phase diagrams for simple-cubic nearest-neighbor Kossel crystal with nearest-neighbor as well as (weak) next nearest-neighbor (NNN) attractions. The angular variables θ and ϕ (not to be confused with φ, cf. Section 5.2.1) interfacial orientation ($\widehat{\mathbf{m}}$) and equilibrium crystal shape ($\widehat{\mathbf{h}}$), respectively, in an equatorial section of the full 3D phase diagram. (A) The T–θ phase diagram (b) shows the locus of cusps in the Wulff plot along the symmetry directions below the respective roughening temperatures. For no NNN interaction ($\epsilon_2 = 0$), there are only cusps at vertical lines at 0 and $\pi/2$. (B) The $T - \widehat{\mathbf{h}}$ phase diagram gives the faceted areas of the crystal shape. The NNN attraction leads to additional (111) (not seen in the equatorial plane) and (110) facets at low enough temperature. Thus, for $\epsilon_2 = 0$ the two bases of the (100) and (010) phases meet and touch each other at (and only at) $\phi = \pi/4$ (at $T = 0$), with no intervening (110) phase. Each type of facet disappears at its own roughening temperature. Above the phase boundaries enclosing those regions, the crystal surfaces are smoothly curved (i.e., thermodynamically "rough"). This behavior is consistent with the observed phase diagram of hcp ^{4}He. *From Ref. [4].*

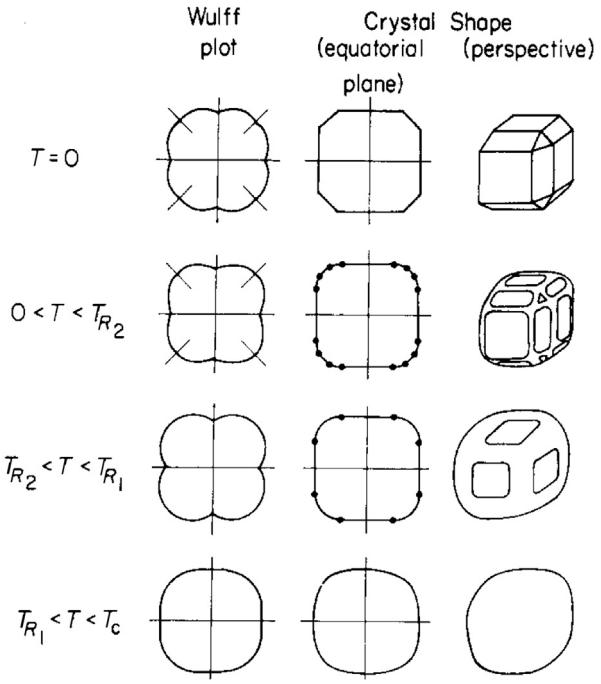


FIGURE 5.13 Representative Wulff plots and equilibrium crystal shapes for the crystal with weak next nearest-neighbor attractions whose phase diagram is shown in Figure 5.12. At low enough temperature there are (100), (110), and (111) facets. For weak attraction, the (110) and (111) facets roughen away below the (100) roughening temperature. For $\epsilon_2 = 0$, $T_{R_2} = 0$, so that the configurations in the second row do not occur; in the first row, the octagon becomes a square and the perspective shape is a cube. Facets are separated at $T > 0$ by curved surfaces, and all transitions are second order. Spherical symmetry obtains as T approaches melting at T_c. *From Ref. [4].*

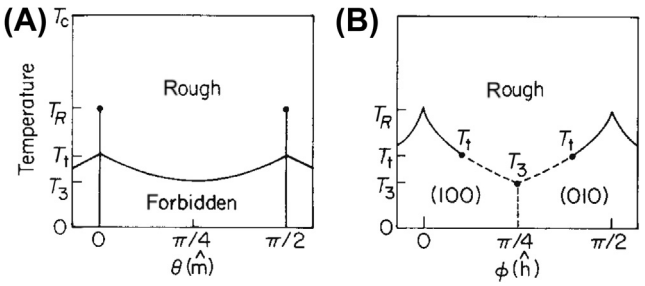


FIGURE 5.14 Interfacial phase diagram with (weak) next nearest-neighbor (NNN) repulsion rather than attraction as in Figure 5.12. The NNN repulsion stabilizes the (100) facets. Curved surfaces first appear at the cube corners and then reach the equatorial plane at T_3. The transition at the equator remains first order until a higher temperature T_t. The dotted boundaries are first order. A forbidden (coexistence) region appears in the $T - \hat{\mathbf{h}}$ phase diagram. *From Ref. [4].*

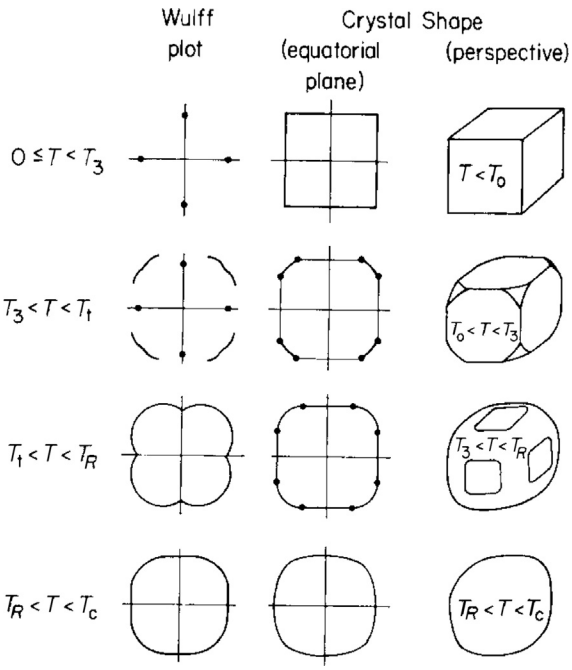


FIGURE 5.15 Representative Wulff plots and equilibrium crystal shapes for the crystal with weak next nearest-neighbor repulsions whose phase diagram is shown in Figure 5.14. Curved surfaces appear first at the cube corners. Junctions between facets and curved surfaces may be either first or second order (sharp or smooth), depending on orientation and temperature. *From Ref. [4].*

5.4.3 2D Studies

Exploring the details is far more transparent in 2D than in 3D. The 2D case is physically relevant in that it describes the shape of islands of atoms of some species at low fractional coverage on an extended flat surface of the same or another material. An entire book is devoted to 2D crystals [76]. The 2D perspective can also be applied to cylindrical surfaces in 3D, as shown by Ref. [7]. Formal proof is also more feasible, if still arduous, in 2D: An entire book is devoted to this task [25]; see also Refs [34,35].

For the 2D nearest-neighbor Kossel crystal described above [1] notes that at $T = 0$ a whole class of Wulff planes pass through a corner. At finite T, thermal fluctuations lift this degeneracy and the corner rounds, leading to type A behavior. To gain further insight, we now include a next nearest-neighbor (NNN) interaction ϵ_2, so that

$$f_i(\theta) = \frac{\epsilon_1 + \epsilon_2}{2}(|\cos\theta| + |\sin\theta|) + \frac{\epsilon_2}{2}(|\cos\theta| - |\sin\theta|) \tag{5.17}$$

For favorable NNN bonds, i.e., $\epsilon_2 > 0$, one finds new {11} facets but still type A behavior with sharp edges, while for unfavorable NNN bonds, i.e., $\epsilon_2 < 0$, there are no new facets but for finite T, the edges are no longer degenerate so that type B behavior obtains. Again recalling that $f_i(\theta) = f_p(\theta)\,|\cos\theta|$, we can identify $f_0 = \epsilon_2 + \epsilon_1/2$ and $\beta/h = \epsilon_1/2$, as noted in other treatments, e.g., Ref. [77]. That work, however, finds that such a model cannot adequately account for the orientation-dependent stiffness of islands on Cu(001).

Attempts to resolve this quandary using 3-site non-pairwise (trio) interactions [78,79] did not prove entirely satisfactory. In contrast, on the hexagonal Cu(111) surface, only NN interactions are needed to account adequately for the experimental data [79,80]. In fact, for the NN model on a hexagonal grid, [81] found an exact and simple, albeit implicit, expression for the ECS. However, on such (111) surfaces (and basal planes of hcp crystals), lateral pair interactions alone cannot break the symmetry to produce a difference in energies between the two kinds of step edges, viz. {100} and {111} microfacets (A and B steps, respectively, with no relation to types A and B!). The simplest viable explanation is an orientation-dependent trio interaction; calculations of such energies support this idea [79,80].

Strictly speaking, of course, there should be no 2D facet (straight edge) and accompanying sharp edges (corners) at $T > 0$ (see Refs [82–85] and references therein) since that would imply 1D long-range order, which should not occur for short-range interactions. Measurements of islands at low temperatures show edges that appear to be facets and satisfy Wulff corollaries such as that the ratio of the distances of two unlike facets from the center equals the ratio of their f_i [86]. Thus, this issue is often just mentioned in passing [87] or even ignored. On the other hand, sophisticated approximations for $f_i(\theta)$ for the 2D Ising model, including NNN bonds, have been developed, e.g., Ref. [88], allowing numerical tests of the degree to which the ECS deviates from a polygon near corners of the latter. One can also gauge the length scale at which deviations from a straight edge come into play by using that the probability per atom along the edge for a kink to occur is essentially the Boltzmann factor associated with the energy to create the kink [89].

Especially for heteroepitaxial island systems (when the island consists of a different species from the substrate), strain plays an important if not dominant role. Such systems have been investigated, e.g., by Liu [90], who points out that for such systems the shape does not simply scale with λ, presumably implying the involvement of some new length scale[s]. A dramatic manifestation of strain effects is the island shape transition of Cu on Ni(001), which changes from compact to ramified as island size increases [91]. For small islands, additional quantum-size and other effects lead to favored island sizes (magic numbers).

5.5 Vicinal Surfaces–Entrée to Rough Regions Near Facets

In the rough regions, the ECS is a vicinal surface of gradually evolving orientation. To the extent that a local region has a particular orientation, it can be approximated as an infinite vicinal surface. The direction perpendicular to the terraces (which are densely packed facets) is typically called $\hat{\mathbf{z}}$. In "Maryland notation" (cf. Section 5.2.1) the normal to the vicinal surface lies in the x–z plane, and the distance ℓ between steps is measured along $\hat{x}$, while the steps run along the $\hat{y}$ direction. In the simplest and usual approximation, the repulsions between adjacent steps arise from two sources: an entropic or

steric interaction due to the physical condition that the steps cannot cross, since overhangs cannot occur in nature. The second comes from elastic dipole moments due to local atomic relaxation around each step, leading to frustrated lateral relaxation of atoms on the terrace plane between two steps. Both interactions are $\propto 1/\ell^2$.

The details of the distribution $\breve{P}(\ell)$ of spacings between steps have been reviewed in many works [60,92,93,97]. The average step separation $\langle\ell\rangle$ is the only characteristic length in the $\hat{x}$ direction. N.B., $\langle\ell\rangle$ need not be a multiple of, or even simply related to, the substrate lattice spacing. Therefore, we consider $P(s) = \langle\ell\rangle^{-1}\breve{P}(\ell)$, where $s \equiv \ell/\langle\ell\rangle$, a dimensionless length. For a "perfect" cleaved crystal, $P(s)$ is just a spike $\delta(s-1)$. For straight steps placed randomly at any position with probability $1/\langle\ell\rangle$, $P(s)$ is a Poisson distribution $\exp(-s)$. Actual steps do meander, as one can study most simply in a terrace step kink (TSK) model. In this model, the only excitations are kinks (with energy ϵ) along the step. (This is a good approximation at low temperature T since adatoms or vacancies on the terrace cost several ϵ_1 ($4\epsilon_1$ in the case of a simple-cubic lattice). The entropic repulsion due to steps meandering dramatically decreases the probability of finding adjacent steps at $\ell \ll \langle\ell\rangle$. To preserve the mean of one, $P(s)$ must also be smaller than $\exp(-s)$ for large s.

If there is an additional energetic repulsion A/ℓ^2, the magnitude of the step meandering will decrease, narrowing $P(s)$. As $A \to \infty$, the width approaches 0 ($P(s) \to \delta(s-1)$, the result for perfect crystals). We emphasize that the energetic and entropic interactions do not simply add. In particular, there is no negative (attractive) value of A at which the two cancel each other (cf. Eqn (5.30) below.) Thus, for strong repulsions, steps rarely come close, so the entropic interaction plays a smaller role, while for $A < 0$, the entropic contribution increases, as illustrated in Figure 5.16 and explicated below. We emphasize that the potentials of both interactions decay as ℓ^{-2} (cf. Eqn (5.27)

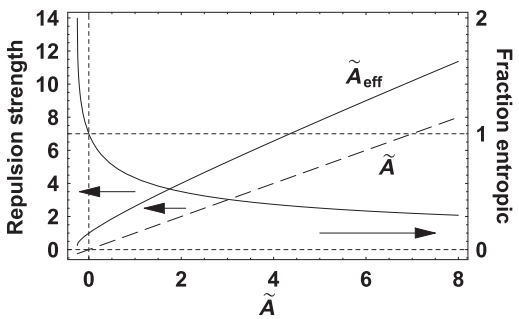


FIGURE 5.16 Illustration of how entropic repulsion and energetic interactions combine, plotted versus the dimensionless energetic interaction strength $\tilde{A} \equiv A\tilde{\beta}/(k_B T)^2$. The dashed straight line is just $\tilde{A}$. The solid curve above it is the combined entropic and energetic interactions, labeled $\tilde{A}_{\text{eff}}$ for reasons explained below. The difference between the two curves at any value of the abscissa is the dimensionless entropic repulsion for that $\tilde{A}$. The decreasing curve, scaled on the right ordinate, is the ratio of this entropic repulsion to the total dimensionless repulsion $\tilde{A}_{\text{eff}}$. It falls monotonically with $\tilde{A}$, passing through unity at $\tilde{A} = 0$. See the discussion accompanying Eqn (5.26) for more information and explicit expressions for the curves. *From Ref. [92].*

below), in contrast to some claims in the literature (in papers analyzing ECS exponents) that entropic interactions are short range while energetic ones are long range.

Investigation of the interaction between steps has been reviewed well in several places [60,94–97]. The earliest studies seeking to extract A from terrace-width distributions (TWDs) used the mean-fieldlike Gruber–Mullins [96] approximation, in which a single active step fluctuates between two fixed straight steps $2\langle\ell\rangle$ apart. Then the energy associating with the fluctuations $x(y,t)$ is

$$\Delta\mathcal{E} = -\beta(0)L_y + \int_0^{L_y} \beta(\theta(y))\sqrt{1+\left(\frac{\partial x}{\partial y}\right)^2}\,dy, \tag{5.18}$$

where L_y is the size of the system along the mean step direction (i.e., the step length with no kinks). We expand $\beta(\theta)$ as the Taylor series $\beta(0)+\beta'(0)\theta+{}^1\!/\!2\beta''(0)\theta^2$ and recognize that the length of the line segment has increased from dy to $dy/\cos\theta \approx dy(1+{}^1\!/_2\,\theta^2)$. For close-packed steps, for which $\beta'(0)=0$, it is well known that (using $\theta \approx \tan\theta = \partial x/\partial y$)

$$\Delta\mathcal{E} \approx \frac{\tilde{\beta}(0)}{2}\int_0^{L_y}\left(\frac{\partial x}{\partial y}\right)^2 dy, \quad \tilde{\beta}(0)\equiv\beta(0)+\beta''(0), \tag{5.19}$$

where $\tilde{\beta}$ is the step stiffness [97]. N.B., the stiffness $\tilde{\beta}(\theta)$ has the same definition for steps with arbitrary in-plane orientation—for which $\beta'(\theta)\neq 0$—because to create such steps, one must apply a "torque" [98] which exactly cancels $\beta'(\theta)$. (See Refs [88,99] for a more formal proof.)

Since $x(y)$ is taken to be a single-valued function that is defined over the whole domain of y, the 2D configuration of the step can be viewed as the worldline of a particle in 1D by recognizing y as a timelike variable. Since the steps cannot cross, these particles can be described as spinless fermions in 1D, as pointed out first by de Gennes [100] in a study of polymers in 2D [220]. Thus, this problem can be mapped into the Schrödinger equation in 1D: $\partial x/\partial y$ in Eqn (5.19) becomes $\partial x/\partial t$, with the form of a velocity, with the stiffness playing the role of an inertial mass. This correspondence also applies to domain walls of adatoms on densely covered crystal surfaces, since these walls have many of the same properties as steps. Indeed, there is a close correspondence between the phase transition at smooth edges of the ECS and the commensurate-incommensurate phase transitions of such overlayer systems, with the rough region of the ECS corresponding to the incommensurate regions and the local slope related to the incommensurability [101–105]. Jayaprakash et al. [37] provide the details of the mapping from a TSK model to the fermion picture, complete with fermion creation and annihilation operators.

In the Gruber–Mullins [96] approximation, a step with no energetic interactions becomes a particle in a 1D infinite-barrier well of width $2\langle\ell\rangle$, with well-known groundstate properties:

$$\psi_0(\ell)\propto\sin\left(\frac{\pi\ell}{2\langle\ell\rangle}\right);\ P(s)=\sin^2\left(\frac{\pi s}{2}\right);\ E_0=\frac{(\pi k_B T)^2}{8\tilde{\beta}\langle\ell\rangle^2} \tag{5.20}$$

Thus, it is the kinetic energy of the ground state in the quantum model that corresponds to the entropic repulsion (per length) of the step. In the exact solution for the free energy expansion of the ECS [106], the numerical coefficient in the corresponding term is 1/6 rather than 1/8. Note that $P(s)$ peaks at $s = 1$ and vanishes for $s \geq 2$.

Suppose, next, that there is an energetic repulsion $U(\ell) = A/\ell^2$ between steps. In the 1D Schrödinger equation, the prefactor of $-\partial^2\psi(\ell)/\partial\ell^2$ is $(k_BT)^2/2\tilde{\beta}$, with the thermal energy k_BT replacing $\hbar$. (Like the repulsions, this term has units ℓ^{-2}.) Hence, A only enters the problem in the dimensionless combination $\tilde{A} \equiv A\tilde{\beta}/(k_BT)^2$ [107]. In the Gruber–Mullins picture, the potential (per length) experienced by the single active particle is (with $\breve{\ell} \equiv \ell - \langle\ell\rangle$)

$$\tilde{U}\left(\breve{\ell}\right) = \frac{\tilde{A}}{\left(\breve{\ell} - \langle\ell\rangle\right)^2} + \frac{\tilde{A}}{\left(\breve{\ell} + \langle\ell\rangle\right)^2} = \frac{2\tilde{A}}{\langle\ell\rangle^2} + \frac{6\tilde{A}\breve{\ell}^2}{\langle\ell\rangle^4} + \mathcal{O}\left(\frac{\tilde{A}\breve{\ell}^4}{\langle\ell\rangle^6}\right) \tag{5.21}$$

The first term is just a constant shift in the energy. For $\tilde{A}$ sufficiently large, the particle is confined to a region $\left|\breve{\ell}\right| \ll \langle\ell\rangle$, so that we can neglect the fixed walls and the quartic term, reducing the problem to the familiar simple harmonic oscillator, with the solution:

$$\psi_0(\ell) \propto \mathrm{e}^{-\breve{\ell}^2/4w^2};\ P_{\mathrm{G}}(s) \equiv \frac{1}{\sigma_G\sqrt{2\pi}} \exp\left[-\frac{(s-1)^2}{2\sigma_G^2}\right] \tag{5.22}$$

where $\sigma_G = (48\tilde{A})^{-1/4}$ and $w = \sigma_G\langle\ell\rangle$.

For $\tilde{A}$ of 0 or 2, the TWD can be computed exactly (See below). For these cases, Eqns (5.20) and (5.22), respectively, provide serviceable approximations. It is Eqn (5.22) that is prescribed for analyzing TWDs in the most-cited resource on vicinal surfaces [58]. Indeed, it formed the basis of initial successful analyses of experimental scanning tunneling microscopy (STM) data [108]. However, it has some notable shortcomings. Perhaps most obviously, it is useless for small but not vanishing $\tilde{A}$, for which the TWD is highly skewed, not resembling a Gaussian, and the peak, correspondingly, is significantly below the mean spacing. For large values of $\tilde{A}$, it significantly underestimates the variance or, equivalently, the value of $\tilde{A}$ one extracts from the experimental TWD width [109]: in the Gruber–Mullins approximation the TWD variance is the same as that of the active step, since the neighboring step is straight. For large $\tilde{A}$, the fluctuations of the individual steps on an actual vicinal surface become relatively independent, so the variance of the TWD is the *sum* of the variance of each, i.e., twice the step variance. Given the great (quartic) sensitivity of $\tilde{A}$ to the TWD width, this is problematic. As experimentalists acquired more high-quality TWD data, other approximation schemes were proposed, all producing Gaussian distributions with widths $\propto \tilde{A}^{-1/4}$, but with proportionality constants notably larger than $48^{-1/4} = 0.38$.

For the "free-fermion" ($\tilde{A} = 0$) case, [110] developed a sequence of analytic approximants to the exact but formidable expression [111,112] for $P(s)$. They, as well as a slightly earlier paper [113], draw the analogy between the TWD of vicinal surfaces and

the distribution of spacings between interacting (spinless) fermions on a ring, the Calogero–Sutherland model [113,114], which, in turn for three particular values of the interaction—in one case repulsive ($\tilde{A} = 2$), in another attractive ($\tilde{A} = -1/4$), and lastly the free-fermion case ($\tilde{A} = 0$)—could be solved exactly by connecting to random matrix theory [92,111,115]; Figure 5.5 of Ref. [117] depicts the three resulting TWDs.

These three cases can be well described by the Wigner surmise, for which there are many excellent reviews [111,117,118]. Explicitly, for $\varrho = 1$, 2, and 4:

$$P_\varrho(s) = a_\varrho s^\varrho \exp\left(-b_\varrho s^2\right), \tag{5.23}$$

where the subscript of P refers to the exponent of s. In random matrix literature, the exponent of s, viz. 1, 2, or 4, is called β, due to an analogy with inverse temperature in one justification. However, to avoid possible confusion with the step free energy per length β or the stiffness $\tilde{\beta}$ for vicinal surfaces, I have sometimes named it instead by the Greek symbol that looked most similar, ϱ, and do so in this chapter. The constants b_ϱ, which fixes its mean at unity, and a_ϱ, which normalizes $P(s)$, are

$$b_\varrho = \left[\frac{\Gamma\left(\frac{\varrho+2}{2}\right)}{\Gamma\left(\frac{\varrho+1}{2}\right)}\right]^2 \qquad a_\varrho = \frac{2\left[\Gamma\left(\frac{\varrho+2}{2}\right)\right]^{\varrho+1}}{\left[\Gamma\left(\frac{\varrho+1}{2}\right)\right]^{\varrho+2}} = \frac{2b_\varrho^{(\varrho+1)/2}}{\Gamma\left(\frac{\varrho+1}{2}\right)} \tag{5.24}$$

Specifically, $b_\varrho = \pi/4$, $4/\pi$, and $64/9\pi$, respectively, while $a_\varrho = \pi/2$, $32/\pi^2$, and $(64/9\pi)^3$, respectively.

As seen most clearly by explicit plots, e.g., Figure 4.2(a) of Haake's text [118], $P_1(s)$, $P_2(s)$, and $P_4(s)$ are excellent approximations of the exact results for orthogonal, unitary, and symplectic ensembles, respectively, and these simple expressions are routinely used when confronting experimental data in a broad range of physical problems [118,119]. (The agreement is particularly outstanding for $P_2(s)$ and $P_4(s)$, which are the germane cases for vicinal surfaces, significantly better than any other approximation [120].)

Thus, the Calogero–Sutherland model provides a connection between random matrix theory, notably the Wigner surmise, and the distribution of spacings between fermions in 1D interacting with dimensionless strength $\tilde{A}$. Specifically:

$$\tilde{A} = \frac{\varrho}{2}\left(\frac{\varrho}{2} - 1\right) \quad \Leftrightarrow \quad \varrho = 1 + \sqrt{1 + 4\tilde{A}}. \tag{5.25}$$

For an arbitrary system, there is no reason that $\tilde{A}$ should take on one of the three special values. Therefore, we have used Eqn (5.28) for arbitrary ϱ or $\tilde{A}$, even though there is no symmetry-based justification of distribution based on the Wigner surmise of Eqn (5.26), and refer hereafter to this formula, Eqns (5.26, 7.27), as the generalized Wigner distribution (GWD). Arguably the most convincing argument is a comparison of the predicted variance with numerical data generated from Monte Carlo simulations. See Ref. [92] for further discussion.

There are several alternative approximations that lead to a description of the TWD as a Gaussian [109]; in particular, focus on the limit of large ϱ, neglecting the entropic

interaction in that limit. The variance $\sigma^2 \propto \tilde{A}^{-1/2}$, the proportionality constant is 1.8 times that in the Gruber–Mullins case. This approximation is improved, especially for repulsions that are not extremely strong, by including the entropic interaction in an average way. This is done by replacing $\tilde{A}$ by

$$\tilde{A}_{\text{eff}} = \left(\frac{\varrho}{2}\right)^2 = \tilde{A} + \frac{\varrho}{2}. \tag{5.26}$$

Physically, $\tilde{A}_{\text{eff}}$ gives the full strength of the inverse-square repulsion between steps, i.e., the modification due to the inclusion of entropic interactions. Thus, in Eqn (5.1)

$$g(T) = \frac{(\pi k_B T)^2}{6h^3\tilde{\beta}}\tilde{A}_{\text{eff}} = \frac{(\pi k_B T)^2}{24h^3\tilde{\beta}}\left[1 + \sqrt{1 + 4\tilde{A}}\right]^2. \tag{5.27}$$

From Eqn (5.29) it is obvious that the contribution of the entropic interaction, viz. the difference between the total and the energetic interaction, as discussed in conjunction with Figure 5.16, is $\varrho/2$. Remarkably, the ratio of the entropic interaction to the total interaction is $(\varrho/2)/(\varrho/2)^2 = 2/\varrho$; this is the fractional contribution that is plotted in Figure 5.16.

5.6 Critical Behavior of Rough Regions Near Facets

5.6.1 Theory

Assuming (cf. Figure 5.17) $\hat{\mathbf{z}}$ the direction normal to the facet and (x_0,z_0) denote the facet edge, $z \sim z_0 - (x - x_0)^{\vartheta}$ for $x \geq x_0$. We show that the critical exponent ϑ[3] has the value 3/2 for the generic smooth edge described by Eqn (5.1) (with the notation of Eqn (5.13)):

$$f_p(p) = f_0 + Bp + gp^3 + cp^4. \tag{5.28}$$

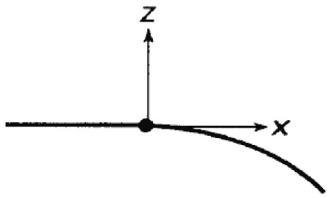


FIGURE 5.17 Critical behavior of the crystal shape near a smooth (second-order) edge, represented by the dot at (x_0,z_0). The temperature is lower than the roughening temperature of the facet orientation, so that the region to the left of the dot is flat. The curved region to the right of the dot correspond to a broad range of rough orientations. In the thermodynamic limit, the shape of the smoothly curved region near the edge is described by the power law $z \sim z_0 - (x - x_0)^{\vartheta}$. Away from the edge there are "corrections to scaling", i.e., higher order terms (cf. Eqn (5.33)). For an actual crystal of any finite size, there is "finite-size rounding" near the edge, which smooths the singular behavior. *Adapted from Ref. [120].*

[3]The conventional designation of this exponent is λ or θ. However, these Greek letters are the Lagrange multiplier of the ECS and the polar angle, respectively. Hence, we choose ϑ for this exponent.

Then we perform a Legendre transformation [55] as in Refs [125,126]; explicitly:

$$\frac{f_p(p)-\tilde{f}(\eta)}{p}=\left[\frac{df_p}{dp}\equiv\eta\right]=B+3gp^2+4cp^3 \tag{5.29}$$

Hence:

$$\tilde{f}(\eta)=f_0-2gp^3(\eta)-3cp^4(\eta) \tag{5.30}$$

But from Eqn (5.29):

$$p=\left(\frac{\eta-B}{3g}\right)^{1/2}\left[1-\frac{2c}{3g}\left(\frac{\eta-B}{3g}\right)^{1/2}+\ldots\right] \tag{5.31}$$

Inserting this into Eqn (5.30) gives

$$\tilde{f}(\eta)=f_0-2g\left(\frac{\eta-B}{3g}\right)^{3/2}+c\left(\frac{\eta-B}{3g}\right)^{2}+\mathcal{O}\left(\frac{\eta-B}{3g}\right)^{5/2} \tag{5.32}$$

for $\eta\geq B$ and $\tilde{f}(\eta)=f_0$ for $\eta\leq B$. (See Refs [9,120,122].) Note that this result is true not just for the free-fermion case but even when steps interact. Jayaprakash et al. [37] further show that the same ϑ obtains when the step–step interaction decreases with a power law in ℓ that is greater than 2. We identify $\tilde{f}(\eta)$ with $r(\widehat{\mathbf{h}})$, i.e., the magnetic-fieldlike variable discussed corresponds to the so-called Andreev field η. Writing $z_0=f_0/\lambda$ and $x_0=B/\lambda$, we find the shape profile

$$\begin{aligned}\frac{z(x)}{z_0}=1-2\left(\frac{f_0}{g}\right)^{1/2}\left(\frac{x-x_0}{z_0}\right)^{3/2}\\ +\frac{cf_0}{g^2}\left(\frac{x-x_0}{z_0}\right)^{2}+\mathcal{O}\left(\frac{x-x_0}{z_0}\right)^{5/2}\end{aligned} \tag{5.33}$$

Note that the edge position depends only on the step free energy B, not on the step repulsion strength; conversely, the coefficient of the leading $(x-x_0)^{3/2}$ term is independent of the step free energy but varies as the inverse root of the total step repulsion strength, i.e., as $g^{-1/2}$.

If, instead of Eqn (5.31), one adopts the phenomenological Landau theory of continuous phase transitions [121] and performs an analytic expansion of $f_p(p)$ in p [123, 124] (and truncate after a quadratic term f_2p^2), then a similar procedure leads $\vartheta=2$, which is often referred to as the "mean-field" value. This same value can be produced by quenched impurities, as shown explicitly for the equivalent commensurate-incommensurate transition by [125].

There are some other noteworthy results for the smooth edge. As the facet roughening temperature is approached from below, the facet radius shrinks like $\exp[-\pi^2T_R/4\{2\ln 2(T_R-T)\}^{1/2}]$ [122], in striking contrast to predictions by mean field theory. The previous discussion implicitly assumes that the path along x for which $\vartheta=3/2$ in Eqn (5.36) is normal to the facet edge. By mapping the crystal surface onto the asymmetric 6-vertex model, using its exact solution [53,54], and employing the Bethe ansatz to

expand the free energy close to the facet edge [127], find that $\vartheta = 3/2$ holds for any direction of approach along the rounded surface toward the edge, *except* along the tangential direction (the contour that is tangent to the facet edge at the point of contact x_0). In that special direction, they find the new critical exponent $\vartheta_y = 3$ (where the subscript y indicates the direction perpendicular to the edge normal, x [128]). Also, Akutsu and Akutsu [128] confirmed that this exact result was universally true for the Gruber–Mullins–Prokrovsky–Talapov free-energy expansion. (The Prokrovsky-Talapov argument was for the equivalent commensurate-incommensurate transition.) They also present numerical confirmation using their transfer-matrix method based on the product-wave-function renormalization group (PWFRG) [129,130]. Observing ϑ_y experimentally will clearly be difficult, perhaps impossible; the nature and breadth of crossover to this unique behavior has not, to the best of my knowledge, been published. A third result is that there is a jump (for $T < T_R$) in the curvature of the rounded part near the facet edge that has a universal value [106,131], distinct from the universal curvature jump of the ECS at T_R [122].

5.6.2 Experiment on Leads

Noteworthy initial experimental tests of $\vartheta = 3/2$ include direct measurements of the shape of equilibrated crystals of ^{4}He [132] and Pb [133]. As in most measurements of critical phenomena, but even harder here, is the identification of the critical point, in this case the value of x_0 at which rounding begins. Furthermore, as is evident from Eqn (5.36), there are corrections to scaling, so that the "pure" exponent 3/2 is seen only near the edge and a larger effective exponent will be found farther from the edge. For crystals as large as a few mm at temperatures in the range 0.7–1.1 K, ^{4}He $\vartheta = 1.55 \pm 0.06$ was found, agreeing excellently with the Prokrovsky–Talapov exponent. The early measurements near the close-packed (111) facets of Pb crystallites, at least two orders of magnitude smaller, were at least consistent with 3/2, stated conservatively as $\vartheta = 1.60 \pm 0.15$ after extensive analysis. Sáenz and García [134] proposed that in Eqn (5.31) there can be a quadratic term, say $f_2 p^2$ (but neglect the possibility of a quartic term). Carrying out the Legendre transformation then yields an expression with both $x - B$ and $(x - B + f_2^2/3g)^{3/2}$ terms, which they claim will lead to effective values of ϑ between 3/2 and 2. This approach provided a competing model for experimentalists to consider but in the end seems to have produced little fruit.

As seen in Figure 5.18, STM allows detailed measurement of micron-size crystal height contours and profiles at fixed azimuthal angles. By using STM to locate the initial step down from the facet, first done by Surnev et al. [135] for supported Pb crystallites, x_0 can be located independently and precisely. However, from the 1984 Heyraud–Métois experiment [133] it took almost two decades until the Bonzel group could fully confirm the $\vartheta = 3/2$ behavior for the smooth edges of Pb(111) in a painstaking study [137]. There were a number of noteworthy challenges. While the close-packed 2D network of spheres has six-fold symmetry, the top layer of a (111) facet of an fcc crystal (or of an (0001) facet

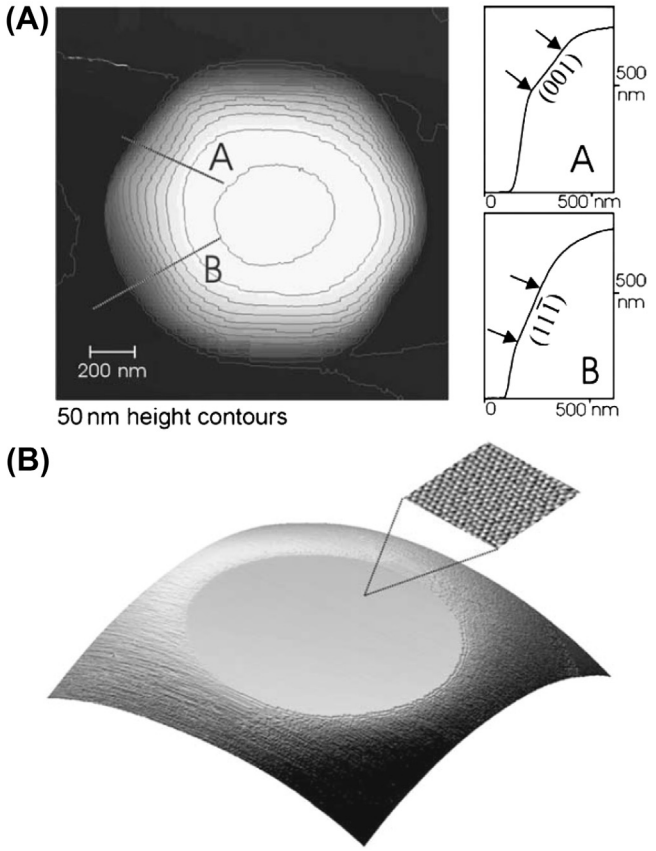


FIGURE 5.18 (A) Micron-size lead crystal (supported on Ru) imaged with a variable-temperature scanning tunneling microscopy at $T = 95$ °C. Annealing at $T = 95$ °C for 20 h allowed it to obtain its stable, regular shape. Lines marked A and B indicate location of profiles. Profile A crosses a (0 0 1)-side facet, while profile B a (1 1 1)-side facet. (B) 770 × 770 nm section of the top part of a Pb crystal. The insert shows a 5.3 × 5.3 nm area of the top facet, confirming its $(1\,1\,\bar{1})$-orientation. Both the main image and the insert were obtained at $T = 110$ °C. *From Ref. [141].*

of an hcp crystal) has only three-fold symmetry due to the symmetry-breaking role of the second layer. There are two dense straight step edges, called A and B, with {100} and {111} microfacets, respectively. In contrast to noble metals, for Pb there is a sizable (of order 10%) difference between their energies. Even more significant—when a large range of polar angles is used in the fitting—is the presence of small (compared to {111}) {112} facets for equilibration below 325 K. Due to the high atomic mobility of Pb that can lead to the formation of surface irregularities, Bonzel's group [135] worked close to room temperature. One then finds strong (three-fold) variation of ϑ with azimuthal angle, with ϑ oscillating between 1.4 and 1.7. With a higher annealing temperature of 383 K, [137] report the azimuthal averaged value $\vartheta = 1.487$ (but still with sizable oscillations of about ± 0.1); in a slightly earlier short report [137], they give a value $\vartheta = 1.47$ for annealing at

room temperature. Their attention shifted to deducing the strength of step–step repulsions by measuring g [138,139]. In the most recent review of the ECS of Pb [140], the authors rather tersely report that the Prokrovsky–Talapov value of 3/2 for ϑ characterizes the shape near the (111) facet and that imaging at elevated temperature is essential to get this result; most of their article relates to comparison of measured and theoretically calculated strengths of the step–step interactions.

Few other systems have been investigated in such detail. Using scanning electron microscopy (SEM) [142] the researchers considered In, which has a tetragonal structure, near a (111) facet. They analyzed the resulting photographs from two different crystals, viewed along two directions. For polar angles $0^\circ \leq \theta < \sim 5^\circ$ they find $\vartheta \approx 2$ while for $5^\circ \leq \theta \leq 15^\circ$ determine $\vartheta \approx 1.61$, concluding that in this window $\vartheta = 1.60 \pm 0.10$; the two ranges have notably different values of x_0. This group [143] also studied Si, equilibrated at 900° C, near a (111) facet. Many profiles were measured along a high-symmetry $\langle 111 \rangle$ zone of samples with various diameters of the order a few μm, over the range $3^\circ \leq \theta \leq 17^\circ$. The results are consistent with $\vartheta = 3/2$, with an uncertainty estimated at 6%. Finally, [144] studied large (several mm) spherical cuprous selenide (Cu_{2-x} Se) single crystals near a (111) facet. Study in this context of metal chalcogenide superionic conductors began some dozen years ago because, other than ^{4}He, they are the only materials having sub-cm size crystals with an ECS form that can be grown on a practical time scale (viz. over several days) because their high ionic and electronic conductivity enable fast bulk atomic transport. For $14.0^\circ \leq \theta \leq 17.1^\circ$ [144] find $\vartheta = 1.499 \pm 0.003$. (They also report that farther from the facet $\vartheta \approx 2.5$, consistent with the Andreev mean field scenario.)

5.6.3 Summary of Highlights of Novel Approach to Behavior Near Smooth Edges

Digressing somewhat, we note that Ferrari, Prähofer, and Spohn [145] found novel static scaling behavior of the equilibrium fluctuations of an atomic ledge bordering a crystalline facet surrounded by rough regions of the ECS in their examination of a 3D Ising corner (Figure 5.19). This boundary edge might be viewed as a "shoreline" since it is the edge of an islandlike region—the crystal facet—surrounded by a "sea" of steps [146].

Spohn and coworkers assume that there are no interactions between steps other than entropic, and accordingly the step configurations can be mapped to the worldlines of free spinless fermions, as in treatments of vicinal surfaces [37]. However, there is a key new feature that the step number operator is weighted by the step number, along with a Lagrange multiplier λ^{-1} associated with volume conservation of the crystallite. The asymmetry of this term leads to the novel behavior found by the researchers. They then derive an exact result for the step density and find that, near the shoreline:

$$\lim_{\lambda \to \infty} \lambda^{1/3} \rho_\lambda \left(\lambda^{1/3} x \right) = -x(\mathrm{Ai}(x))^2 + (\mathrm{Ai}'(x))^2, \tag{5.34}$$

where ρ_λ is the step density (for the particular value of λ).

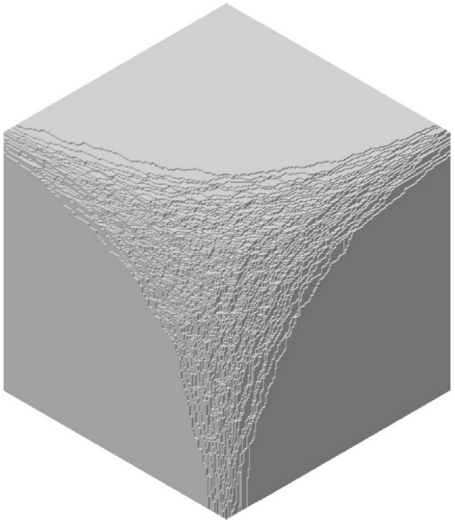


FIGURE 5.19 Simple-cubic crystal corner viewed from the {111} direction. *From Ref. [145].*

The presence of the Airy function Ai results from the asymmetric potential implicit in $\mathcal{H}_F$ and preordains exponents involving 1/3. The variance of the wandering of the shoreline, the top fermionic worldline in Figure 5.20 and denoted by b, is given by

$$\text{Var}\left[b_\lambda(t) - b_\lambda(0)\right] \cong \lambda^{2/3} g\left(\lambda^{-2/3} t\right) \tag{5.35}$$

where t is the fermionic "time" along the step; $g(s) \sim 2|s|$ for small s (diffusive meandering) and $\sim 1.6264 - 2/s^2$ for large s. 1.202… is Apery's constant and N is the number of atoms in the crystal. They find:

$$\text{Var}\left[b_\ell(\ell\tau + x) - b_\ell(\ell\tau)\right] \cong (\mathcal{A}\ell)^{2/3} g\left(\mathcal{A}^{1/3}\ell^{-2/3}x\right), \tag{5.36}$$

where $\mathcal{A} = (1/2)b''_\infty$. This leads to their central result that the width $w \sim \ell^{1/3}$, in contrast to the $\ell^{1/2}$ scaling of an isolated step or the boundary of a single-layer island and to the $\ln \ell$ scaling of a step on a vicinal surface, i.e., in a step train. Furthermore, the

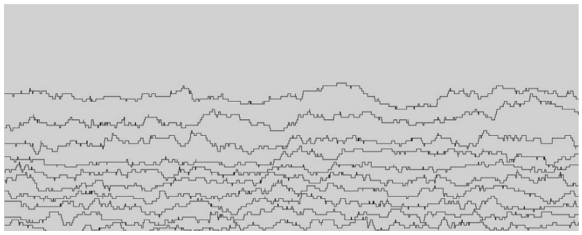


FIGURE 5.20 Snapshot of computed configurations of the top steps (those near a facet at the flattened side portion of a cylinder) for a terrace-step-kink (TSK) model with volume constraint. *From Ref. [145].*

fluctuations are non-Gaussian. The authors also show that near the shoreline, the deviation of the equilibrium crystal shape from the facet plane takes on the Pokrovsky–Talapov [101,104] form with $\vartheta = 3/2$.

From this seminal work, we could derive the dynamic exponents associated with this novel scaling and measure them with STM, as reviewed in Ref. [150].

5.7 Sharp Edges and First-Order Transitions—Examples and Issues

5.7.1 Sharp Edges Induced by Facet Reconstruction

Si near the (111) plane offers an easily understood entry into sharp edges [68,69]. As Si is cooled from high temperatures, the (111) plane in the "(1×1)" phase reconstructs into a (7×7) pattern [150] around 850 °C, to be denoted T_7 to distinguish it clearly from the other subscripted temperatures. (The notation "(1×1)" is intended to convey the idea that this phase differs considerably from a perfect (111) cleavage plane but has no superlattice periodicity.) For comparison, the melting temperature of Si is ∼1420 °C, and the T_R is estimated to be somewhat higher. As shown in Figure 5.21, above T_7, surfaces of

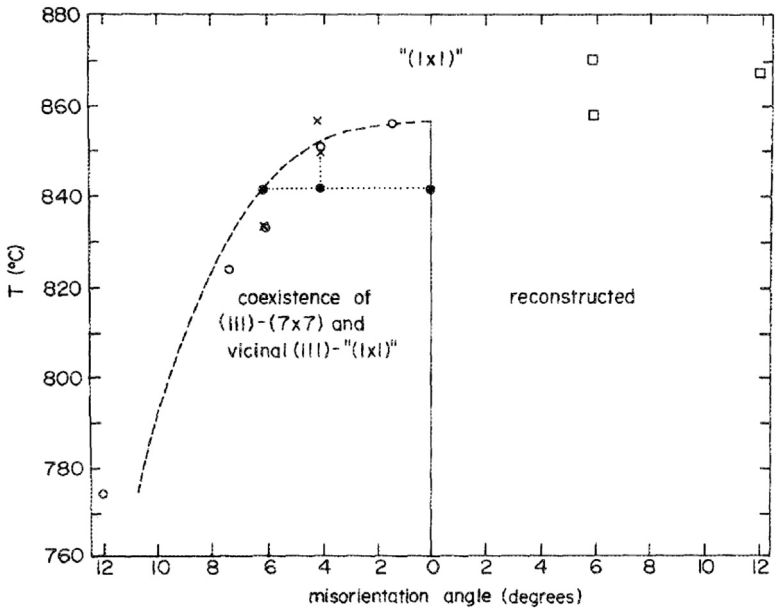


FIGURE 5.21 Summary of experimental results for vicinal Si (111) surface: ○ denotes the temperature at which faceting begins for surfaces misoriented toward the (110) direction, × the faceting temperatures for surfaces misoriented toward the $[11\bar{2}]$, and □ the temperatures at which the step structure of surfaces misoriented toward the $[\bar{1}\bar{1}2]$ direction change. The dashed line displays a fit of the $[\bar{1}10]$ data to Eqn (5.43). The dotted lines show how a four sample phase separates into the states denoted by ● as it is further cooled. *From Ref. [69].*

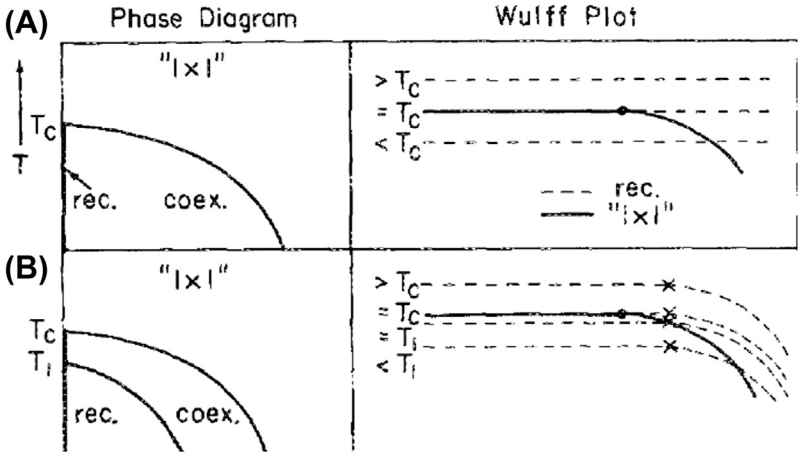


FIGURE 5.22 Wulff plots illustrating the effect of a reconstructive transition on the equilibrium crystal shape (ECS), and corresponding temperature-[mis]orientation phase diagrams. The solid curves represent the ECS with an unreconstructed ["(1 × 1)"] facet, while the dashed curves give the ECS with a reconstructed facet. As temperature decreases, the free energy of the reconstructed facet, relative to that of the unreconstructed facet, decreases. Below the transition temperature T_c (called T_7 in the text), the two shapes intersect, giving a "net" ECS that is the inner envelope of the two. The phase diagram shows regions where all orientations $\tan\theta$ (or $\widehat{\mathbf{m}}$) are allowed for the unreconstructed crystal ["(1 × 1)"], regions of phase separation (labeled "coex."), and regions where the reconstruction (labeled "rec.") is allowed for ranges of orientation. The relative size of the reconstructed and unreconstructed facets depends on the free energy to create a step on the reconstructed (111) face, compared to its unreconstructed counterpart: (A) the behavior for extremely large energy to create steps on the (7 × 7) terrace and (A) a smaller such energy. Solid circles mark the sharp edge at the temperature at which the crystal shapes cross. Crosses show the intersection of the facet and the curved part (i.e., the smooth edge) of the crystal shape for the reconstructed phase. *From Ref. [69].*

all orientations are allowed and are unreconstructed. At T_7, a surface in the (111) direction reconstructs but all other orientations are allowed and are unreconstructed. Below T_7, surfaces misoriented toward $[\bar{1}\,\bar{1}2]$ remain stable during cooling (although the step structure changes). On the other hand, on surfaces misoriented toward $[\bar{1}10]$ and $[11\bar{2}]$, the temperature at which the (7 × 7) occurs decreases with increasing misorientation angle $\widehat{\mathbf{m}}$. Furthermore, just as the (7 × 7) appears, the surface begins to separate into two phases, one a perfectly oriented (7 × 7) plane $\widehat{\mathbf{m}} = 0$ and the second an unreconstructed phase with a misorientation greater than that at higher temperature. As the temperature further decreases, the misorientation of the unreconstructed phase increases. Figure 5.21 depicts this scenario with solid circles and dotted lines for a 4° misoriented sample at 840 °C. This behavior translates into the formation of a sharp edge on the ECS between a flat (7 × 7) line and a rounded "(1 × 1)" curve.

To explain this behavior, one coplots the ECS for the two phases, as in Figure 5.22 [69]. The free energy to create a step is greater in the (7 × 7) than in the "(1 × 1)" phase. In the top panels (A), the step energy for the (7 × 7) is taken as infinite, i.e., much larger than that of the "(1 × 1)" phase, so its ECS never rounds. At T_7 (T_c in the figure), the free energies per area f_0 of the two facets are equal, call them f_7 and f_1, with associated

energies u_7 and u_1 and entropies s_7 and s_1 for the (7×7) and "(1×1)" phases, respectively, near T_7. Then $T_7 = (u_1 - u_7)/(s_1 - s_7)$ and, assuming the internal energies and entropies are insensitive to temperature:

$$f_1 - f_7 = (T_7 - T)(s_1 - s_7). \tag{5.37}$$

Since $s_1 > s_7$ because the (7×7) phase is so highly ordered, we find that $f_1 - f_7 > 0$ below T_7, as illustrated in Figure 5.22. Making connection to thermodynamics, we identify

$$\frac{L}{T_7} = (s_1 - s_7)_{T_7} = \left(\frac{\partial f_7}{\partial T}\right)_{T_7} - \left(\frac{\partial f_1}{\partial T}\right)_{T_7} \tag{5.38}$$

where L is the latent heat of the first-order reconstruction transition.

Corresponding to the minimum of a free energy as discussed earlier, the ECS of the system will be the inner envelope of the dashed and solid traces: a flat (7×7) facet along the dashed line out to the point of intersection, the sharp edge, beyond which it is "(1×1)" with continuously varying orientation. If one tries to construct a surface with a smaller misorientation, it will phase separate into flat (7×7) regions and vicinal unreconstructed regions with the orientation at the curved (rough) side of the sharp edge. Cf. Figure 5.23.

Using the leading term in Eqn (5.35) or (5.36), we can estimate the slope of the coexisting vicinal region and its dependence on temperature[4]: First we locate the sharp edge (recognizing f_0 as f_1 and z_0 as z_1) by noting

$$\begin{aligned} z_7 &= z_1 - 2(\lambda/g)^{1/2}(x - x_0)^{3/2} \\ (T_7 - T)\Delta s \approx (f_1 - f_7)_T &= \lambda^{3/2} g^{-1/2}(x - x_0)^{3/2} \end{aligned} \tag{5.39}$$

Since the slope m there is $-3\lambda(\lambda/g)^{1/2}(x - x_0)^{1/2}$, the temperature dependence of the slope is

$$m = -3\left(\frac{L}{2g}\right)^{1/3}\left(1 - \frac{T}{T_7}\right)^{1/3} \tag{5.40}$$

If the step free energy of the reconstructed phase were only modestly greater than that of the "(1×1)", then, as shown in the second panel in Figure 5.22, the previous high-T behavior obtains only down to the temperature T_1 at which the "(1×1)" curve intersects the (7×7) curve at its [smooth] edge. For $T < T_1$ the sharp edge associated with the interior of the curves is between a misoriented "(1×1)" phase and a differently misoriented (7×7) phase, so that it is these two which coexist. All orientations with smaller misorientation angles than this (7×7) plane are also allowed, so that the forbidden or coexistence regime has the depicted slivered crescent shape. Some other, but physically improbable, scenarios are also discussed by Bartelt et al. [69]. Phaneuf and Williams [68] show (their Figure 3) the LEED-beam splitting for a surface misoriented by 6.4° is $\propto (T_7 - T)^{1/3}$ once the surface is cooled below the temperature (which is $< T_7$) when this orientation becomes unstable to phase separation; however, by changing the

[4]There are some minor differences in prefactors from Ref. [69].

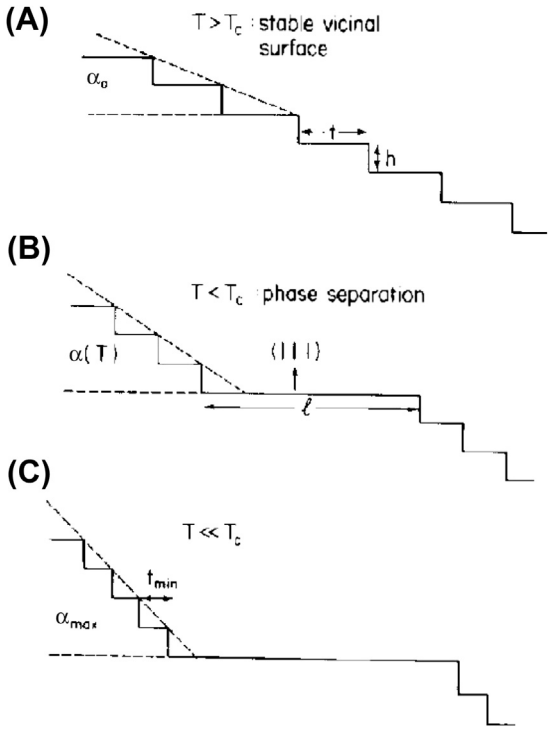


FIGURE 5.23 Microscopic view of what happens to a misoriented surface in Figure 5.22 as temperature decreases. (A) At high temperature, the Si(lll) vicinal surface is a single, uniform phase. Initial terrace widths t are typically a few nm, as determined by the net angle of miscut α_0 (i.e., θ_0), and the step-height h, which is one interplanar spacing (~0.31 nm). (B) Below the (7 × 7) reconstruction temperature (~850 °C), the steps cluster to form a new surface of misorientation angle $a(T)$ (i.e., θ). A facet of (111) orientation with (7 × 7) reconstruction forms simultaneously. The width of the (111) facet, ℓ, is larger than the experimentally resolvable width of 500 Å. (C) Well below the transition, the step separation reaches a minimum distance, t_{min} ~ 1 nm. No further narrowing occurs, perhaps because surface diffusion is too slow $T \leq 600$ °C. *From Ref. [3].*

range of fitting, they could also obtain agreement with $(T_7 - T)^{1/2}$, i.e., $\vartheta = 2$. With high-resolution LEED, [151] conclude that the exponent $\overline{\beta} \equiv (\vartheta - 1)/\vartheta = 0.33 \pm 0.05$ (i.e., that $\vartheta = 3/2$. The result does depend somewhat on what thermal range is used in the fit, but they can decisively rule out the mean-field value $\vartheta = 2$. Williams et al. [152] give a more general discussion of vicinal Si, with treatment of azimuthal in addition to polar misorientations. In contrast, synchrotron X-ray scattering experiments by Noh et al. [153,154] report the much larger $\vartheta = 2.3^{+0.8}_{-0.3}$. However, subsequent synchrotron X-ray scattering experiments by [155] obtain a decent fit of data with $\vartheta = 3/2$ and a best fit with $\vartheta = 1.75$ (i.e., $\overline{\beta} = 0.43 \pm 0.07$). (They also report that above 1159 K, the surface exists as a single, logarithmically rough phase.) The origin of the curious value of ϑ in the Noh et al. experiments is not clear. It would be possible to attribute the behavior to impurities, but there is no evidence to support this excuse, and indeed for the analogous behavior near the reconstructing (331) facet of Si (but perhaps a different sample), Noh

et al. [156] found $\vartheta = 1.47 \pm 0.1$. It is worth noting that extracting information from X-ray scattering from vicinal surfaces requires great sophistication (cf. the extensive discussion in Ref. [157]) and attention to the size of the coherence length relative to the size of the scattering region [158], as for other diffraction experiments.

Similar effects to reconstruction (viz. the change in f_0) could be caused by adsorption of impurities on the facet [159]. Some examples are given in a review by Somorjai and Van Hove [160]. In small crystals of dilute Pb-Bi-Ni alloys, cosegregation of Bi-Ni to the surface has a similar effect of reversibly changing the crystal structure to form {112} and {110} facets [161]. There is no attempt to scrutinize the ECS to extract an estimate of ϑ. Meltzman et al. [162] considered the ECS of Ni on a sapphire support, noting that, unlike most fcc crystals, it exhibits a faceted shape even with few or no impurities, viz. with {111}, {100}, and {110} facets; {135} and {138} emerged at low oxygen pressure and additionally {012} and {013} at higher pressure.

The phase diagram of Pt(001), shown in Figure 5.24 and studied [163] using synchrotron X-ray scattering, at first seems similar to that of Si near (111) [223,224,225], albeit with more intricate magic phases with azimuthal rotations at lower temperatures, stabilized by near commensurability of the period of their reconstruction and the separation of their constituent steps. In the temperature-misorientation (surface slope) phase diagram, shown in Figure 5.24, the (001) facet undergoes a hexagonal reconstruction at $T_6 = 1820$ K (well below the bulk melting temperature of 2045 K). For samples misoriented from the (100) direction (which are stable at high temperature), there is coexistence between flat reconstructed Pt(001) and a rough phase more highly

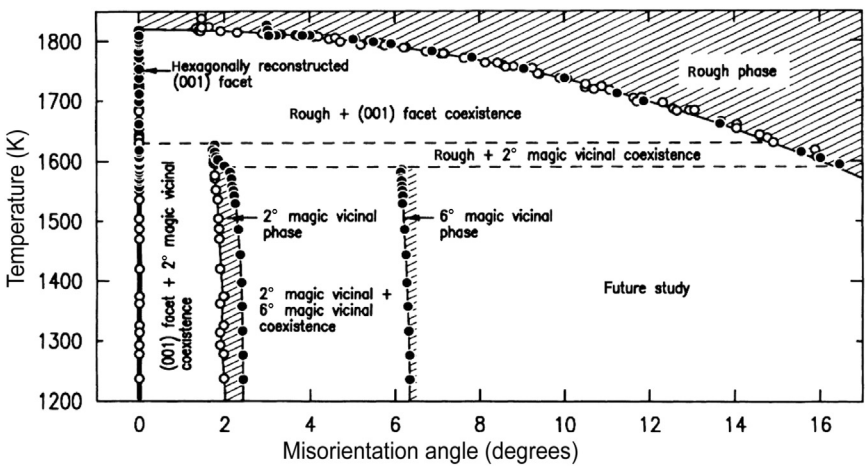


FIGURE 5.24 Orientational phase diagram of vicinal Pt (001) misoriented toward the [110] direction. Single-phase regions are hatched, and two-phase coexistence regions are unhatched. Solid lines are boundaries between two phases. Dashed lines mark triple points. Open circles show misorientation angles measured for a sample miscut by 1.4° toward the [110] direction, while solid circles show tilt angles measured for a sample miscut by 3.0°. *From Ref. [163].*

misoriented than it was at high temperature, with a misorientation that increases as temperature decreases. However, they find $\overline{\beta} = 0.49 \pm 0.05$, or $\vartheta = 1.96$, consistent with mean field and inconsistent with $\overline{\beta} = 1/3$ or $\vartheta = 3/2$ of Prokrovsky–Talapov. The source of this mean-field exponent is that in this case the (001) orientation is rough above T_6. Hence, in Eqn (5.31) B vanishes, leaving the expansion appropriate to rough orientations. Proceeding as before, Eqn (5.35) becomes

$$f_p(p) = f_0 + Dp^2 \Rightarrow \tilde{f}(\eta) = f_0 - \eta^2 \Big/ 4D, \tag{5.41}$$

where the result for $\tilde{f}(\eta)$ is reached by proceeding as before to reach the modification of Eqn (5.35). Thus, there is no smooth edge take-off point (no shoreline) in the equivalent of Figure 5.22, and one finds the reported exponent ϑ near 2.

The effect of reactive and nonreactive gases metal catalysts has long been of interest [211]. Various groups investigated adsorbate-induced faceting. Walko and Robinson [164] considered the oxygen-induced faceting of Cu(115) into O/Cu(104) facets, using Wulff constructions to explain their observations. The researchers found three temperature regimes with qualitatively different faceting processes. Szczepkowicz et al. [165] studied the formation of {211} facets by depositing oxygen and paladium on tungsten, both on (111) facets and on soherical crystals. While the shape of the facets is different for flat and curved surfaces, the distance between parallel facet edges is comparable, although the area of a typical facet on a curved surface is an order of magnitude greater. There is considerable information about facet sizes, width of the facet-size distribution, and surface rms roughness.

For 2D structures on surfaces, edge decoration can change the shape of the islands. A well-documented example is Pt on Pt(111). As little as 10^{-3} ml of CO produces a 60° rotation of the triangular islands by changing the balance of the edge free energies of the two different kinds of steps forming the island periphery [166]. Stasevich et al. [167] showed how decoration of single-layer Ag islands on Ag(111) by a single-strand "necklace" of C_{60} dramatically changes the shape from hexagonal to circular. With lattice-gas modeling combined with STM measurements, they could estimate the strength of C_{60}-Ag and C_{60}-C_{60} attractions. Generalizations to decoration on systems with other symmetries is also discussed [167].

5.8 Gold–Prototype or Anomaly of Attractive Step–Step Interaction?

Much as ^{4}He and Pb are the prototypical materials with smooth edges, Au is perhaps the prime example of a surface with sharp edges, around (1111) and (100) facets (cf. e.g., Ref. [1]). Care must be taken to ensure that the surface is not contaminated by atoms (typically C) from the supporting substrate [168]. (See similar comments by Handwerker et al. [169] for ceramics, which have a rich set of ECS possibilities.) To describe these systems phenomenologically, the projected free energy expansion in Eqn (5.1) requires a negative term to generate a region with negative curvature, as in Figure 5.7, so that the

two orientations joined by the Maxwell double-tangent construction correspond to the two sides of the sharp edge. Thus, for sharp edges around facets, the more-left minimum must be in the high-symmetry facet direction.

In a mean field-based approach, Wang and Wynblatt [168] included a negative quadratic term, with questionable physical basis. Emundts et al. [170] instead took the step–step interaction to be attractive ($g < 0$) in Eqn (5.1). Then, proceeding as above, they find

$$x_0 = \frac{1}{\lambda}\left[B - \frac{4}{27}|g|\left(\frac{g}{c}\right)^2\right], \quad p_c = \frac{2|g|}{3c}, \tag{5.42}$$

where p_c is the tangent of the facet contact angle. Note that both the shift in the facet edge from B and the contact slope increase with $|g|/c$. Emundts et al. [170] obtain estimates of the key energy parameters in the expansion for the sharp edges of both the (111) and (100) facets. They also investigate whether it is the lowering of the facet free energy f_0 that brings about the sharp edges, in the manner of the case of Si(111) discussed above. After reporting the presence of standard step–step repulsions (leading to narrowing of the TWD) in experiments on flame-annealed gold, Shimoni et al. [171] then attribute to some effective long-range attraction—with undetermined dependence on ℓ—the (nonequilibrium) movement of single steps toward step bunches whose steps are oriented along the high-symmetry $\langle 110 \rangle$.

Is it possible to find a generic long-range attractive $A\ell^{-2}$ step–step interaction ($A < 0$) for metals and elemental semiconductors (where there is no electrostatic attraction between oppositely charged atoms)? Several theoretical attempts have only been able to find such attractions when there is significant alternation between "even" and "odd" layered steps. Redfield and Zangwill consider whether surface relaxation can produce such an attraction, pointing out a flaw in an earlier analysis assuming a rigid relaxation by noting that for large step separations, the relaxation must return to its value for the terrace orientation. Since atomic displacements fall off inversely with distance from a step, the contribution to the step interaction can at most go like ℓ^{-2} and tend to mitigate the combined entropic and elastic repulsion. They argue that this nonlinear effect is likely to be small, at least for metals. It is conceivable that on an elastically highly anisotropic surface, the elastic interaction might not be repulsive in special directions, although I am not aware of any concrete examples.

By observing that the elastic field mediating the interaction between steps is that of a dipole applied on a stepped rather than on a flat surface, Kukta et al. [172] deduce a correction to the ℓ^{-2} behavior of the Marchenko–Parshin [41] formula that scales as $\ell^{-3}\ln\ell$. Using what was then a state-of-the-art semiempirical potential, the embedded atom method (EAM) [228], the authors find that this can lead to attractive interactions at intermediate values of ℓ. However, their "roughness correction" term exists only when the two steps have unlike orientations (i.e., one up and one down, such as on opposite ends of a monolayer island or pit). For the like-oriented steps of a vicinal surface or near a facet edge, the correction term vanishes. The oft-cited paper then invokes three-step

interactions, which are said to have the same size as the correction term, as a way to achieve attractive interactions. Although the authors discuss how this idea relates to the interaction between an isolated step and a step bunch, they do not provide the explicit form of the threefold interaction; their promise that it will be "presented elsewhere" has not, to the best of my searching, ever been fulfilled. Prévot and Croset [173] revisited elastic interactions between steps on vicinals and found that with a buried-dipole model (rather than the surface-dipole picture of Marchenko and Parshin), they could achieve "remarkable agreement" with molecular dynamics simulations for vicinals to Cu and Pt (001) and (111), for which data is fit by $E_2^{MD}\ell^{-2} + E_3^{MD}\ell^{-3}$. The tabulated values of E_2^{MD} indeed agree well with their computed results for their improved elastic model, which includes the strong dependence of the interaction energy on the force direction. While there is barely any discussion of E_3, plots of the interaction are always repulsive. Hecquet [174] finds that surface stress modifies the step–step interaction compared to the Marchenko–Parshin result, enhancing the prefactor of ℓ^{-2} nearly threefold for Au(001); again, there is no mention of attractive interactions over any range of step separations.

In pursuit of a strictly attractive ℓ^{-2} step interaction to explain the results of Shimoni et al. [171], Wang et al. [175] developed a model based on the SSH model [176] of polyacetylene (the original model extended to include electron–electron interaction), focusing on the dimerized atom rows of the (2×1) reconstruction of Si(001). The model produces an attractive correction term to the formula derived by Alerhand et al. [177] for interactions between steps on Si(001), where there is ABAB alternation of $(2 \times 1$ and $1 \times 2)$ reconstructions on neighboring terraces joined at single-height steps. For this type of surface, the correction has little significance, being dwarfed by the logarithmic repulsion. It also does not occur for vicinals to high-symmetry facets of metals. However, for surfaces such as Au(110) with its missing row morphology [178] or adsorbed systems with atomic rows, the row can undergo a Peierls [179] distortion that leads to an analogous dimerization and an ℓ^{-2} attraction. There have been no tests of these unsung predictions by electronic structure computation.

Returning to gold, applications of the glue potential (a semiempirical potential rather similar to EAM), Ercolessi et al. [180] were able to account for reconstructions of various gold facets, supporting that the sharp edges on the ECS are due to the model used for Si(111) rather than attractive step interactions. Studies by this group found no real evidence for attractive step interactions [181].

In an authoritative review a decade ago, Bonzel [2]— the expert in the field who has devoted the most sustained interest in ECS experiments on elemental systems— concluded that it was not possible to decide whether the surface reconstruction model or attractive interactions was more likely to prove correct. In my view, mindful of Ockham's razor, the former seems far more plausible, particularly if the assumed attractive interaction has the ℓ^{-2} form.

The phase diagram of surfaces vicinal to Si(113) presents an intriguing variant of that vicinal to Si(111). There is again a coexistence regime between the (113) orientation and progressively more highly misoriented vicinals as temperature is reduced below a

threshold temperature T_t, associated with a first-order transition. However, for higher temperatures $T > T_t$ there is a continuous transition, in contrast to the behavior on (111) surfaces for $T > T_7$. Thus, Song and Mochrie [182] identify the point along (113) at which coexistence vanishes, i.e., T_t, as a tricritical point, the first such point seen in a misorientation phase diagram. To explain this behavior, Song and Mochrie invoke a mean-field Landau theory argument in which the cubic term in p is proportional to $(T - T_t)$, so negative for $T < T_t$, with a positive quartic term. Of course, this produces the observed generic behavior, but the exponent $\overline{\beta}$ is measured as 0.42 ± 0.10 rather than the mean-field value 1. Furthermore, the shape of the phase diagram differs from the mean-field prediction and the amplitude of the surface stiffness below T_t is larger than above it, the opposite of what happens in mean field. Thus, it is not clear in detail what the interactions actually are, let alone how an attractive interaction might arise physically.

5.9 Well-Established Attractive Step–Step Interactions Other Than ℓ^{-2}

For neutral crystals, there are two ways to easily obtain interactions that are attractive for some values of ℓ. In neither case are the interactions monotonic long range. The first is short-range local effects due to chemical properties of proximate steps, while the other is the indirect Friedel-like interaction.

5.9.1 Atomic-Range Attractions

At very small step separations, the long-range ℓ^{-2} monotonic behavior is expected to break down and depend strongly on the local geometry and chemistry. Interactions between atoms near step edges are typically direct, thus stronger than interactions mediated by substrate elastic fields or indirect electronic effects (see below). We saw earlier that a ℓ^{-3} higher order term arises at intermediate separations [42], and further such terms should also appear with decreasing ℓ. On TaC(910) [vicinal to (001) and miscut toward the [010] direction], Zuo et al. [183] explained step bunching using a weak $\ell^{-3[\pm 0.5]}$ attraction in addition to the ℓ^{-2} repulsion. (The double-height steps are electrically neutral.) Density-functional theory (DFT) studies were subsequently performed for this system by Shenoy and Ciobanu [184]. Similarly, Yamamoto et al. [185] used an attractive ℓ^{-3} dipole-quadrupole interaction to explain anomalous decay of multilayer holes on $SrTiO_3$(001).

More interesting than such generic effects are attractions that occur at very short step separations for special situations. A good example is Ciobanu et al. [186], who find an attraction at the shortest separation due to the cancellation of force monopoles of two adjacent steps on vicinal Si(113) at that value of ℓ.

As alluded to above, most of our understanding of the role of ℓ^{-2} step interactions comes from the mapping of classical step configurations in 2D to the worldlines of spinless fermions in 1D. Unlike fermions, however, steps can touch (thereby forming double-height steps), just not cross. Such behavior is even more likely for vicinal fcc or

bcc (001) surfaces, where the shortest possible "terrace," some fraction of a lateral nearest-neighbor spacing, amounts to touching fermions when successive layers of the crystal are described with simple-cubic rather than layer-by-layer laterally offset coordinates. Sathiyanarayanan et al. [187] investigated some systematics of step touching, adopting a model in which touching steps on a vicinal cost an energy ϵ_t. Note that $\epsilon_t = \infty$ recoups the standard fermion model. For simplicity, the short study concentrates on the "free-fermion" case $\tilde{A} = 0$, i.e., $\varrho = 2$ (cf. Eqn (5.28)). Even for $\epsilon_t = 0$, there is an effective attraction, i.e., $\varrho < 2$, since the possibility of touching broadens the TWD. This broadening is even more pronounced for $\epsilon_t < 0$. In other words, such short-range effects can appear, for a particular system, to contribute a long-range attraction. Closer examination shows that this attraction is a finite-size effect that fades away for large values of $\langle \ell \rangle$. In our limited study, we found that fits of simulated data to the GWD expression could be well described by the following finite-size scaling form, with the indicated three fitting parameters:

$$\varrho_{\text{eff}} = 2 - (0.9 \pm 0.1)\langle \ell \rangle^{-0.29 \pm 0.07} \exp\left[- (3.3 \pm 0.2)\epsilon_t \Big/ k_B T \right]. \tag{5.43}$$

While Eqn (5.43) suggests that making the step touching more attractive (decreasing ϵ_t) could decrease β_{eff} without limit, instabilities begin to develop, as expected since Lässig [188] showed that for $\tilde{A} < -1/4$, i.e., $A < -(k_B T)^2/4\tilde{\beta}$, a vicinal surface becomes unstable (to collapse to step bunches). Correspondingly, the lowest value tabulated in Sathiyanarayanan et al. [187] is $\epsilon_t / k_B T = -0.2$.

To distinguish true long-range (ℓ^{-2}) attractions on vicinal surfaces requires measurements of several different vicinalities (i.e., values of $\langle \ell \rangle$). Likewise, in analyses of ECS data, consideration of crystallites of different sizes would seem necessary. Wortis [1] noted the importance of size dependence in other contexts.

Along this theme, an instructive specific case is the "sticky-step" or, more formally, the p-RSOS (restricted solid-on-solid with point-contact attractions between steps) model explored in detail by Akutsu [189] using the product wavefunction renormalization group (PWFRG) method, calculating essentially the ECS (see Figure 5.25) and related properties. Steps are zig-zag rather than straight as in the preceding Sathiyanarayanan model, so her "stickiness" parameter ϵ_{int} is similar but not identical to ϵ_t. She finds that in some temperature regimes, nonuniversal non-Prokrovsky–Talapov values of ϑ occur. Specifically, let $T_{f,111}(\epsilon_{int}/\epsilon, \phi_0)$ and $T_{f,001}(\epsilon_{int}/\epsilon, \phi_0)$ be the highest temperature at which a first-order phase transition (sharp edge) occurs for the (111) and (001) facets, respectively, where ϕ_0 indicates the position along the ECS. Note $T_{f,111}(\epsilon_t/\epsilon, \phi_0) = (0.3610 \pm 0.0005)\epsilon/k_B > T_{f,111}(\epsilon_{int}/\epsilon, \phi_0) = (0.3585 \pm 0.0007)\epsilon/k_B$. For $k_B T/\epsilon = 0.37$, so $T > T_{f,111}(-0.5, \pi/4)$, Akutsu recovers Prokrovsky–Talapov values for ϑ and ϑ_t, but for $k_B T/\epsilon = 0.36$ (shown in Figure 5.26), so $T_{f,111}(-0.5, \pi/4) > T > T_{f,001}(-0.5, \pi/4)$, the values are very different: $\vartheta = 1.98 \pm 0.03$ and $\vartheta_t = 3.96 \pm 0.08$, more like mean field. For $\phi_0 = 0$ (tilting toward the $\langle 100 \rangle$ direction), only standard Prokrovsky–Talapov exponents are found. Upon closer examination with Monte Carlo simulations, Akutsu finds large step bunches for $T < T_{f,100}$ but step droplets for $T_{f,001} < T < T_{f,111}$. The details are beyond the

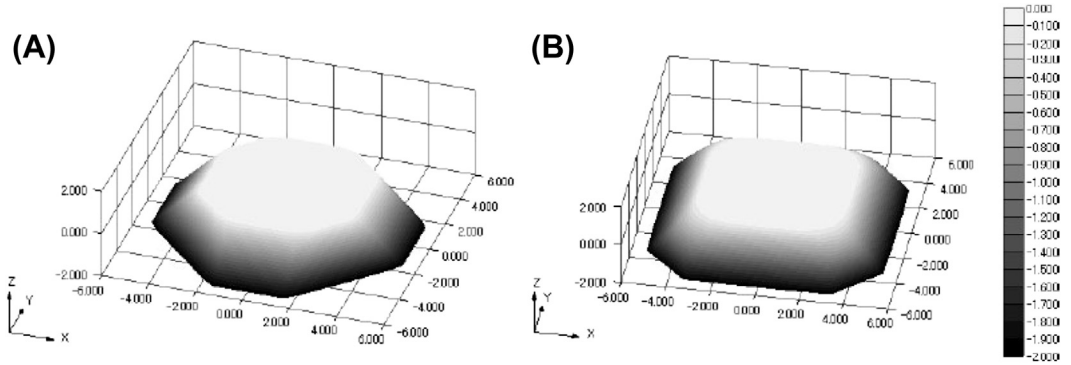


FIGURE 5.25 Perspective views of essentially the equilibrium crystal shape (actually the Andreev surface free energy divided by k_BT) around the (001) facet calculated by the transfer-matrix method with the product-wave-function renormalization group algorithm at $k_BT/\epsilon_1 = 0.3$. (A) Restricted solid-on-solid with point-contact attractions between steps (p-RSOS) model ($\epsilon_{int}/\epsilon_1 = -0.5$). (B) For comparison, the original unsticky RSOS model ($\epsilon_{int} = 0$). *From Ref. [189].*

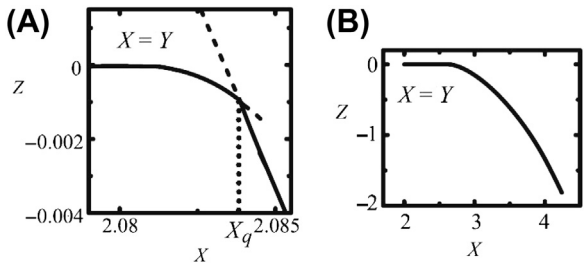


FIGURE 5.26 Profiles in the diagonal direction of the surface in Figure 5.25, still at $k_BT/\epsilon_1 = 0.3$. Broken lines represent metastable lines. (A) $k_BT/\epsilon_1 = 0.36$, $\epsilon_{int}/\epsilon_1 = -0.5$, on a very fine length scale. The edge of the (111) facet is denoted by X_q. (B) The original RSOS model ($\epsilon_{int} = 0$) on a much coarser scale. On this scale (and on an intermediate scale not included here), the profiles are flat until the edge. On the intermediate scale, the region beyond X_q is starts deviating rather smoothly for $k_BT/\epsilon_1 = 0.35$ but looks straight for $k_BT/\epsilon_1 = 0.36$ and 0.37. See text and source. *From Ref. [189].*

scope of this review, but eventually Akutsu deduces an expansion of the projected free energy that includes either a quadratic term or a term after the linear term that has the form $|p|^{\zeta}$, with $\zeta > 1$.

5.9.2 Attractions at Periodic Ranges of Separation via Oscillatory Friedel-Type Interactions

Oscillatory (in sign) interactions, mediated by substrate conduction electrons, between steps ipso facto lead to attractive interactions between steps. As reviewed by Einstein [94], such interactions have been known for many decades to account for the ordered patterns of adsorbates on metal surfaces [190]. While at short range, all electrons

contribute, asymptotically the interaction is dominated by the electron(s) at the Fermi surface or, from another perspective, the nonanalyticity in the response function at the nesting vector. The interaction energy has the form

$$E_{\text{pair}}^{\text{asymp}} \propto \ell^{-n} \cos\left(2k_F\ell + \Phi\right) \tag{5.47}$$

This, or its analogue for interacting local magnetic species, is called the RKKY [191,192] interaction. (The community studying magnetism now labels as RKKY any interaction mediated by substrate electrons, not just the asymptotic limit written down in the RKKY papers.) The phase factor Φ is the nonperturbative result is the scattering phase shifts at the two atoms that are interacting; it is absent in the perturbational approach to this problem used in the RKKY papers. The exponent n indicates the decay envelope. For interacting bulk entities, $n = 3$, the standard RKKY results. On metal surfaces, the leading term in the propagator is canceled by the image charge, leading to $n = 5$, with very rapid decay [94,190]. Such effects are insignificant for adatom interactions but can be more potent when a whole step participates. Redfield and Zangwill [193] show that a line of localized perturbations will generate an interaction with n reduced by subtracting 1/2 and Φ augmented by $\pi/4$. They used this result, with $n = 9/2$, to account for Frohn et al.'s [194] remarkable experimental results on vicinal Cu(001): from their observed bimodal TWD, Frohn et al. deduced that the step–step interaction is attractive for intermediate distances three to five atoms. Indeed, it was their striking observation that led to several of the previously discussed theory papers that claimed to find long-range step attractions.

When there are metallic surface states (i.e., surface states for which their 2D band dispersion relation crosses the Fermi energy E_F) of Shockley nature (lying in a 2D band gap containing E_F), the indirect interaction has a much slower decay, with $n = 2$ [94,195–199]. Furthermore, the Fermi wavevector typically is much smaller than that of bulk states, so the period of the oscillation in real space is much larger. Perhaps the most familiar metallic surface on metals is that at the center of the surface Brillouin zone ($\overline{\Gamma}$) of the (111) surfaces of noble metals, which exist inside the necks of the Fermi surface, discussed in textbooks, e.g. [200]. This is the state that produces the famous wave structure in Eigler's group's dramatic STM images [201] of atoms on metal surfaces. However, there is a less well-known metallic surface state on Cu(001), discovered relatively late (compared to other surface states) by Kevan [202]; it is centered at the zone-edge center $\overline{X}$ rather than $\overline{\Gamma}$, and may provide a better explanation of the Frohn et al. results in the Redfield–Zangwill framework. For surface-state mediated interactions between steps, their formula indicates $n = 3/2$, comparable to the entropic and elastic repulsions.

The effect of surface-state mediated interactions on TWDs was elucidated by Pai et al. [203] in combined experimental and theoretical examination of vicinal Ag(110), which has a metallic surface state centered at $\overline{Y}$, the middle of the shorter edge of the rectangular surface Brillouin zone [204]. In essence, the surface state introduces a second length scale, the Fermi wavelength λ_F, in addition to $\langle\ell\rangle$, with the major consequence that the TWD is no longer a function of the single scaled dimensionless variable s but depends

also on $\langle \ell \rangle$. With a suitable model potential, Pai et al. [203] could then account for the different TWDs at a few different misorientations (i.e., mean step spacings). Indeed, to establish convincingly that this Friedel-like effect is significant, one must measure several different values of $\langle \ell \rangle$. While this paper has been cited with regard to other modifications of TWDs (cf. e.g., Refs [205,206]), I have found no other investigations of Friedel-like effects on TWDs for several misorientations of the same substance.

Patrone and Einstein [207] discuss other issues related to possible anisotropic surface state dispersion as well as showing the insensitivity to the point in the surface Brillouin zone about which the state is centered.

5.10 Conclusions

An aspect of ECS studies on which there has been substantial progress since the 1980s, but which has received little attention in this chapter, is comparing and reconciling the values of the characteristic energies (surface free energy per area, step free energy per length, and step–step repulsion strength) that are extracted from experimental measurements with ever-improving calculations (using density functional theory) of these energies. Bonzel's review [2], as well as Nowicki and Bonzel [140], Bonzel et al. [139], Barreteau et al. [226], Yu et al. [227], contain extensive coverage of this issue for the soft metals to which his group has devoted exhaustive attention. Williams [59] review most results for silicon. Such efforts to find absolute energies has also taken place in studies of island shapes, e.g., of TiN(001) [208] and (111) [209].

There are several significant advances in generic understanding of ECS since the 1980s. The Prokrovsky–Talapov ($\vartheta = 3/2$) critical phenomena near the edge of the smoothly curved region near a facet has proved to be far more robust and general than originally realized, while novel behavior is predicted in a very special direction. Even though invoked in many accounts of sharp edges, long-range attractive ℓ^{-2} do not have an apparent physical basis, except perhaps in idiosyncratic cases. The likely cause is a reconstruction or adsorption that changes the surface free energy of the facet orientation. On the other hand, hill-and-valley structures are widely seen, and the possibility of azimuthal in addition to polar misorientation can lead to astonishingly rich phase diagrams. Of course, nonequilibrium considerations open up a whole new universe of behavior. Furthermore, at the nanoscale, cluster shape is very sensitive to the particulars of a system, with the addition or removal of a single atom leading to a substantial change in shape, rather like biological systems, in contrast to the macroscale phenomena that have been treated in this chapter.

Acknowledgments

My research related to this subject was long supported by the UMD-NSF MRSEC; it is now supported partially by NSF-CHE 13-05892. Much of this paper is based on extensive collaboration with the experimental surface physics group at the University of Maryland, led by Ellen D. Williams until 2010,

with ongoing guidance by Janice Reutt-Robey and William G. Cullen, as well as with Margret Giesen and Harald Ibach, and Hans P. Bonzel, at FZ-Jülich, partially sponsored by a Humboldt Senior U.S. Scientist Award, and, over the last decade, with Alberto Pimpinelli. I also benefited from fruitful interactions at Maryland with John D. Weeks, theory postdocs Olivier Pierre-Louis, Howard L. Richards, Ferenc Szalma, and Kwangmoo Kim, and graduate students Norman C. Bartelt, Raymond C. Nelson, Sanjay V. Khare, Hailu Gebremarian, Timothy J. Stasevich, Rajesh Sathiyanarayanan, Paul N. Patrone, and Josue R. Morales-Cifuentes.

References

[1] Wortis M. Equilibrium crystal shapes and interfacial phase transitions. In: Vanselow R, Howe R, editors. Chemistry and physics of solid surfaces, VII. Berlin: Springer-Verlag; 1988. p. 367–405.

[2] Bonzel HP. Phys Rep 2003;385:1.

[3] Williams ED, Bartelt NC. Ultramicroscopy 1989;31:36.

[4] Rottman C, Wortis M. Phys Rep 1984;103:5979.

[5] Pimpinelli A, Villain J. Physics of crystal growth. Cambridge (England): Cambridge University Press; 1998.

[6] Landau LD, Lifshitz EM. Statistical physics, part 1. revised and enlarged by Lifshitz EM and Pitaevskii LP. 3rd ed. Oxford: Pergamon Press; 1980. pp. 155.

[7] Nozières P. Shape and growth of crystals. In: Godrèche C, editor. Solids far from equilibrium. Cambridge: Cambridge University Press; 1992.

[8] Sekerka RF. Theory of crystal growth morphology. In: Müller G, Métois J-J, Rudolph P, editors. Crystal growth – from fundamentals to technology. Amsterdam: Elsevier; 2004. p. 55.

[9] van Beijeren H, Nolden I. "The roughening transition", structure and dynamics of surfaces II: phenomena, models, and methods. In: Schommers W, von Blanckenhagen P, editors. Topics in current physics, vol. 43. Berlin: Springer; 1987. p. 299.

[10] J. W. Gibbs, "On the equilibrium of heterogeneous substances", Transactions of the Connecticut academy of arts and sciences, 3, 108 248, 343 524, (1874–1878). Reproduced in both the Scientific Papers. 1906; 55–353 and The Collected Works of Gibbs JW. 1928; 55–353. The Collected Works of Gibbs JW, in two volumes, Longley WR, Van Name RG, editors. New Haven: Yale University Press; 1957 (1928).

[11] Herring C. Phys Rev 1951;82:87–93.

[12] Herring C. The use of classical macroscopic concepts in surface energy problems. In: Gomer R, Smith CS, editors. Conference arranged by National Research Council, Lake Geneva, 1952. Chicago: University of Chicago Press; 1953b. p. 5–81 [Chapter 1].

[13] Curie P. Bull Soc Min Fr 1885;8:145.

[14] Wulff G. Z für Kryst Mineral 1901;34:449.

[15] Hilton H. Mathematical crystallography. Oxford; 1903.

[16] Liebmann H. Z Krist 1914;53:171.

[17] von Laue M. Der Wulffsche Satz für die Gleichgewichtsform von Kristallen. Z Krist 1943;105:124.

[18] Dinghas A. Z Krist 1944;105:303–14.

[19] Burton WK, Cabrera N, Frank FC. Phil Trans R Soc A 1951;243:299.

[20] Frank FC. Metal surfaces. Metals Park (OH): American Society for Metals; 1962/3 [Chapter 1]p. 1–15.

[21] Mullins WW. In: Metal surfaces. Metals Park (OH): American Society for Metals; 1962/3. p. 17–62 [Chapter 2].

[22] Jackson KA. In: Ueda R, Mullin JB, editors. Crystal growth and characterization. Amsterdam: North-Holland; 1975. p. 21.

[23] Chernov AA. Sov Phys Uspekhi 1961;4:116.

[24] Cerf R, Picard J. The Wulff crystal in ising and percolation models. Berlin: Springer; 2006.

[25] Dobrushin RL, Kotecký R, Shlosman S. The Wulff construction: a global shape from local interactions. Providence (RI): AMS; 1992. preprint: http://www.cpt.univ-mrs.fr/dobrushin/DKS-book.pdf.

[26] Almgren F, Taylor JE. Fractals 1996;3:713.

[27] Peng D, Osher S, Merriman B, Zhao H-K. The geometry of Wulff crystal shapes and its relation with Riemann problems. In: Nonlinear partial differential equations. Contemp. Math, vol. 238. Providence (RI): AMS; 1999. p. 251–303.

[28] Dobrushin RL, Kotecký R, Shlosman SB. J Stat Phys 1993;12:1.

[29] Dacorogna B, Pfister CE. J Math Pures Appl 1992;71:97–118.

[30] Fonseca I. Proc R Soc Lond A 1991;432:125.

[31] De Coninck J, Dunlop F, Rivasseau V. Commun Math Phys 1989;121:401.

[32] Miracle-Sole S. In: Miracle-Sole S, Ruiz J, Zagrebnov V, editors. Mathematical results in statistical mechanics. Singapore: World Scientific; 1999. p. 83 [arXiv:1206.3736v1].

[33] Miracle-Sole S. Wulff shape of equilibrium crystals. 2013 [arXiv:1307.5180v1].

[34] Pfister CE. Helv Phys Acta 1991;64:953.

[35] Miracle-Sole S, Ruiz J. On the Wulff construction as a problem of equivalence of statistical ensembles. In: Fannes M, Verbeure A, editors. On three levels: micro, meso and macroscopic approaches in physics. New York: Plenum Press; 1994. p. 295–302 [arXiv:1206.3739v1].

[36] Fonseca I, Müller S. Proc R Soc Edinb 1991;119:125.

[37] Jayaprakash C, Rottman C, Saam WF. Phys Rev B 1984;30:6549. in the Hamiltonian in their Eqn 3, the factor t/2 should have been t. See erratum in Ref. [152].

[38] Nelson RC, Einstein TL, Khare SV, Rous PJ. Surf Sci 1993;295:462–84.

[39] Ibach H, Schmickler W. Phys Rev Lett 2003;91:016106.

[40] Stewart J, Pohland O, Gibson JM. Phys Rev B 1994;49:13848.

[41] Marchenko VI, Parshin A Ya. Sov. Phys. JETP 1980;52:129 [Zh Eksp Teor Fiz 1980;79:257.].

[42] Najafabadi R, Srolovitz DJ. Surf Sci 1994;317:221–34.

[43] Duxbury PM, Pence TJ, editors. Dynamics of crystal surfaces and interfaces. New York: Plenum; 1997.

[44] Carlon E, van Beijeren H. Phys Rev E 2000;62:7646.

[45] van Albada SB, Rost MJ, Frenken JWM. Phys Rev B 2002;65:205421.

[46] Bonzel HP, Preuss E. Surf Sci 1995;336:209.

[47] Lieb EH, Wu FY. In: Domb C, Green MS, editors. Phase transitions and critical phenomena, vol. 1. London: Academic Press; 1972. p. 331.

[48] Lieb EH. Phys Rev Lett 1967;19:108.

[49] van Beijeren H. Phys Rev Lett 1977;38:993.

[50] Nolden IM, van Beijeren H. Phys Rev B 1994;49:17224.

[51] Kosterlitz JM, Thouless DJ. J Phys C 1973;6:1181.

[52] Kosterlitz JM. J Phys C 1974;7:1046.

[53] Yang CP. Phys Rev Lett 1967;19:586.

[54] Sutherland B, Yang CN, Yang CP. Phys Rev Lett 1967;19:588.

[55] Callen HB. Thermodynamics and an introduction to the mostatistics. 2nd ed. New York: Wiley; 1985.

[56] Kossel W. Nachr Ges Wiss Göttingen 1927;143.

[57] Kossel W. Ann Phys 1934;21:457–80.

[58] García N, Sáenz JJ, Cabrera N. Physica 1984;124B:251.

[59] Jeong H-C, Williams ED. Surf Sci Rep 1999;34:171.

[60] Cahn JW, Carter WC. Metall Mater Trans A Phys Metall Mater Sci 1996;27:1431–40 [arXiv:cond-mat/0703564v1].

[61] Hoffman DW, Cahn JW. Surf Sci 1972;31:368. Cahn JW, Hoffman DW. Acta Met 1974;22:1205.

[62] Wheeler AA. J Stat Phys 1999;95:1245–80.

[63] Balibar S, Castaing B. J Phys Lett 1980;41:L329.

[64] Keshishev KO, Parshin AY, Babkin AV. Sov Phys JETP 1981;53:362.

[65] Wolf PE, Balibar S, Gallet F. Phys Rev Lett 1983;51:1366.

[66] Wolf PE, Gallet F, Balibar S, Rolley E, Nozières P. J Phys Fr 1985;46:1987.

[67] Balibar S, Alles H, Parshin A Ya. Rev Mod Phys 2005;77:317.

[68] Phaneuf RJ, Williams ED. Phys Rev Lett 1987;58:2563.

[69] Bartelt NC, Williams ED, Phaneuf RJ, Yang Y, Das Sarma S. J Vac Sci Technol A 1989;7:1898.

[70] Phaneuf RJ, Bartelt NC, Williams ED, Swiech W, Bauer E. Phys Rev Lett 1993;71:2284.

[71] Rousset S, Berroir JM, Repain V, Garreau Y, Etgens VH, Lecoeur J, et al. Surf Sci 1999;443:265.

[72] Bonczek F, Engel T, Bauer E. Surf Sci 1980;97:595.

[73] Madey TE, Guan J, Nien C-H, Tao H-S, Dong C-Z, Campbell RA. Surf Rev Lett 1996;3:1315.

[74] Madey TE, Nien C-H, Pelhos K, Kolodziej JJ, Abdelrehim IM, Tao H-S. Surf Sci 1999;438:191.

[75] Fisher DS, Weeks JD. Phys Rev Lett 1983;50:1077.

[76] Lyuksyutov I, Naumovets AG, Prokrovsky V. Two-dimensional crystals. San Diego: Academic Press; 1992.

[77] Dieluweit S, Ibach H, Giesen M, Einstein TL. Phys Rev B 2003;67:121410(R).

[78] Stasevich TJ, Einstein TL, Zia RKP, Giesen M, Ibach H, Szalma F. Phys Rev B 2004;70:245404.

[79] Stasevich TJ, Einstein TL, Stolbov S. Phys Rev B 2006;73:115426.

[80] Stasevich TJ, Gebremariam H, Einstein TL, Giesen M, Steimer C, Ibach H. Phys Rev B 2005;71:245414.

[81] Zia RKP. J Stat Phys 1986;45:801.

[82] Avron JE, van Beijeren H, Schulman LS, Zia RKP. J Phys A Math Gen 1982;15:L81–6.

[83] Gallavotti G. Commun Math Phys 1972;27:103.

[84] Abraham DB, Reed P. Phys Rev Lett 1974;33:377.

[85] Abraham DB, Reed P. Commun Math Phys 1976;49:35.

[86] Kodambaka S, Khare SV, Petrov I, Greene JE. Surf Sci Rep 2006;60:55–77.

[87] Michely T, Krug J. Berlin: Springer; 2004 [Chapter 3].

[88] Stasevich TJ, Einstein TL. [SIAM-]Multiscale Model Simul 2007;6:90.

[89] Weeks JD. Private discussions, 2014.

[90] Liu F. Modeling and simulation of strain-mediated nanostructure formation on surface. In: Rieth M, Schommers W, editors. Handbook of theoretical and computational nanotechnology, vol. 4. American Scientific Publishers; 2006. p. 577–625.

[91] Müller B, Nedelmann L, Fischer B, Brune H, Barth JV, Kern K. Phys Rev Lett 1998;80:2642.

[92] Einstein TL. Appl Phys A 2007;87:375384 [cond-mat/0612311].

[93] Einstein TL, Richards HL, Cohen SD, Pierre-Louis O. Surface Sci 2001;493:460.

[94] Einstein TL. Interactions between adsorbate particles. In: Unertl WN, editor. Physical structure of solid surfaces. Handbook of surface science, vol. 1. Amsterdam: Elsevier; 1996. Holloway S, Richardson NV, series editors, 577–650.

[95] Giesen M. Prog. Surface Sci. 2001;68:1.

[96] Gruber E, Mullins WW. J Phys Chem Solids 1967;28:875.

[97] Fisher MPA, Fisher DS, Weeks JD. Phys Rev Lett 1982;48:368.

[98] Leamy HJ, Gilmer GH, Jackson KA. Statistical thermodynamics of clean surfaces. In: Blakely JM, editor. Surface physics of materials, vol. 1. New York: Academic Press; 1975. p. 121 [Chapter 3].

[99] Stasevich TJ. 2006 [Ph.D. dissertation]. University of Maryland [unpublished].

[100] de Gennes PG. J Chem Phys 1968;48:2257.

[101] Prokrovsky VL, Talapov AL. Phys Rev Lett 1979;42:65. Sov Phys-JETP 1980;51:134.

[102] Haldane FDM, Villain J. J Phys Paris 1981;42:1673.

[103] Schulz HJ, Halperin BI, Henley CL. Phys Rev B 1982;26:3797.

[104] Prokrovsky VL, Talapov AL. Theory of incommensurate crystals, soviet scientific reviews supplement series physics, vol. 1. Chur: Harwood Academic Publishers; 1984 [and references therein].

[105] Villain J. In: Riste T, editor. Ordering in strongly fluctuating condensed matter systems. New York: Plenum; 1980. p. 221.

[106] Akutsu Y, Akutsu N, Yamamoto T. Phys Rev Lett 1988;61:424.

[107] Jeong H-C, Weeks JD. Surf Sci 1999;432:101 [and references therein].

[108] Wang X-S, Goldberg JL, Bartelt NC, Einstein TL, Williams ED. Phys Rev Lett 1990;65:2430.

[109] Ihle T, Misbah C, Pierre-Louis O. Phys Rev B 1998;58:2289.

[110] Joós B, Einstein TL, Bartelt NC. Phys. Rev. B 1991;43:8153.

[111] Mehta ML. Random matrices. 3rd ed. New York: Academic; 2004.

[112] Dyson FJ. Statistical Theory of the Energy Levels of Complex Systems. III. J. Math. Phys. 1962;3:166.

[113] Bartelt NC, Einstein TL, Williams ED. Surf Sci 1990;240:L591.

[114] Calogero F. J Math Phys 1969;10:2191. 2197.

[115] Sutherland B. J Math Phys 1971;12:246. Phys Rev A 1971;4:2019.

[116] Dyson FJ. Commun. Math. Phys. 1970;19:235.

[117] Haake F. Quantum signatures of chaos. 2nd ed. Berlin: Springer; 1991.

[118] Guhr T, Müller-Groeling A, Weidenmüller HA. Phys. Rept 1998;299:189.

[119] Gebremariam H, Cohen SD, Richards HL, Einstein TL. Phys Rev B 2004;69:125404.

[120] Jayaprakash C, Saam WF. Phys Rev B 1984;30:3916.

[121] Andreev AF. Sov. Phys.-JETP 1982;53:1063 [Zh. Eksp. Teor. Phys 1981;80:2042].

[122] Jayaprakash C, Saam WF, Teitel S. Phys Rev Lett 1983;50:2017.

[123] Cabrera N, García N. Phys Rev B 1982;25:6057.

[124] Cabrera N. The equilibrium of crystal surfaces. Surf Sci 1964;2:320.

[125] Kardar M, Nelson DR. Phys Rev Lett 1985;55:1157.

[126] Dahmen SR, Wehefritz B, Albertini G. A novel exponent in the equilibrium shape of crystals. 1998. arXiv:cond-mat/9802152.

[127] Akutsu Y, Akutsu N, Yamamoto T. Universality of the tangential shape exponent at the facet edge of a crystal. 1998. arXiv:cond-mat/9803189.

[128] Akutsu Y, Akutsu N. Prog Theor Phys 2006;116:983.

[129] Nishino T, Okunishi K. J Phys Soc Jpn 1995;64:4084.

[130] Okunishi K, Hieida Y, Akutsu Y. Phys Rev B 1999;59:6806. 1999;60:R6953.

[131] Sato R, Akutsu Y. J Phys Soc Jpn 1995;64:3593.

[132] Carmi Y, Lipson SG, Polturak E. Phys Rev B 1987;36:1894.

[133] Rottman C, Wortis M, Heyraud JC, Métois JJ. Phys Rev Lett 1984;52:1009.

[134] Saénz JJ, García N. Surf Sci 1985;155:24.

[135] Surnev S, Arenhold K, Coenen P, Voigtländer B, Bonzel HP. J Vac Sci Technol A 1998;16:1059.

[136] Nowicki M, Bombis C, Emundts A, Bonzel HP, Wynblatt P. New J Phys 2002;4:60.

[137] Nowicki M, Bombis C, Emundts A, Bonzel HP, Wynblatt P. Eur Phys Lett 2002;59:239.

[138] Nowicki M, Bombis C, Emundts A, Bonzel HP. Phys Rev B 2003;67:075405.

[139] Bonzel HP, Nowicki M. Phys Rev B 2004;70:245430.

[140] Bonzel HP, Yu DK, Scheffler M. Appl Phys A 2007;87:391.

[141] Thürmer K, Reutt-Robey JE, Williams ED. Surf Sci 2003;537:123.

[142] Métois JJ, Heyraud JC. Surf Sci 1987;180:647.

[143] Bermond JM, Métois JJ, Heyraud JC, Floret F. Surf Sci 1998;416:430.

[144] Gladić J, Vučić Z, Lovrić D. J Cryst Growth 2002;242:517–32.

[145] Ferrari PL, Prähofer M, Spohn H. Phys Rev E 2004;69:035102.

[146] Pimpinelli A, Degawa M, Einstein TL, Williams ED. Surf Sci 2005;598:L355.

[147] Degawa M, Stasevich TJ, Cullen WG, Pimpinelli A, Einstein TL, Williams ED. Phys Rev Lett 2006;97:080601.

[148] Degawa M, Stasevich TJ, Pimpinelli A, Einstein TL, Williams ED. Surface Sci. 2007;601:3979 [Proc. ECOSS 2006].

[149] Einstein TL, Alberto Pimpinelli. J Statistical Phys 2014;155:1178. longer version posted at arXiv http://arxiv:1312.4910v1.

[150] Farnsworth HE, Schlier RE. J Chem Phys 1959;31:89.

[151] Hwang RQ, Williams ED, Park RL. Phys Rev B 1989;40:11716.

[152] Williams ED, Phaneuf RJ, Wei Jian, Bartelt NC, Einstein TL. Surface Sci 1993;294:219. Erratum 1994:310:451.

[153] Noh DY, Blum KI, Ramstad MJ, Birgeneau RJ. Phys Rev B 1991;44:10969.

[154] Noh DY, Blum KI, Ramstad MJ. Phys Rev B 1993;48:1612.

[155] Held GA, Goodstein DM, Brock JD. Phys Rev B 1995;51:7269.

[156] Noh DY, Liang KS, Hwu Y, Chandavarkar S. Surf Sci 1995;326:L455.

[157] Dietrich S, Haase A. Phys Rep 1995;260:1.

[158] Ocko B. Private discussion, 2014.

[159] Cahn JW. J de Physique 1982;43(C6):199–213. Proceedings of Conference on the Structure of Grain Boundaries, Caen, France.

[160] Somorjai GA, Van Hove MA. Prog Surf Sci 1989;30:201.

[161] Cheng W-C, Wynblatt P. Surf Sci 1994;302:185.

[162] Meltzman H, Chatain D, Avizemer D, Besmann TM, Kaplan WD. Acta Mater 2011;59:3473.

[163] Yoon M, Mochrie SGJ, Zehner DM, Watson GM, Gibbs D. Phys Rev B 1994;49:16702.

[164] Walko DA, Robinson IK. Phys Rev B 2001;64:045412.

[165] Szczepkowicz A, Ciszewski A, Bryl R, Oleksy C, Nien C-H, Wu Q, et al. Surf Sci 2005;599:55.

[166] Kalff M, Comsa G, Michely T. Phys Rev Lett 1998;81:1255.

[167] Stasevich TJ, Tao C, Cullen WG, Williams ED, Einstein TL. Phys Rev Lett 2009;102:085501.

[168] Wang Z, Wynblatt P. Surf Sci 1998;398:259.

[169] Handwerker CA, Vaudin MD, Blendell JE. J Phys Colloq 1988;49:C5-367.

[170] Emundts A, Bonzel HP, Wynblatt P, Thrmer K, Reutt-Robey J, Williams ED. Surf Sci 2001;481:13.

[171] Shimoni N, Ayal S, Millo O. Phys Rev B 2000;62:13147.

[172] Kukta RV, Peralta A, Kouris D. Phys Rev Lett 2002;88:186102.

[173] Prévot G, Croset B. Phys Rev Lett 2004;92:256104.

[174] Hecquet P. Surf Sci 2008;602:819.

[175] Wang G, Webb JF, Zi J. Surf Sci 2007;601:1944.

[176] Su WP, Schrieffer JR, Heeger AJ. Phys Rev Lett 1979;42:1698. Phys Rev B 1980;22:2099.

[177] Alerhand OL, Vanderbilt D, Meade RD, Joannopoulos JD. Phys Rev Lett 1988;61:1973.

[178] Copel M, Gustafsson T. Phys Rev Lett 1986;57:723.

[179] Peierls RF. Quantum theory of solids. Oxford: Clarendon; 1955. p. 108 [Also Fröhlich H. Proc R Soc A 1954;223:296].

[180] Ercolessi F, Bartolini A, Garofalo M, Parrinello M, Tosatti E. Surf Sci 1987;189/190:636.

[181] Tosatti E. Private communication, March 2014.

[182] Song S, Yoon M, Mochrie SGJ. Surf Sci 1995;334:153.

[183] Zuo J-K, Zhang T, Wendelken JF, Zehner DM. Phys Rev B 2001;63:033404.

[184] Shenoy VB, Ciobanu CV. Phys Rev B 2003;67(081402).

[185] Yamamoto M, Sudoh K, Iwasaki H, Williams ED. Phys Rev B 2010;82:115436.

[186] Ciobanu CV, Tambe DT, Shenoy VB, Wang CZ, Ho KM. Phys Rev B 2003;68:201302R.

[187] Sathiyanarayanan R, Hamouda ABH, Einstein TL. Phys Rev B 2009;80:153415.

[188] Lässig M. Phys Rev Lett 1996;77:526.

[189] Akutsu N. J Phys Condens Matter 2011;23:485004.

[190] Einstein TL, Schrieffer JR. Phys Rev B 1973;7:3629.

[191] Ruderman MA, Kittel C. Phys Rev 1954;96:99.

[192] Yosida K. Phys Rev 1957;106:893.

[193] Redfield AC, Zangwill A. Phys Rev B 1992;46:4289.

[194] Frohn J, Giesen M, Poensgen M, Wolf JF, Ibach H. Phys Rev Lett 1991;67:3543.

[195] Lau KH, Kohn W. Surf Sci 1978;75:69.

[196] Repp J, Moresco F, Meyer G, Rieder K-H, Hyldgaard P, Persson M. Phys Rev Lett 2000;85:2981.

[197] Hyldgaard P, Persson M. J Phys Condens Matter 2000;12:L13.

[198] Knorr N, Brune H, Epple M, Hirstein A, Schneider MA, Kern K. Phys Rev B 2002;65:115420.

[199] Hyldgaard P, Einstein TL. J Cryst Growth 2005;275:e1637 [cond-mat/0408645].

[200] Ashcroft NW, Mermin ND. Solid state physics. Cengage Learning; 1976.

[201] Crommie MF, Lutz CP, Eigler DM. Nature 1993;363:524. Science 1993;262:218.

[202] Kevan SD. Phys Rev B 1983;28:2268(R).

[203] Pai WW, Ozcomert JS, Bartelt NC, Einstein TL, Reutt-Robey JE. Surf Sci 1994;309:747.

[204] Liu SH, Hinnen C, van Huong CN, de Tacconi NR, Ho K-M. J Electroanal Chem 1984;176:325.

[205] Mugarza A, Schiller F, Kuntze J, Cordón J, Ruiz-Osés M, Ortega JE. J Phys Condens Matter 2006;18:S27.

[206] Li F, Allegretti F, Surnev S, Netzer FP. Surf Sci 2010;604:L43.

[207] Patrone PN, Einstein TL. Phys Rev B 2012;85:045429.

[208] Kodambaka S, Khare SV, Petrova V, Vailionis A, Petrov I, Greene JE. Surf Sci 2002;513:468474.

[209] Kodambaka S, Khare SV, Petrova V, Johnson DD, Petrov I, Greene JE. Phys Rev B 2003;67:035409.

[210] Stranski I. Z. Phys Chem Leipz 1928;136:259.

[211] Flytzani-Stephanopoulos M, Schmidt LD. Prog Surf Sci 1979;9:83.

[212] Rottman C, Wortis M. Phys Rep 1984;103:5979.

[214] Gibbs JW. Trans Conn Acad 1877;3:108–248. see Collected Works 1928, p. 343–524.

[215] Zia RKP. Anisotropic surface tension and equilibrium crystal shapes. In: Hu CK, editor. Progress in Statistical Mechanics. Singapore: World Scientific; 1988. p. 303–57.

[216] Weeks JD. In: Riste T, editor. Ordering in strongly fluctuating condensed matter systems. New York: Plenum; 1980. p. 293.

[217] Rottman C, Wortis M. Phys Rev B 1984;29:328.

[218] von Laue M. Z. Krist Min 1944;105:124.

[219] Andreev AF. Sov Phys-JETP 1982;53:1063.

[220] For a review of fermionic methods, see den Nijs M. In: Domb C, Lebowitz JL, editors. Phase transitions and critical phenomena, vol. 12. London: Academic; 1989.

[221] Williams ED, Bartelt NC. Thermodynamics and statistical mechanics of surfaces. In: Unertl WN, editor. Physical structure of solid surfaces. Holloway S, Richardson NV, editors. Handbook of surface science, vol. 1. Amsterdam: Elsevier; 1996. p. 51–99.

[222] Einstein TL, Pimpinelli A. Dynamical scaling implications of Ferrari, Prähofer, and Spohn's remarkable spatial scaling results for facet-edge fluctuations, arXiv 1312.4910.

[223] Song S, Mochrie SGJ. Phys Rev Lett 1994;73:995.

[224] Song S, Mochrie SGJ. Phys Rev B 1995;51:10068.

[225] Shimoni N, Ayal S, Millo O. Phys Rev B 2000;62:13147.

[226] Barreteau C, Raouafi F, Desjonquères MC, Spanjaard D. J Phys Condens Matter 2003;15:3171.

[227] Yu DK, Bonzel HP, Scheffler M. The stability of vicinal surfaces and the equilibrium crystal shape of Pb by first principles theory. New J Phys 2006;8:65.

[228] Daw MS, Foiles SM, Baskes MI. Mater Sci Rep 1993;9:251.

6

Rough–Smooth Transition of Step and Surface

Noriko Akutsu[1], Takao Yamamoto[2]

[1]FACULTY OF ENGINEERING, OSAKA ELECTRO-COMMUNICATION UNIVERSITY, NEYAGAWA, OSAKA, JAPAN; [2]DIVISION OF PURE AND APPLIED SCIENCE, FACULTY OF SCIENCE AND TECHNOLOGY, GUNMA UNIVERSITY, KIRYU, GUNMA, JAPAN

CHAPTER OUTLINE

6.1 Introduction: Universal Features 266
6.2 Background 270
6.2.1 Rough Surface in Crystal Growth—Viewpoint of BCF 270
6.2.2 Entropy Effect—Jackson's Argument 271
6.2.3 "Drumhead Wandering" due to Capillary Waves 272
6.3 Rough Surface 275
6.3.1 Definition of an Interface 275
6.3.1.1 Existence of a Roughening Transition Temperature 275
6.3.1.2 Ising Model and the Equivalent Lattice Gas Model 276
6.3.1.3 Definition of Interface Tension 277
6.3.2 Definition of Surface Roughness 279
6.3.2.1 Surface Width—Variance of Surface Height 279
6.3.2.2 Height–Height Correlation Function of a Surface 280
6.3.2.3 Thin Film-Like Surface in the Macroscopic Scale 281
6.3.3 Relationship between Energy and Shape 281
6.3.4 Roughness of a Single Step 282
6.4 Roughening Transition and Faceting Transition as Critical Phenomena 284
6.4.1 Microscopic Models for Studying Surface Roughness 284
6.4.1.1 BCSOS Model 285
6.4.1.2 ASOS Model and RSOS Model 286
6.4.2 Summary of KT-Type Critical Phenomena of Surfaces 286
6.4.3 Diffuseness for Atomically Rough Surfaces 290
6.4.3.1 New Picture of Roughening 290
6.4.3.2 Intrinsic Width of a Step 291
6.4.3.3 Preroughening Phenomena 291

Handbook of Crystal Growth. http://dx.doi.org/10.1016/B978-0-444-56369-9.00006-X

6.4.4 Changes in the Roughening Temperature on Complex Surfaces 292
6.4.4.1 Inverse Roughening 292
6.4.4.2 Surface Modified by Langmuir Adsorption 292
6.5 Vicinal Surface 294
6.5.1 Rough or Smooth? The Terrace-Step-Kink Picture 294
6.5.2 1D Free-Fermion Universal Features—Gruber-Mullins-Pokrovsky-Talapov Behavior 295
6.5.3 Logarithmic Behavior on the Width of a Single Step 296
6.5.3.1 Height–Height Correlation Function of a Single Step 296
6.5.3.2 Elastic Step–Step Repulsion 297
6.6 Step Faceting 298
6.6.1 Stability of a Macrostep 298
6.6.2 Discontinuous Surface Tension 300
6.6.2.1 p-RSOS Model 300
6.6.2.2 Discontinuity in Surface Tension 300
6.7 Summary 302
Appendix A. Transfer Matrix Method 303
Appendix B. Driving Force for Crystal Growth 304
Appendix C. Example of the Anisotropy of the Entropy of a Step 304
Appendix D. IPW Method 304
Appendix E. Calculation of Surface Width 306
Appendix F. Derivation of the Capillary Wave Hamiltonian 307
Appendix G. Other Microscopic Models 308
Appendix G.1. Discrete Gaussian Model 308
Appendix G.2. RSOS-I Model 308
Acknowledgment 309
References 309

6.1 Introduction: Universal Features

During the years 1970–2000, much progress was made in understanding roughening and faceting transitions [1–39]. Roughening transitions are particularly notable because they are a typical example of the Kosterlitz–Thouless (KT) transition [40–43]. The KT transition was first presented as a magnetic phase transition of the two-dimensional (2D) XY model [40]. Later, 2D crystals [40,41], a 2D Coulomb gas [42], and a superfluid film [43] of helium (^{4}He) were also shown to exhibit KT transitions.

Van Beijeren provided the correct understanding of the roughening transition by his exact calculation of the body-centered cubic solid-on-solid (BCSOS) model [1]. The

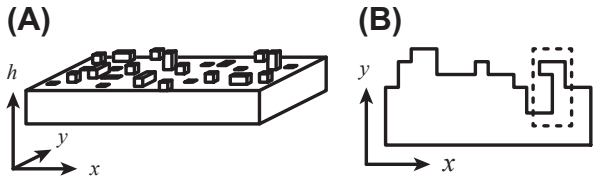


FIGURE 6.1 Microscopic diagram of a crystal surface. (A) Perspective view of a solid-on-solid model; (B) top view of a step. The pattern inside the broken rectangle shows the overhanging structure.

solid-on-solid (SOS) model, Figure 6.1(A), is a model for studying surface roughness when it is inhibited by an overhanging structure (the overhang structure is shown in the dotted square of Figure 6.1(B)).[1] Van Beijeren showed by exact calculations that the step internal energy e_{step} reduces to zero at the roughening transition temperature T_{R} as $e_{\text{step}} \propto \exp(-C/\sqrt{T_{\text{R}} - T})$. This also means that the step tension decreases to zero at T_{R} as $\gamma_{\text{step}} \propto \exp(-C'/\sqrt{T_{\text{R}} - T})$. It had not been known previously if the step tension became zero at T_{R}. His exact calculation showed us, for the first time, the close connection between surface roughening and the KT transition. Soon after, Knops [10] and Jose et al. [11] showed that any 2D SOS models in 3D can be mapped to the XY model. In Figure 6.2, we show this observed singular property of the step tension on Si(001) [32].

Without the long-range order in 2D, the KT phase transition has special characteristics. When the number of dimensions is less than or equal to two, thermal fluctuations destroy the long-range order [47]. In the XY model, the KT transition occurs at the temperature T_{KT}, where the entropy of forming a special spin configuration called a "vortex" exceeds the excitation energy of the vortex [43]. For $T < T_{\text{KT}}$, the 2D XY model forms a quasi-long-range order for the pairs of $+1$ and -1 vortices; for $T > T_{\text{KT}}$, the proliferation of the vortex monomers destroys the quasi-long-range order. The correlation length ξ_{KT}, which characterizes the size of the coherent domain, is infinite for $T < T_{\text{KT}}$, and $\xi_{\text{KT}} \propto \exp(A/\sqrt{T - T_{\text{KT}}})$ for $T > T_{\text{KT}}$. Thus, the critical exponents are infinite, and ξ_{KT} depends on temperature differently than does a typical phase transition, such as that in the 2D Ising model.[2]

The KT transition, however, is such a subtle phase transition that it is difficult to directly detect the singularity. Fortunately, in the roughening transition, the universal quantities specific to the KT transition are measurable as geometrical quantities on an equilibrium crystal shape (ECS) [48–58], which is the shape of a crystal droplet with the least surface free energy. The shape change accompanied by the roughening transition is

[1]From several numerical studies of a 2D interface in the 3D Ising model, the roughening transition temperature T_{R} is estimated to be almost half of the transition temperature T_c of the 3D Ising model. Hence, up to around T_{R}, the frequency at which the overhang structure forms is thought to be sufficiently low. As a step, however, an overhang such as the one shown within the broken line in Figure 6.1(B) frequently appears near T_c in the 2D Ising model.

[2]The correlation lengths of the 2D and 3D Ising models diverge only at the temperature of the phase transition.

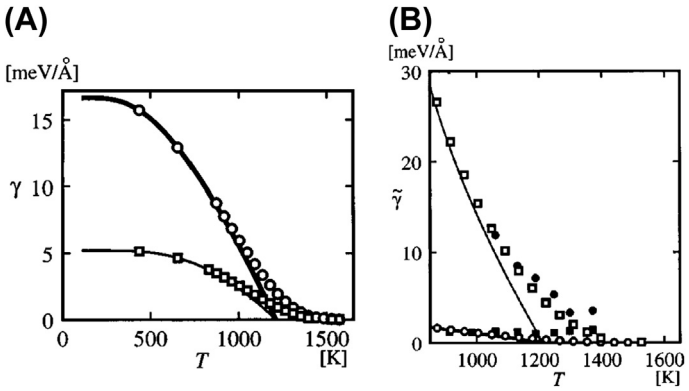


FIGURE 6.2 Step tension γ and step stiffness $\tilde{\gamma}$ of an Si(001) surface [32]. (A) Step tension; (B) step stiffness. Open squares: values of the S_A step calculated by the PWFRG method [44] (see Appendix A). Open circles: values of the S_B step calculated by the PWFRG method. Thick and thin solid curves: one-dimensional (1D) interface in a two-dimensional (2D) next-nearest-neighbor Ising model for the S_B step and S_A steps, respectively [31]. Solid circles: low-energy electron microscopy (LEEM) [45] results for the S_A step. Solid squares: LEEM results for the S_B step [46]. *From Ref. [32].*

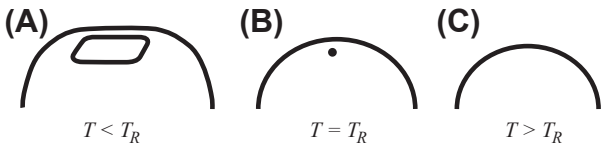


FIGURE 6.3 Faceting transition on an equilibrium crystal shape.

called the faceting transition [18] (Figure 6.3). A facet is a plane with a low Miller index that appears on the ECS (Figure 6.3(A)). The facet appears at $T < T_R$, where T_R represents that of the facet plane. The shape of the facet represents the 2D ECS with respect to the step tension [24,33]. The area of the facet shrinks as the temperature increases. At T_R, the area of the facet becomes zero (Figure 6.3(B)). For $T > T_R$, the facet disappears (Figure 6.3(C)).

The exact study of the vicinal surface in the BCSOS model [18] again leads to correct understanding of the faceting transition. When the temperature rises to just above T_R, the Gaussian curvature[3] K_G [59] at the topmost point in Figure 6.3(B) jumps from 0 to $\Delta K_G = [\lambda a_z^2/(k_B T_R)]^2 K_R^2$, where a_z represents the height of a single step, λ represents the Lagrange multiplier relating to the volume of the crystal droplet, T represents the temperature, and k_B represents the Boltzmann constant. It is surprising that the *universal quantity* $K_R = 2/\pi$ appears in the expression of the Gaussian curvature jump, where K_R is proportional to the inverse of the KT transition temperature of the XY

[3]The Gaussian curvature is defined as $K_G = \kappa_1 \kappa_2$, where κ_1 and κ_2 are the principal values of the curvature.

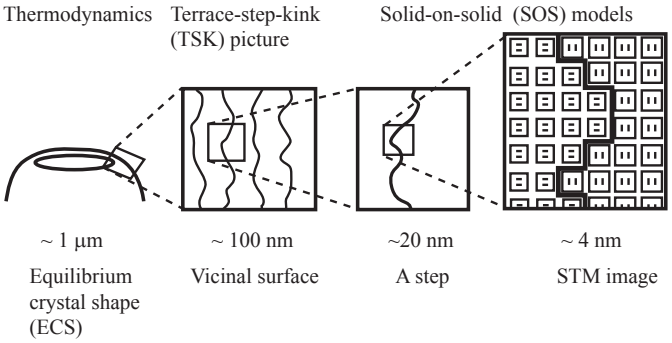


FIGURE 6.4 Models in various scales. *From Ref. [86].*

model: $K_R = J/(k_B T_{KT})$, where J is the coupling constant of the XY model [18]. This relation is a result of the duality relationship between the XY model and the SOS model, which Knops [10] and Jose et al. [11] pointed out.

The vicinal surface for $T < T_R$, where T_R represents the roughening transition temperature of the *terrace* plane, is also interesting because the system can be mapped to a 1D system of free fermions [13–33]. For $T < T_R$, the vicinal surface is described by a regular train of steps with a zigzag structure (Figure 6.4). This image of the vicinal surface is called the terrace-step-kink (TSK) or the terrace-ledge-kink (TLK) picture [60–63]. The free energy of the vicinal surface $f(\rho)$ can be obtained by the ground-state energy of the 1D free fermion (FF), as follows [13–33]:

$$f(\rho) = f(0) + \gamma\rho + B\rho^3 + \cdots \tag{6.1}$$

where ρ represents the step density, γ represents the step tension, and B represents the step interaction coefficient. This correspondence of the surface steps to quantum particles provides another universal type of surface; in the field of surface studies, a surface for which the energy is described in Eqn (6.1) is said to be of the Gruber-Mullins-Pokrovsky-Talapov (GMPT) type [13]. In addition, for $T < T_R$, the vicinal surface near a facet edge has another universal quantity reminiscent of the KT transition: the universal Gaussian curvature jump at the facet edge [25,26]. The Gaussian curvature jump can be expressed exactly in the 1D FF system, independent of the orientation of the vicinal surface: $\Delta K'_G = [\lambda a_z^2/(k_B T)]^2 (K_R/2)^2$. We will discuss this in detail in Section5. We add here that these singularities on the ECS are observed in systems of ^{4}He [64–72], Pb [73–76], Ag_2S [77], Si [78–82], and other materials [83–85].

This chapter gives an overview of the universal features of the roughening and smoothing phenomena of surfaces and steps. As background, we mention the treatment of a surface by Burton, Cabrera, and Frank (BCF) [87] for vapor growth, the treatment of the surface for melt growth by Jackson [88,89], and the "drumhead wandering" [90–92] that is caused by capillary waves. We begin with surfaces at temperatures higher than T_R and then consider those at lower temperatures. This also means that we begin with models at macroscopic scales and proceed to those at microscopic scales. We do this

because surfaces and their interfaces are so complex that many models in various scales need to be considered (Figure 6.4). We begin with an explanation of the surface thermodynamic quantities, including definitions of surface, interface, and surface roughness (Section 3). Next, we show the critical phenomena when the surface is near the temperature T_R (Section 4), after which we mention several topics concerning roughening transitions. We then discuss the vicinal surface when $T < T_R$ (Section 5). Finally, we mention the smoothing of steps and the formation of macrosteps, which is due to an anomalous polar-angle dependence of the surface tension (Section 6).

6.2 Background

6.2.1 Rough Surface in Crystal Growth—Viewpoint of BCF

Before we proceed further, we briefly review the classical but important work of Burton, Cabrera and Frank [87], and Jackson [88].

BCF discussed the importance of the roughening transition of a surface[4] [87], and they showed that the growth mechanisms are very different on a rough surface than they are on a smooth surface.

The growth rate of a rough surface is proportional to the driving force $\Delta\mu$ of the crystal growth ($\Delta\mu$ is the difference between the chemical potentials of the bulk crystal phase and the ambient phase[5]). A smooth surface, however, does not grow linearly with $\Delta\mu$. Hence, BCF had to introduce the 2D nucleation process, a surface with a regular train of steps (a vicinal surface), and a surface with screw dislocations [87,93] in order to explain a realistic growth rate of a crystal (for example, see Figure 6.5). To see this, let us consider a curved step on a surface that is near equilibrium. As mentioned in Appendix D.12 in BCF [87], the normal velocity of a step v_n on a surface is described as follows:

$$v_n = \nu\left[\frac{\Delta\mu}{\Omega} - \frac{\tilde{\gamma}}{R}\right], \quad \tilde{\gamma} = \gamma + \frac{\partial^2\gamma}{\partial\phi^2}, \tag{6.2}$$

where ν represents a kinetic coefficient, $\Delta\mu$ represents the driving force,[6] Ω represents a unit volume, R represents the radius of curvature, $\tilde{\gamma}$ represents the step stiffness, γ represents the step tension, and ϕ represents the tilt angle of the mean tangential line of a step, relative to the y-axis (Figure 6.5(B)). Equation (6.2) is known as the Gibbs-Thomson equation for a curved step.

[4]BCF studied the roughening transition by adopting a two-dimensional (2D) Ising model by means of the Bethe approximation. The limits of this approximation method misled them to thinking that the step tension had a finite value at the temperature of the roughening transition.

[5]For vapor growth, $\Delta\mu = k_B T \ln\alpha = k_B T \ln P/P_0$, in the BCF notation, where k_B represents the Boltzmann constant, T represents the temperature, P represents the pressure in the ambient phase, and P_0 represents the equilibrium pressure at some temperature T_0. For other cases, please see Appendix B [94].

[6]$\Delta\mu/\Omega = kT\ln\alpha$ in equation D.12 of BCF, and it expresses the case of $v_n = 0$.

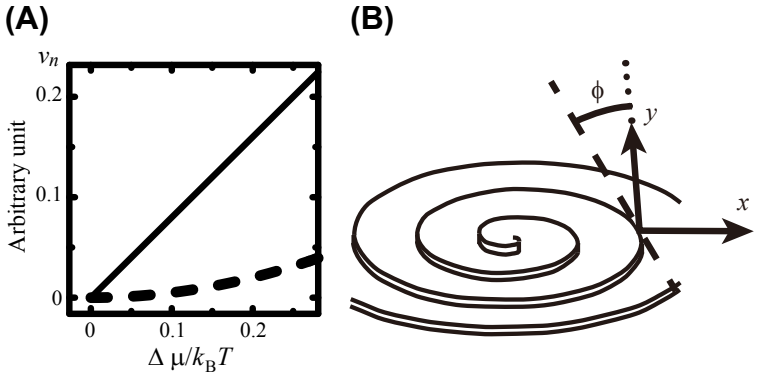


FIGURE 6.5 (A) Example of crystal growth at a surface at a normal velocity v_n. Solid line: a rough surface. Broken line: a smooth surface with a screw dislocation. (B) Schematic illustration of a screw step on a surface (perspective view) [87].

As a natural extension of Eqn (6.2), the normal velocity of a surface v_n is described as follows [95,96]:

$$v_n = \nu\left[\frac{\Delta\mu}{\Omega} - \frac{f_1}{R_1} - \frac{f_2}{R_2}\right], \tag{6.3}$$

where R_1 and R_2 represent the radii of curvature in the principal direction, and f_1 and f_2 represent the surface stiffness tensors in the principal direction. It should be noted that the surface of the crystal whose growth behavior is expressed by Eqn (6.3) is implicitly assumed to be rough. For a smooth surface, the value of $f_1 \times f_2$ diverges [23], so Eqn (6.3) cannot be applied to describe the growth velocity. On a smooth surface, crystal growth occurs under conditions that are far from equilibrium. Hence, the growth process of a smooth surface depends on the microscopic details of the surface.

6.2.2 Entropy Effect—Jackson's Argument

In 1958, Jackson [88] pointed out the relationship between the surface roughness and the entropy of melting, and introduced the parameter α, which is called Jackson's parameter. Jackson's parameter indicates the relative degree to which the surface energy contributes to the free energy, compared to the contribution of the surface entropy. In this subsection, we discuss the surface entropy of roughness.

According to the calculations on the two-level (Jackson) model [55,88,94], in which the interface is considered in the "atomic" scale,[7] the interface between the liquid and the crystal is described by a 2D lattice. The free energy of the interface between the liquid and the crystal $G(C,T)$ is given by a mean-field approximation, as follows:

$$G(C, T_m)/(Nk_BT_m) \approx \alpha C(1-C) + C\ln C + (1-C)\ln(1-C), \tag{6.4}$$

[7]What was meant by the "atomic" scale in 1958 is equivalent to a scale that is more than a unit cell of a crystal.

where T_m represents the melting temperature, N represents the total number of lattice points in the 2D lattice, k_B represents the Boltzmann constant, C represents the concentration of the "solid-like atom," and $\alpha = z'\phi_{eff}/(2k_BT_m)$ represents Jackson's α. Here, z' represents the number of nearest-neighbor (nn) sites for a 2D lattice on a surface, ϕ_{eff} represents the effective bond energy. Since the bond energy is approximately measured by the heat of melting per molecule Δh, Jackson's α can be rewritten using the entropy of melting, as follows:

$$\alpha = \frac{z'}{z}\frac{\Delta h}{k_BT_m} = \frac{z'}{z}\frac{\Delta s}{k_B}, \tag{6.5}$$

where z represents the number of nn sites in a 3D lattice and Δs represents the entropy of melting per molecule. Using the roughening transition temperature T_R, Jackson's α can be rewritten as follows: $\alpha \approx 2T_R/T_m$ [97]. Therefore, α can be used to estimate the roughening transition temperature of a specific surface.

The original idea of introducing α in Eqn (6.4) was to consider the relative contributions of the energy and the entropy to the interface free energy (Figure 6.6). The first term on the right-hand side (r.h.s.) of Eqn (6.4) represents the energy cost due to interface roughness, and the second and the third terms on the r.h.s. represent the contributions of the interface entropy. Hence, for large α, energy wins and the interface is smooth, while for small α, entropy wins and the interface becomes rough. Usually, the entropy does not depend on the substance,[8] but the bond energy strongly depends on the substance. Therefore, for each substance, Jackson's α parameter gives us information about whether the roughening transition will be observable.[9]

The anisotropy of interface entropy due to the interface roughness is sometimes ignored, and we note that there is also anisotropy due to the structure of the crystal lattice. As shown in C, the ground-state structure of the interface is generally degenerated except for the several interfaces that have a low Miller index. When there is an in-plane bond network that connects all of the atoms, the interface will have a unique structure [51].

6.2.3 "Drumhead Wandering" due to Capillary Waves

In Figure 6.7, we show an intuitive picture of phase coexistence for a bulk crystal and a bulk ambient phase. The two phases are separated by a narrow interface region of width w.

[8]We recall Trouton's rule for the vaporization of a liquid. This rule states that the entropy of vaporization is almost constant, irrespective of the substance.

[9]If the loss of the long-range periodicity of a crystal has a dominant contribution to the increase of entropy, as in metals, the melting of entropy will be small and so α is small (2–3). The interface is rough near the melting temperature. On the other hand, in the case of orientational molecules, for example, the entropy of melting is large because disordering of the molecular orientation occurs together with the disordering of the periodicity of the crystal. In this case, α becomes large ($\sim$10). Then, the surfaces at several azimuths become facets at the melting temperature.

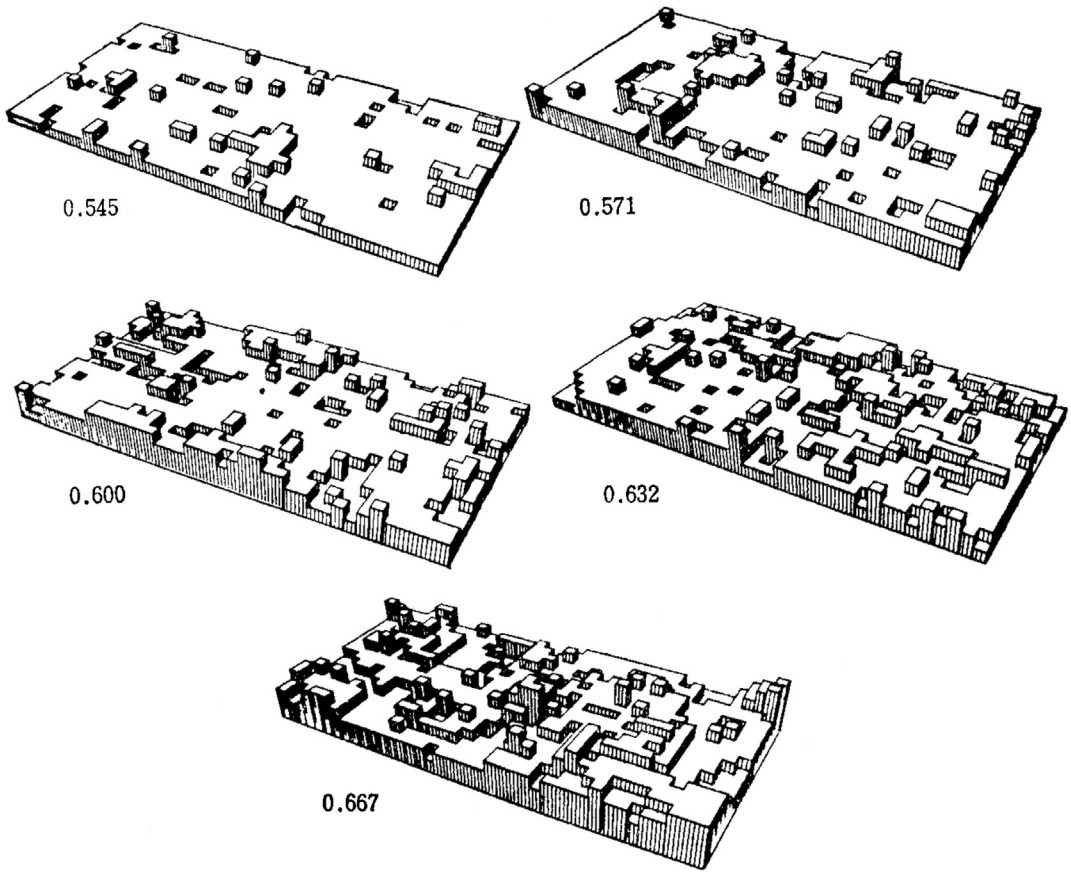


FIGURE 6.6 Snapshots of a computer simulation of the roughening transition on a solid-on-solid model [89]. *From Ref. [55].*

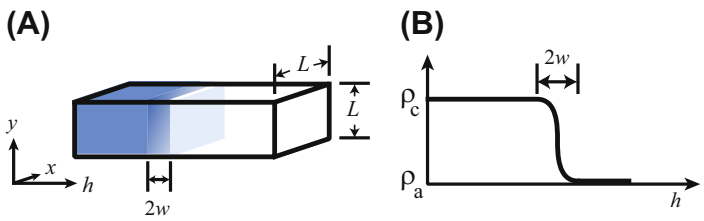


FIGURE 6.7 Schematic illustration of crystal-vapor coexistence, showing the diffuse interfacial region. (A) Shaded area: crystal; *w*: width of interface. (B) Density profile. ρ_c: density of crystal; ρ_a: density of the ambient phase (vapor phase).

If we try to consider the interface/surface precisely, we find that the definition of an interface (surface) is not trivial. At a crystal-melt interface, for example, the location of the interface is rather ambiguous when we consider the model at a microscopic scale. The liquid phase is generally distinguished from the crystal phase by the lack of a

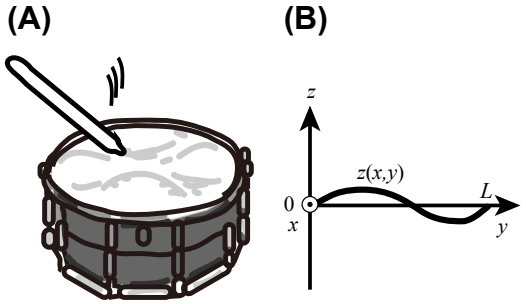


FIGURE 6.8 (A) A drum with a drumhead. (B) Surface height of the drumhead.

long-range periodic structure. This means that we cannot tell the difference between "liquid" and "crystal" by observing only a single atom. The same situation occurs in the magnetic domain walls of the XY model and the Heisenberg model. The microscopic parameters at the interface or in the domain wall change continuously, as shown in Figure 6.7. In this case, the interface (or the crystal surface) is unstable against the long-wavelength distortions that are excited by thermal fluctuations with infinitesimal energy costs [90–92]. That is, the interface is rough.

In this subsection, using the drumhead model (Figure 6.8), we explain how the long-wavelength distortions called "capillary wave" destabilize the continuous interface/surface.

The work functional against the surface tension for the instantaneous distortion of the surface dividing phase 1 from phase 2 is given by Buff, Lovett, and Stillinger in the following equation [90–92]:

$$\mathcal{H}_{DH} = \int \gamma_0 \sqrt{g} dxdy + \frac{1}{2} m_0^2 \int z(x,y)^2 dxdy, \tag{6.6}$$

where $z(x,y)$ represents the instantaneous "dividing surface" between phases 1 and 2 (the surface of the drumhead), $\sqrt{g}dxdy$ represents a small surface area dA, γ_0 represents the surface tension, and m_0 represents the force necessary to stabilize the surface (e.g., gravity for a liquid–gas interface). Here, g represents a geometrical factor and is defined by $g = 1 + p_x^2 + p_y^2$, where $p_x = \partial z/\partial x$ and $p_y = \partial z/\partial y$. This work functional is also called the "drumhead" Hamiltonian [92]. The thermodynamic equations and the correlations follow from $Z_{DH} = \int Dz[\exp\{-\beta\mathcal{H}_{DH}\}]$ with $\beta = 1/k_B T$, where $\int Dz$ describes the integral of all possible surface distortions. The surface area term in Eqn (6.6) can be expanded as

$$\sqrt{g} = 1 + \frac{1}{2}\left(p_x^2 + p_y^2\right) + \mathcal{O}\left[\left(p_x^2 + p_y^2\right)^2\right] + \cdots. \tag{6.7}$$

The Hamiltonian of Eqn (6.6) is usually approximated by the term that has lowest order with respect to $\sqrt{g}$: $1 + (1/2)(p_x^2 + p_y^2)$.

The local strength of the fluctuations from the mean shape of the surface $\langle z(x,y)\rangle = 0$ is measured by the variance, where $<\cdot>$ represents the thermal average. This becomes

$$\left\langle z(x,y)^2 \right\rangle = \frac{k_B T}{(2\pi)^2} \int \frac{dq_x dq_y}{\gamma_0 \left(q_x^2 + q_y^2\right) + m_0^2}, \tag{6.8}$$

where q_x and q_y are the wave numbers introduced by the Fourier decomposition (similar to the argument in Appendix E). In the thermodynamic limit ($L \to \infty$), this integral becomes

$$\left\langle z(x,y)^2 \right\rangle = \frac{k_B T}{4\pi\gamma_0} \ln \frac{\gamma_0 \Lambda^2 + m_0^2}{\gamma_0 k_0^2 + m_0^2}, \tag{6.9}$$

where $\Lambda = 2\pi/a$ represents the upper cutoff, a is the lattice spacing, and $k_0 = 2\pi/L$ represents the low-wavenumber (infrared) cutoff for the integral. Therefore, for $m_0^2 = 0$, the variance diverges logarithmically in the thermodynamic limit $L \to \infty$, as follows:

$$\left\langle z(x,y)^2 \right\rangle = \frac{k_B T}{4\pi\gamma_0} \ln L. \tag{6.10}$$

The long-wavelength fluctuations, the capillary waves, destabilize the surface. This surface instability is referred to as "drumhead wandering" [92].

It is important to note that the drumhead wandering depends on the dimensionality of the space. Extending the above argument to a general $(d-1)$-dimensional interface in a d-dimensional space, the variance of the interface fluctuations Eqn (6.8) becomes as follows:

$$\langle z^2 \rangle = \frac{k_B T}{(2\pi)^{d-1}} \int \frac{d^{d-1}q}{\gamma_0 q^2 + m_0^2} \propto \int_{k_0}^{\Lambda} \frac{q^{d-2} dq}{\gamma_0 q^2 + m_0^2}. \tag{6.11}$$

In the thermodynamic limit ($L \to \infty$), $\langle z^2 \rangle$ becomes $(m_0^2)^{\frac{1}{2}(d-3)}$. Hence, for $d < 3$, $\langle z^2 \rangle$ obeys a power law, and for $d = 3$, $\langle z^2 \rangle$ becomes $\sim -\ln m_0^2$. While for $d > 3$, there is no divergence around $k_0 = 0$. As for the surface with $m_0 \to 0$,[10] for $d = 2$, the 1D surface (a step) is always rough; for $d = 3$, the surface is marginal; and for $d > 3$ (if it exists), the surface is always smooth [4,5].

6.3 Rough Surface

6.3.1 Definition of an Interface

6.3.1.1 Existence of a Roughening Transition Temperature

Rigorously speaking, at equilibrium, the crystal surface is an interface between the crystal phase and the ambient phase. However, we will use the terms *surface* and *interface* interchangeably. We begin this section with the definition of an interface.

[10] As a candidate of m_0 on a surface, we may consider the potential to form a lattice structure. Such an effect can be taken into consideration by the discrete Gaussian model [98]. In the case of metals, a quantum effect with respect to the electrons may be a candidate for the case of m_0.

When a crystal is surrounded by a vapor phase, the microscopic separation between the crystal phase and the vapor phase is relatively clear.[11] The discrete order parameter, therefore, can be used to describe the configuration of the interface.[12] In this case, the surface of the crystal can be described as the interface in a 3D Ising model. For the interface in a 3D Ising model, it has been rigorously proven that the interface causes a roughening transition at a temperature less than the transition temperature of the bulk 3D Ising model $T_{c,3D}$ [4,5]. In addition, the roughening transition temperature T_R is near the transition temperature of the 2D Ising model $T_{c,2D}$, but has been rigorously proven to be higher than $T_{c,2D}$ [4]. Since the existence of the roughening transition is guaranteed by these rigorous results for the interface in a 3D Ising model, several solid-on-solid (SOS) models, which are specialized for surfaces, will be used to study surface roughness.

A coarse-grained rough surface of a crystal is diffuse, and it will have a surface profile that appears similar to that shown in Figure 6.7. In the following subsubsections, we will give a brief overview of how to assign tension to a coarse-grained interface when using the Ising model. To simplify this, we will use the 2D Ising model.

6.3.1.2 *Ising Model and the Equivalent Lattice Gas Model*

The 2D and the 3D Ising models are often adopted for the microscopic study of surface roughness. The Hamiltonian of the Ising model is written as follows[13]:

$$\mathcal{H}_{\text{Ising}} = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - H \sum_{i=1}^{\mathcal{N}} \sigma_i, \tag{6.12}$$

where $\sigma = \{\pm 1\}$ represents the Ising spin, J represents the coupling constant between spins, H represents an external magnetic field, and $\mathcal{N}$ represents the total number of lattice points. The summation of $\langle i,j \rangle$ is taken over all the nearest-neighbor (nn) pairs. The equivalent Hamiltonian of the lattice gas model is written as follows:

$$\mathcal{H}_{\text{LG}} = -\widehat{\phi} \sum_{\langle i,j \rangle} C_i C_j - (\mu_{2\text{D}} - \mu_{\text{ambient}}) \sum_{i=1}^{\mathcal{N}} C_i, \tag{6.13}$$

where $\widehat{\phi}$ represents the bond energy between nn atoms, $C = \{0,1\}$ indicates the presence of lattice gas ($C = 1$ when it is present and $C = 0$ otherwise), $\mu_{2\text{D}}$ represents the chemical potential of the atoms in the 2D lattice, and μ_{ambient} represents the chemical potential of the atoms in the ambient phase. $\mu_{2\text{D}}$ is approximately written by $\mu_{2\text{D}} = -\widehat{\phi}/2 + \mu_{\text{crystal}}$, where μ_{crystal} represents the chemical potential of the atoms in the bulk crystal. The Ising

[11] Even in the case of a crystal-vapor interface, the location of the phase separation sometimes becomes ambiguous for multicomponent crystals, such as can be seen in stones.

[12] The crystal-melt interface can be described by the discrete order parameter when another parameter, such as one for dielectric polarization, accompanies the crystal-melt transition. It is possible to identify melt or crystal at the microscopic scale by observing the dielectric polarization parameter.

[13] Figure 6.13(A) shows a side view of a 2D lattice gas model that is equivalent to the 2D Ising model.

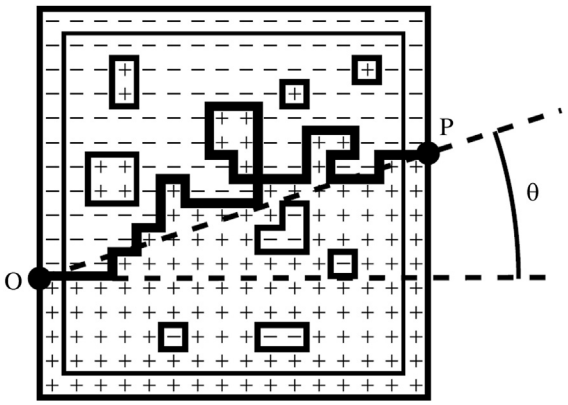


FIGURE 6.9 A 1D interface as a phase separation line under an antiphase boundary condition in the 2D Ising model [4,6–8,38,39]. In the lattice gas model that is equivalent to this 2D Ising model, the phase separation line corresponds to a step on the surface. *From Ref. [39].*

Hamiltonian (6.12) is translated into the lattice gas Hamiltonian by substituting $\sigma = 2C - 1$, as follows:

$$\begin{aligned} &\mathcal{H}_{\text{Ising}} = \mathcal{H}_{\text{LG}} + \Delta E \\ &\Delta E = \Big(2H - zJ\Big)\mathcal{N}/2, \quad 4J = \widehat{\phi}, \quad 2H = \mu_{\text{crystal}} - \mu_{\text{ambient}}, \end{aligned} \tag{6.14}$$

where z represents the number of nn sites.[14]

In the case of $J = 0$, the Ising model (the lattice gas model) reduces to a two-level model with independent elements (Jackson's model).

6.3.1.3 *Definition of Interface Tension*

The partition function of the Ising system is obtained as $Z_{\text{Ising}} = \sum_{\sigma_i} \exp[-\beta \mathcal{H}_{\text{Ising}}]$, where $\beta = 1/k_{\text{B}}T$. The summation $\{\sigma_i\}$ is taken over all possible spin configurations. The free energy (density) of the system is obtained as $f(T) = -(k_{\text{B}}T/\mathcal{N}) \ln Z_{\text{Ising}}$. The exact form of $f(T)$ was given by Onsager for the 2D Ising model with $H = 0$ [99].

The free energy of an interface is generally defined as the *excess* free energy due to the coexistence of the two phases. We will consider an example of this for a 1D interface in a 2D Ising model. For a temperature $T < T_{c,2\text{D}}$, where $T_{c,2\text{D}}$ represents the Curie temperature of the 2D Ising model, a finite magnetized spin configuration will be self-organized. A phase separation line, such as the one between end points O and P in Figure 6.9, can be formed by applying a special boundary condition called an antiphase boundary condition (Figure 6.9). We denote the partition function of the Ising model system with a uniform boundary condition by Z_{++}, and the partition function with the antiphase boundary condition by $Z_{+-}(\theta)$. The 1D interface free energy per length (in the horizontal direction) is defined as follows [4]:

$$f_{1\text{D}}(\theta) = \lim_{N_x \to \infty} \frac{-k_{\text{B}}T}{N_x a} \ln \frac{Z_{+-}(\theta)}{Z_{++}}, \tag{6.15}$$

[14]This $zJ(= z\widehat{\phi}/4)$ relates to the "half crystal site."

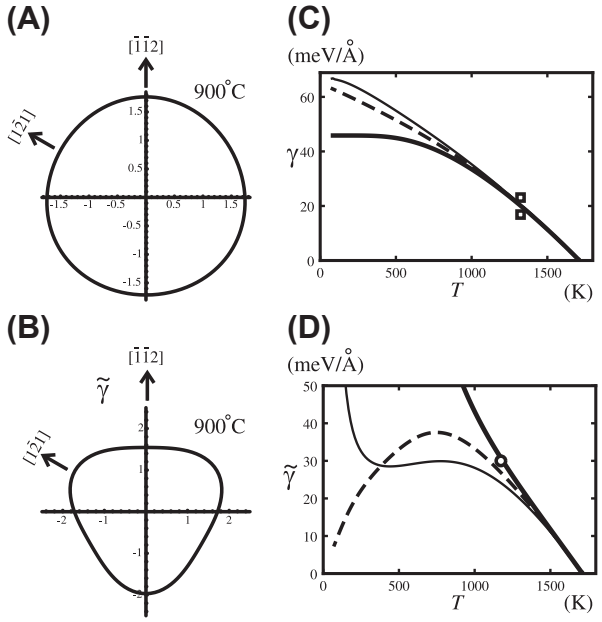


FIGURE 6.10 Step quantities for a double layer of a (111) surface of a diamond structure [39]. The microscopic coupling constants were chosen to reproduce the experimental observations [81,101] on Si(111), which are denoted by open squares and an open circle. (A) The equilibrium shape of an island at 900 °C; (B) a polar graph of the step stiffness at 900 °C; (C) temperature dependence of the step tension; (D) temperature dependence of the step stiffness. In (C) and (D), the thick lines correspond to the $\{2\bar{1}\bar{1}\}$ step and the thin lines to the $\{10\bar{1}\}$ step. *From Ref. [39].*

where N_x is the number of lattice points in the horizontal direction. The interface tension of a 1D interface is given by

$$\gamma_{1D}(\theta) = f_{1D}(\theta)|\cos\theta| = f_{1D}(\theta)\Big/\sqrt{1+p_x^2}, \tag{6.16}$$

where $p_x = \tan\theta$ [6,94].

In the case of the 1D interface, we can go further by using the imaginary-path-weight random walk (IPW) method (Appendix D) [100]. In Figure 6.2, the calculated step quantities for Si(001) are shown [31]. The Ising values were calculated by the IPW method. As seen from Figure 6.2, the Ising model provides good approximations for the step tensions and step stiffness at low temperatures. Figure 6.10 shows the step tension and step stiffness for the Si(111) surface, as calculated by using the honeycomb Ising model with a staggered magnetic field [39]. The calculated values were obtained by the IPW method.

The random walk picture of the interface that connects O and P locates the 1D interface at the line connecting O and P in the limit of $L \to \infty$. The existence of the limit is guaranteed by the central limit theorem. Similarly, a 2D interface in the 3D Ising model can be introduced under the antiphase boundary condition. The idea of the phase separation line can be easily extended to that of a phase separation surface in the 3D

Table 6.1 Surface Width and Surface Roughness ($d = 3$)

Squared Surface Width W^2_{surf}	Squared Surface Roughness w^2_{surf}	Remarks
$\frac{k_B T}{2\pi\sqrt{\det(f^{ij})}} \ln L$	$\frac{k_B T}{2\pi\sqrt{\det(f^{ij})}}$	$T > T_R$ (Section 3.2.1, Section 3.3, Appendix E)
$\frac{a_z^2}{2\pi} K_R \ln L$	$\frac{a_z^2}{2\pi} K_R$	$T = T_R$, $K_R = \frac{2}{\pi}$ (Section 3.2.1, Section 4.2)
w^2	0	$T < T_R$ smooth surface (Section 3.2.1) w^2: a finite value of the variance of the surface height
$\frac{a_z^2}{2\pi}\left(\frac{K_R}{2}\right) \ln L$	$\frac{a_z^2}{2\pi}\left(\frac{K_R}{2}\right)$	$T < T_R$ for vicinal surface with small slope limit (Section 5.2)
$\frac{k_B T}{2\pi\sqrt{\det(f^{ij})}} \ln L$	$\frac{k_B T}{2\pi\sqrt{\det(f^{ij})}}$	$T < T_R$ for vicinal surface with large slope (Section 3.3)
$\frac{a_z^2}{2\pi\tilde{\lambda}(g)}\left(\frac{K_R}{2}\right) \ln L$	$\frac{a_z^2}{2\pi\tilde{\lambda}(g)}\left(\frac{K_R}{2}\right)$	$T < T_R$ for vicinal surface with small slope and with elastic step–step repulsion (Section 5.3)

Ising model. Therefore, the location of the interface as a mathematical plane $h(x,y)$ is determined by the boundary condition in the Ising model.[15]

6.3.2 Definition of Surface Roughness

6.3.2.1 Surface Width—Variance of Surface Height

The variance of the surface height is defined as the *squared surface width*, as follows:

$$W^2_{\text{surf}} = \left\langle [h(\boldsymbol{x}) - \langle h(\boldsymbol{x})\rangle]^2 \right\rangle, \tag{6.17}$$

where $\langle \cdot \rangle$ represents the thermal average. Examples of the squared surface width in the limit as $L \to \infty$ are listed in Table 6.1, where a_z represents the unit height of a step and d is the number of space dimensions.[16] When $T < T_R$, the squared surface width becomes a certain finite value in the limit of $L \to \infty$. Let the finite value be denoted by w^2. Hence, we have

$$w^2 = \lim_{L\to\infty} \left\langle [h(\boldsymbol{x}) - \langle h(\boldsymbol{x})\rangle]^2 \right\rangle. \tag{6.18}$$

[15]For more general cases, the location of the interface within the area w in Figure 6.7 is ambiguous. For simplicity, however, we will consider that the interface is located at the Gibbs' dividing surface, and we will denote it by $h(x,y)$. In the case of multicomponent materials, the interface may have multiple dividing surfaces.

[16]This is done in accordance with the customary notation of the statistical mechanics of an interface [92].

When $T \geq T_R$, the surface width diverges logarithmically as the limit of $L \to \infty$, as follows (Appendix E, Eqn (E.6)) [23]:

$$W_{surf}^2 = \frac{k_B T}{2\pi\sqrt{\det(f^{ij})}} \ln L, \tag{6.19}$$

where (f^{ij}) $(i,j = x,y)$ represents the surface stiffness tensor, which will be explained in the following subsection (Section 3.3). This divergence is caused by the "drumhead wandering" explained in Section 2.3. Though the surface is well defined in the microscopic scale, a rough surface behaves like the liquid–vapor interface in macroscopic scale. That is, long-wavelength distortions of the surface, called capillary waves, destabilize the surface.

Keeping in mind Eqn (6.19), a nondivergent measure of the surface roughness w_{surf} can be defined as follows:

$$w_{surf} = \left[W_{surf}^2 \Big/ \ln L\right]^{1/2} \quad (L \to \infty), \tag{6.20}$$

which we call a *scaled surface width*. The scaled surface width is suitable for defining the "surface roughness." Hence, the squared surface roughness is expressed as follows:

$$w_{surf}^2 = \begin{cases} \dfrac{k_B T}{2\pi\sqrt{\det(f^{ij})}} & T > T_R \\ a_z^2 K_R / 2\pi & T = T_R \\ 0 & T < T_R \end{cases}, \quad (d = 3), \tag{6.21}$$

where we use the universal relationship $\sqrt{\det(f^{ij})} = k_B T/(K_R a_z^2)$ at T_R (Section 4.2) with the universal value $K_R = 2/\pi$.

6.3.2.2 Height–Height Correlation Function of a Surface

There is another quantity that may be suitable for defining the surface roughness. The height–height correlation function is defined as follows:

$$G(\boldsymbol{r}) = \left\langle [h(\boldsymbol{x}+\boldsymbol{r}) - h(\boldsymbol{x})]^2 \right\rangle. \tag{6.22}$$

Similarly, after some calculations, in the limit of $r = |\boldsymbol{r}| \to \infty$, we obtain [22]

$$G(\boldsymbol{r}) = \begin{cases} \dfrac{k_B T}{\pi\sqrt{\det(f^{ij})}} \ln r & T > T_R \\ (a_z^2 K_R/\pi) \ln r & T = T_R \\ 2w^2 & T < T_R \end{cases}, \quad (d = 3). \tag{6.23}$$

where w^2 represents a finite value calculated by Eqn (6.18). Hence, we can also define the surface roughness as $[G(\boldsymbol{r})/\ln r]^{1/2}$, which is the value multiplied by $\sqrt{2}$ of the value defined by Eqn (6.20).

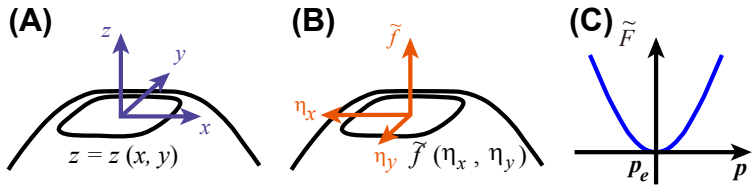


FIGURE 6.11 (A) An equilibrium crystal shape; (B) Andreev free energy [57]; (C) Extended free energy for the Hamiltonian of the capillary wave.

6.3.2.3 *Thin Film-Like Surface in the Macroscopic Scale*

It should be noted that the macroscopic appearance of a rough surface looks like a continuous thin surface. The surface width W_{surf}, the square root of the variance of the surface height, diverges at the rough surface. The ratio of the volume of the surface region ($\sim L^2 \times W_{\mathrm{surf}}$) and the bulk ($\sim L^3$) becomes W_{surf}/L and converges to zero in the limit as $L \to \infty$. Therefore, in the thermodynamic limit, a rough surface looks like a thin elastic film.

6.3.3 Relationship between Energy and Shape

In this subsection, we show how the thermodynamic quantities on the surface relate to the geometry of the equilibrium crystal shape, which is the shape of a crystal droplet that has the least surface free energy (Figure 6.11(A)). The final relationship we would like to present in this subsection is the following [23]:

$$w_{\mathrm{surf}}^2 = \frac{1}{2\pi}\frac{k_{\mathrm{B}}T}{\sqrt{\det(f^{ij})}} = \frac{1}{2\pi}\frac{k_{\mathrm{B}}T}{\lambda} g\sqrt{K_G}, \quad g = 1 + |\boldsymbol{p}|^2, \tag{6.24}$$

where g is the geometrical constant[17] and K_G is the Gaussian curvature [59] of the surface.

The quantities defined on crystal surfaces are anisotropic; this is because of the structure of the crystal lattice. The ECS can be obtained by using the polar graph of the surface tension and the Wulff construction method [49–55]. Andreev and Landau [56,57] proposed another method in which the ECS may be obtained analytically [56,57]; in their method, the ECS and the surface gradient $\boldsymbol{p}$ are described by $z(x,y)$ and $\boldsymbol{p}=(\partial z(x,y)/\partial x, \partial z(x,y)/\partial y)$, respectively. Andreev introduced the Andreev free energy $f(\boldsymbol{\eta})$ [57] as the Legendre transformed thermodynamic potential of a surface with respect to the surface gradient $\boldsymbol{p}$, as follows: $f(\boldsymbol{\eta}) = f(\boldsymbol{p}) - \boldsymbol{p}\cdot\boldsymbol{\eta}$, where $f(\boldsymbol{p}) = \gamma_{\mathrm{surf}}(\boldsymbol{n})\sqrt{g}$ represents the surface free energy per projected xy area, $\gamma_{\mathrm{surf}}(\boldsymbol{n})$ represents the surface tension, $\boldsymbol{n}$ represents the surface normal unit vector (see Eqn (F.1)). Hereafter, we will call $f(\boldsymbol{p})$ the vicinal surface free energy. Here, $\boldsymbol{\eta}$ is the Andreev field, the thermodynamic force conjugate to the surface gradient, which causes the surface to tilt.

[17] More precisely, $g = \det(g_{\mu\nu})$, where $g_{\alpha\beta} = \delta_{\alpha\beta} + p_\alpha p_\beta$ ($\alpha, \beta = \{x,y\}$), $p_x = \partial z(\boldsymbol{x})/\partial x$, and $p_y = \partial z(\boldsymbol{x})/\partial y$ [59].

For the least free energy condition on the crystal droplet, the thermodynamic equations for the ECS [57] are obtained as follows:

$$\begin{aligned} &\tilde{f}(\eta_x,\eta_y)=\lambda z(x,y), \quad \eta_x=-\lambda x, \quad \eta_y=-\lambda y \\ &p_{e,\alpha}=-\frac{\partial\tilde{f}(\eta_x,\eta_y)}{\partial\eta_\alpha}, \quad \eta_\alpha=\frac{\partial f\left(p_x,p_y\right)}{\partial p_\alpha}, \quad (\alpha=x,y), \end{aligned} \tag{6.25}$$

where $\boldsymbol{p}_e$ ($p_{e,x}$, $p_{e,y}$) is the surface gradient in equilibrium

It is interesting that the shape of the Andreev free energy as a function of $\boldsymbol{\eta}$ is similar to the ECS (Figure 6.11(A) and (B)). Therefore, determining a thermodynamic quantity on a surface can be transformed to the problem of determining the geometry of the ECS.

Next, let us consider the fluctuations around the ECS. The Gaussian-type capillary-wave Hamiltonian is obtained from the extended free energy with respect to the slope fluctuations around the equilibrium surface slope (Appendix F) [23,102]:

$$\mathcal{H}_{CW}=\frac{1}{2}\int_0^L dx\int_0^L dy\left[f^{xx}(\boldsymbol{p}_e)\Delta p_x^2+f^{yy}(\boldsymbol{p}_e)\Delta p_y^2+\left(f^{xy}(\boldsymbol{p}_e)+f^{yx}(\boldsymbol{p}_e)\right)\Delta p_x\Delta p_y\right], \tag{6.26}$$

where (f^{ij}) represents the surface stiffness tensor and is defined as follows:

$$(f^{ij})=\begin{pmatrix} f^{xx}(\boldsymbol{p}_e) & f^{xy}(\boldsymbol{p}_e) \\ f^{yx}(\boldsymbol{p}_e) & f^{yy}(\boldsymbol{p}_e)\end{pmatrix}=\left.\begin{pmatrix} \dfrac{\partial^2 f(\boldsymbol{p})}{\partial p_x^2} & \dfrac{\partial^2 f(\boldsymbol{p})}{\partial p_y\partial p_x} \\ \dfrac{\partial^2 f(\boldsymbol{p})}{\partial p_x\partial p_y} & \dfrac{\partial^2 f(\boldsymbol{p})}{\partial p_y^2}\end{pmatrix}\right|_{\boldsymbol{p}=\boldsymbol{p}_e}. \tag{6.27}$$

Using this capillary-wave Hamiltonian, we can determine the thermal average explicitly (Appendix E), and then we obtain

$$W_{\text{surf}}^2=\frac{k_{\text{B}}T}{2\pi\sqrt{\det(f^{ij})}}\ln L. \tag{6.28}$$

Further, from the definition of the Gaussian curvature and Eqn (6.25), we can obtain an expression for the Gaussian curvature K_G, as follows:

$$\begin{aligned} &K_G=\frac{1}{g^2}\det\left(\frac{\partial^2 z}{\partial x^\mu\partial x^\nu}\right)=\frac{\lambda^2}{g^2}\frac{1}{\det(f^{ij})}, \\ &(\mu,\nu=1\ or\ 2, \quad x^1=x, \quad x^2=y). \end{aligned} \tag{6.29}$$

From Eqns (6.20), (6.28), and (6.29), we obtain the final equation, Eqn (6.24).

Equation (6.24) connects the quantities of the surface roughness (the scaled surface width), the surface stiffness tensor, and the Gaussian curvature of the surface.

6.3.4 Roughness of a Single Step

In this subsection, we consider the roughness of a single step on a surface when $T<T_{\text{R}}$. As shown in Figure 6.9 [6] for the Ising model, we will consider a step for which the mean running direction is $\overline{\text{OP}}$. We assign the height of the step $h(x)$ in the vertical

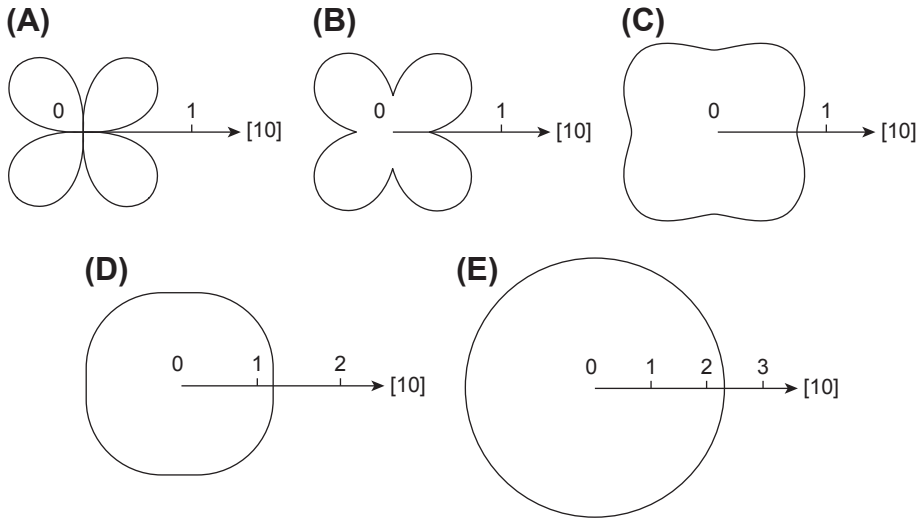


FIGURE 6.12 Polar graphs of $w_{\text{step}}(\phi)\sqrt{\cos\phi} = W_{\text{step}}(\phi)\sqrt{\cos\phi/L}$ for the 2D square-lattice Ising model. The temperatures are chosen as (A) $T/T_c = 0.1$; (B) $T/T_c = 0.3$; (C) $T/T_c = 0.5$; (D) $T/T_c = 0.7$; and (E) $T/T_c = 0.9$. *From Ref. [6].*

direction. Similar to Eqn (6.17), we define the width of the step as the variance of the height of the step, as follows:

$$W_{\text{step}}^2 = \left\langle [h(x) - \langle h(x)\rangle]^2 \right\rangle. \tag{6.30}$$

We can then obtain, for $T < T_{\text{R}}$:

$$W_{\text{step}}^2 = (k_{\text{B}}T/\tilde{\gamma})L, \quad \tilde{\gamma} = \gamma + \frac{\partial^2\gamma}{\partial\phi^2} \tag{6.31}$$

for an isolated step, where $\tilde{\gamma}$ represents the step stiffness, and ϕ represents tilt angle of the step. Hence, it is natural to define the step roughness as $w_{\text{step}} = W_{\text{step}}/\sqrt{L}$. In Figure 6.12, examples are shown of w_{step} for a 1D interface in the 2D Ising model [6]. The 1D interface in the 2D Ising model is considered to be an approximate model of a step on a (001) surface of a cubic Kossel crystal (Figure 6.2). As seen from Figure 6.12, the roughness of a step is strongly anisotropic at low temperatures. This is because of the degeneracy of the ground states with respect to the configuration of a step.

We can obtain a relationship similar to Eqn (6.24) between the step stiffness, the step roughness, and the curvature of the 2D ECS [6,7]:

$$w_{\text{step}}^2 = \frac{k_{\text{B}}T}{\tilde{\gamma}} = \frac{k_{\text{B}}T}{\lambda}\kappa, \tag{6.32}$$

where κ represents the curvature at a point of the 2D ECS.

It should be noted that the variance of a step in the *vicinal surface* is different from Eqn (6.31) but becomes as follows [103–105]:

$$W^2_{\mathrm{step,vicinal}} = \begin{cases} \dfrac{k_\mathrm{B}T}{2\pi a_z^2 \rho^2 \sqrt{\det(f^{ij})}} \ln L & (L \to \infty) \\ \dfrac{1}{2\pi^2 \rho^2} \ln L & (\rho \to 0, L \to \infty) \end{cases}, \tag{6.33}$$

where ρ represents the step density on a vicinal surface. Since a single step collides with the adjacent steps, the step width here diverges more weakly than for an isolated step. In this case, the squared "roughness of a step" $w^2_{\mathrm{step,vicinal}}$ should be defined as $w^2_{\mathrm{step,vicinal}} = W^2_{\mathrm{step,vicinal}}/\ln L$. We will return to this below in Section 5.3.

6.4 Roughening Transition and Faceting Transition as Critical Phenomena

6.4.1 Microscopic Models for Studying Surface Roughness

In this subsection, we will demonstrate some typical microscopic models for studying the surface roughness of the (001) surface of a crystal (Figure 6.13). The known roughening transition temperatures T_R are listed in Table 6.2. Since the transition temperature T_c of the 2D Ising model gives a lower bound for T_R, T_c for several lattice structures are also given in Table 6.2.

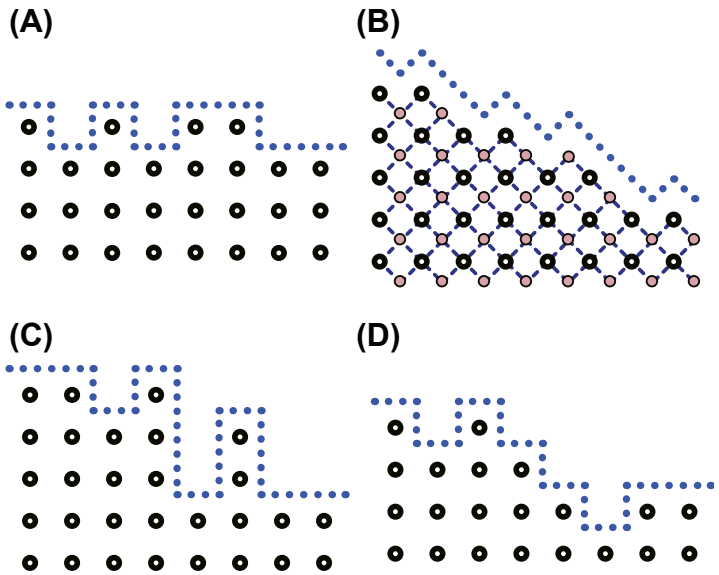


FIGURE 6.13 Side view of surfaces. (A) 2D lattice gas model equivalent to the 2D Ising model, (B) BCSOS model, (C) ASOS model, (D) RSOS model.

Table 6.2 Phase Transition Temperatures

Model	T_R or T_c	Hamiltonian	Remarks
2D Ising model	$k_B T_c/J = 2/\ln(1+\sqrt{2})$ ≈ 2.269	Equation (6.12)	Exact. Square lattice.
	$k_B T_c/J = 2/\cos h^{-1}(2)$ ≈ 1.519	Equation (6.12)	Exact. Honeycomb lattice [106].
	$k_B T_c/J = 4/\ln(3)$ ≈ 3.641	Equation (6.12)	Exact. Triangular lattice [106].
	$k_B T_c/J = 4/\ln(3+2\sqrt{3})$ ≈ 2.143	Equation (6.12)	Exact. Kagome lattice [106].
	$k_B T_c/J = 4/\ln(3)$ ≈ 2.405	Equation (6.12)	Exact. Diced lattice [106].
3D Ising model	$T_R > T_{c,2D}$	Equation (6.12)	
BCSOS model	$k_B T_R/\varepsilon = 1/\ln 2$ ≈ 1.443	Equation (6.34)	Exact [1].
ASOS model	$k_B T_R/\varepsilon \approx 1.21$	Equation (6.37)	Square lattice [2].
RSOS model	$k_B T_R/\varepsilon \approx 1.580$	Equation (6.38)	Square lattice [3,107].
Discrete Gaussian model	$k_B T/J \approx 1.44$	Equation (G.1)	Square lattice [2].

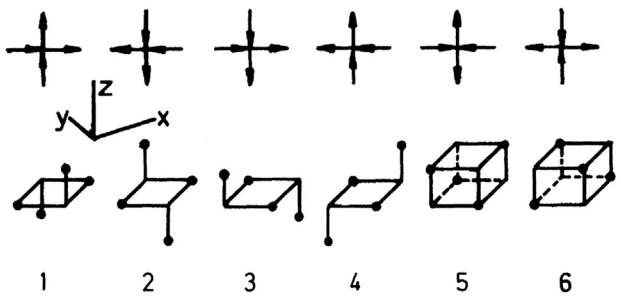


FIGURE 6.14 Mapping of the body-centered cubic solid-on-solid model to a 6-vertex model [1]. *From Ref. [1].*

6.4.1.1 BCSOS Model

In Figure 6.13(B), the BCSOS model [1] is shown. The BCSOS model is a microscopic model of the (001) surface of the body-centered cubic crystal. This model is usually described by the 6-vertex model (Figure 6.14). The Hamiltonian is as follows:

$$\mathcal{H}_{\mathrm{BCSOS}} = \varepsilon_1/2 \sum_{(n,m)} \sum_{\delta} [|h_A(n,m) - h_B\{(n,m)+\delta\}|] + \varepsilon_2 \sum_{\langle i,j \rangle} \left[|h_{A,i} - h_{A,j}| + |h_{B,i} - h_{B,j}| \right], \tag{6.34}$$

where ε_1 represents half of the nn bond energy, and ε_2 represents half of next nn (nnn) bond energy. Since the crystal structure of the model is body-centered cubic (BCC), the lattice points can be divided into two cubic lattices, the A-sublattice and B-sublattice. We will distinguish the sites in the A- and B-sublattices by adding an A or B to the variables. The summation (n,m) is taken over all A-sublattice points, and the summation $\langle i,j \rangle$ is taken over all pairs of the nn sublattice points.

In the BCSOS model, ε_1 is set to infinity in order to avoid the overhang structure. Hence, the height difference between the nn sites, $|\Delta h| = |h_{A,i} - h_{B,j}|$, is fixed to be 1/2. The partition function of the BCSOS model Z_{BCSOS} is defined as

$$Z_{\mathrm{BCSOS}} = \sum_{\{h_A(m,n)\}} \sum_{\{h_B(m,n)\}} \exp[-\beta \mathcal{H}_{\mathrm{BCSOS}}], \tag{6.35}$$

where $\beta = 1/k_{\mathrm{B}}T$, and the summation with respect to $\{h_A(m,n)\}$ and $\{h_B(m,n)\}$ means the summation over all possible surface configurations. This partition function is usually calculated for the equivalent 6-vertex model (Figure 6.14), and it was obtained exactly by van Beijeren [1]. The surface tension $\gamma_{\mathrm{surf}}(0)$ and the surface free energy per area $f_{\mathrm{surf}}(0)$ for the (001) surface are calculated as follows:

$$\gamma_{\mathrm{surf}}(0) = f_{\mathrm{surf}}(0) = \lim_{\mathcal{N}\to\infty} \frac{k_{\mathrm{B}}T}{\mathcal{N}a^2} \ln Z_{\mathrm{BCSOS}}, \tag{6.36}$$

where a represents the lattice constant of a cubic lattice. As mentioned in Section 1, the exact calculations for the BCSOS model greatly contributed to the understanding of roughening and faceting transitions.

6.4.1.2 ASOS Model and RSOS Model

Figure 6.13(C) shows a side view of the absolute SOS (ASOS) model [89]. A perspective view of the ASOS model was already shown in Figure 6.1(A). This ASOS model is a natural model and is based on the Kossel crystal [108]. The height variable $h(x,y)$ takes an integer. The Hamiltonian of the ASOS model is given by

$$\mathcal{H}_{\mathrm{ASOS}} = \varepsilon \sum_{m,n} [|h(m+1,n) - h(m,n)| + |h(m,n+1) - h(m,n)|], \tag{6.37}$$

where ε represents the microscopic energy cost required to make the nearest-neighbor (nn) height difference Δh. The summation of (m,n) covers all lattice points on the square lattice. Since ε is half of the lateral bond energy, ε corresponds to $2J$ in the 2D Ising model.

Figure 6.13(D) shows a side view of the restricted SOS (RSOS) model [109], where "restricted" means that the nn height difference is restricted to $\{0,\pm 1\}$. The Hamiltonian of the RSOS model is similar to that of the ASOS model, as follows:

$$\mathcal{H}_{\mathrm{RSOS}} = \varepsilon \sum_{m,n} [|h(m+1,n) - h(m,n)| + |h(m,n+1) - h(m,n)|]. \tag{6.38}$$

The RSOS restriction is implicitly assumed.

The partition function of the ASOS model Z_{ASOS} and the RSOS model Z_{RSOS} are defined in a way similar to Eqn (6.35). The exact solutions for these models, however, have not yet been obtained.

6.4.2 Summary of KT-Type Critical Phenomena of Surfaces

According to Knops' correspondence [10,11] that was based on the duality arguments, T^*_{KT} is related to the roughening transition temperature T_{R}; the quasi-long-range ordered

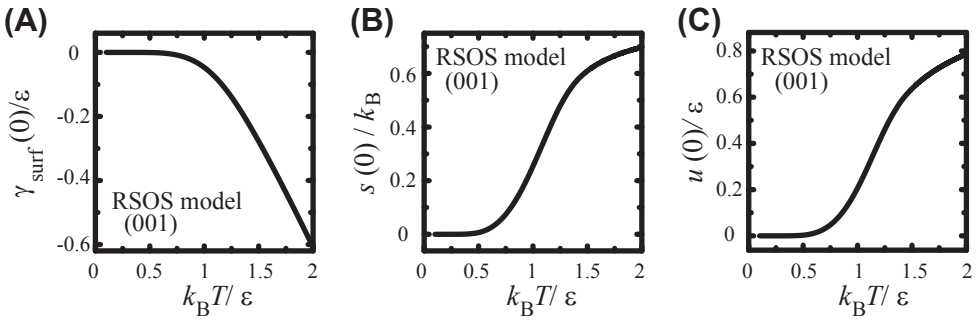


FIGURE 6.15 Thermodynamic quantities of the (001) surface of a cubic lattice: RSOS model and PWFRG calculations (Appendix A). $k_BT_R/\varepsilon = 1.580$. (A) Surface free energy per unit cell area. (B) Surface entropy per unit cell area. (C) Surface internal energy per unit cell area [110]. *From Ref. [110].*

phase for $T^* < T^*_{KT}$ in the XY model corresponds to the rough phase for $T > T_R$ in the surface model, and the phase for $T^* > T^*_{KT}$ in the XY model corresponds to the smooth phase for $T < T_R$. The transition temperature of the XY model is given by the zero of the free energy for a single vortex creation as follows: $k_BT^*_{KT}/J = 1/K_R = \pi/2$ [11,40].[18] The surface structure corresponding to the vortex in the XY model is not easily seen.

The features of the roughening transition are listed below [18,22,25,26]:

1. Correlation length

$$\xi(T) = \begin{cases} \infty & T \geq T_R \\ \xi_0 \exp\left(\dfrac{A}{\sqrt{T_R - T}}\right) & T < T_R \end{cases}, \tag{6.39}$$

   where T_R (Table 6.2), ξ_0, and A are nonuniversal constants.
2. Surface tension: The singular part of the surface tension becomes

$$\gamma_{surf,sing} \approx B \exp\left(-\frac{C}{\sqrt{|T - T_R|}}\right), \tag{6.40}$$

   where B and C are nonuniversal constants. In Figure 6.15(A), we show the surface tension of the (001) surface for the RSOS model. We also show the surface entropy and the surface internal energy in Figure 6.15(B) and (C), respectively. The singularity around T_R, however, is too subtle to be discerned. The steep increase of the surface entropy occurs near at $T_{c,Ising}$[19] of the 2D Ising model. This increase of the surface entropy results from the increase of the kink density on the surface.

[18] J represents the coupling constant of the XY model.
[19] $k_BT_{c,Ising}/\varepsilon = 1/\ln(1+\sqrt{2}) \sim 1.135$.

3. Height–height correlation function:

$$G(\boldsymbol{r}) = \begin{cases} \dfrac{k_B T}{\pi\sqrt{\det(f^{ij})}} \ln r & T > T_R \\ (a_z^2 K_R/\pi)\ln r & T = T_R \\ 2w^2 & T < T_R \end{cases} \quad (d = 3), \tag{6.41}$$

where w^2 represents a finite value calculated by Eqn (6.18).

4. Step free energies: Near T_R, the step free energy behaves as follows:

$$f_{step}(T) = \begin{cases} 0 & T \geq T_R \\ f_{step,0} \exp\left(-\dfrac{A}{\sqrt{T_R - T}}\right) & T < T_R \end{cases} \tag{6.42}$$

where $f_{step,0}$ and A are nonuniversal constants. For the BCSOS model, $A = \pi^2/(4\sqrt{2}\ln 2) \approx 2.52$ [18]. Since the 2D ECS obtained by the step tension coincides with the facet shape of the ECS [24], a decrease in the step tension leads to a faceting transition (Figure 6.3).

In Figure 6.16, we show the step tension γ_{step} and the step stiffness $\tilde{\gamma}_{step}$. As expected, the behavior of step tension at high temperature is well described by Eqn (6.42). Unexpectedly the value of A for the BCSOS model could fit the data for the RSOS model, though the value of A is thought to be nonuniversal.

We also show the step entropy and the step internal energy for the RSOS model in Figure 6.17. Both the step internal energy and the step entropy reduce to zero in the manner of Eqn (6.42). Here, a step entropy of zero does not mean a smooth surface. From Eqn (6.32), a step stiffness of zero causes a divergence in the step

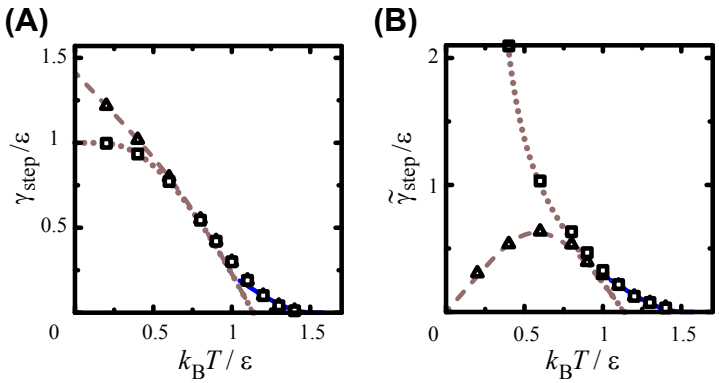


FIGURE 6.16 Thermodynamic quantities of a step on a (001) surface [110]. (A) Step tension. (B) Step stiffness. Open squares: $\phi = 0$,(01) step. Open triangles: $\phi = \pi/4$,(11) step. Dotted line: 1D interface for the 2D nn Ising model $\phi = 0$,Broken line: 1D interface for the 2D nn Ising model $\phi = \pi/4$. Solid line: (A) $\gamma_{step}/\varepsilon = 6 \times \exp(-2.52/\sqrt{1.580 - k_B T/\varepsilon})$; (B) $\tilde{\gamma}_{step}/\varepsilon = 8\exp(-2.52/\sqrt{1.580 - k_B T/\varepsilon})$. $k_B T_R/\varepsilon$ is assumed to be 1.580. *From Ref. [110].*

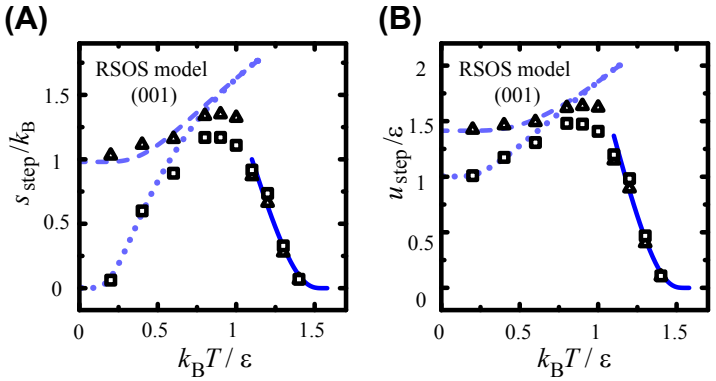


FIGURE 6.17 Thermodynamic quantities of a step on the (001) surface [110]. (A) Step entropy. (B) Internal energy of a step. Open squares: $\phi = 0$ ((01) step). Open triangles: $\phi = \pi/4$ ((11) step). Dotted line: 1D interface for the 2D nn Ising model $\phi = 0$. Broken line: 1D interface for the 2D nn Ising model $\phi = \pi/4$. Solid line: (A) $s_{\mathrm{step}}/k_{\mathrm{B}} = 38\exp(-2.52/\sqrt{1.580 - k_{\mathrm{B}}T/\varepsilon})$; (B) $u_{\mathrm{step}}/\varepsilon = 52\exp(-2.52/\sqrt{1.580 - k_{\mathrm{B}}T/\varepsilon})$. $k_{\mathrm{B}}T_{\mathrm{R}}/\varepsilon$ is assumed to be 1.580. *From Ref. [110].*

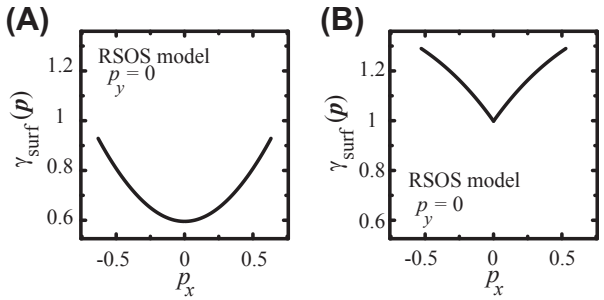


FIGURE 6.18 Profile of surface tension: RSOS model and PWFRG calculations (A) $k_{\mathrm{B}}T/\varepsilon = 1.7$ ($T > T_{\mathrm{R}}$). (B) $k_{\mathrm{B}}T/\varepsilon = 0.6$ ($T < T_{\mathrm{R}}$).

roughness (the scaled step width) w_{step} at T_{R}. Hence, the steps dissolve into the bulk, as will be discussed in the following Section 4.3.2 for the 2D Ising model. Since the step free energy is defined as the excess part of the free energy from the bulk (Section 3.1.3), the step free energy decreases to zero as the temperature increases to T_{R}.

5. Surface free energy: Just above T_{R}, the vicinal surface free energy density and the Andreev free energy become as follows (Figure 6.18(A)) [18]:

$$f(\boldsymbol{p}) = f(0) + \frac{k_{\mathrm{B}}T_{\mathrm{R}}}{2K_R a_z^2}|\boldsymbol{p}|^2 + \mathcal{O}\left(|\boldsymbol{p}|^3\right), \tag{6.43}$$

$$\tilde{f}(\boldsymbol{\eta}) = f(0) - \frac{a_z^2 K_R}{2k_{\mathrm{B}}T_{\mathrm{R}}}|\boldsymbol{\eta}|^2 + \mathcal{O}\left(|\boldsymbol{\eta}|^3\right), \tag{6.44}$$

where $K_R = 2/\pi$ and $f(0) = \gamma_{\mathrm{surf}}(0)$. Hence, we have,

$$|\boldsymbol{\eta}| = \frac{k_B T}{K_R a_z^2} |\boldsymbol{p}|. \tag{6.45}$$

Using this equation, we can determine T_R numerically [3].

6. Universal curvature jump: The faceting transition on the ECS (Figure 6.3) accompanies a jump in the Gaussian curvature at the faceted surface as follows:

$$K_G = \begin{cases} \left(\frac{\lambda a_z^2}{k_B T_R}\right)^2 K_R^2 & T = T_R \\ 0 & T < T_R \end{cases}. \tag{6.46}$$

The Gaussian curvature is expressed as $K_G = \kappa_1 \kappa_2$, where κ_1 and κ_2 are the principal values of the curvature at that point on the surface.

7. Universal jump in the surface stiffness tensor: The faceting transition also accompanies a jump in the determinant of the surface stiffness tensor at the faceted surface as follows:

$$\det(f^{ij}) = \begin{cases} \left(\frac{k_B T_R}{a_z^2}\right)^2 \left(\frac{1}{K_R}\right)^2 & T = T_R \\ \infty & T < T_R \end{cases}. \tag{6.47}$$

6.4.3 Diffuseness for Atomically Rough Surfaces

6.4.3.1 New Picture of Roughening

Since the step tension becomes zero at $T \geq T_R$(Eqn (6.42)), surface roughening is said to occur as a result of the step proliferation without any excess cost of free energy (Figure 6.19). According to this picture, some finite structures of the excited states, such as the adatoms and islands that form on the surface, do not significantly contribute to the free energy of the surface (i.e., they are irrelevant).

As shown in Figure 6.6, however, a surface simulated by Monte Carlo calculations with an ASOS model at high temperature is slightly different from the picture in Figure 6.19. The surface appears rough in the small scale. Hence, the terms *surface diffuseness* or *atomically rough surface* have been used in order to describe the roughness in the small scale. This difference is considered to be important in dynamic phenomena such as crystal growth, but it seems that the problem has not yet been clarified

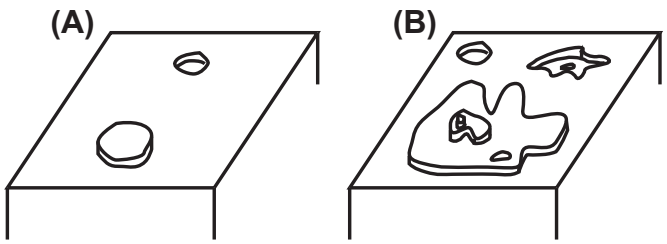


FIGURE 6.19 Proliferation of steps. (A) $T < T_R$. (B) $T > T_R$.

sufficiently. However, the following two approaches may help us to connect an atomically rough structure and the rough phase with drumhead wandering.

One approach is to study the intrinsic interface width [4,111,112], and the other approach is to study the preroughening phenomena [113]. We will discuss them briefly in the following subsubsections.

6.4.3.2 Intrinsic Width of a Step

To understand the microscopic structure of a phase separation line (1D interface) in the 2D Ising model, Bricmont, Lebowitz, and Pfister [111] presented the notion of "deformations." They decomposed a 1D interface into a bone-line and "blobs." In Figure 6.20(A), the blobs are shown. The blobs are replaced by kinks of size D (Figure 6.20(B)). The kinks are deformations in the phase separation line. Then, d_n, which is n-th moment of $|D|$, is defined as follows [112]:

$$d_n = \lim_{N_x \to \infty} \frac{1}{N_x} \sum_i < |D_i|^n > , \tag{6.48}$$

where N_x represents the number of lattice points in the direction of the bone-line. Based on the results of a Monte Carlo study [112], d_1 is kept finite for $T \leq T_c$ and d_2 coincides with $w^2_{\text{step}} = k_B T/\tilde{\gamma}$. Since $\tilde{\gamma} \propto T_c - T$, d_2 and w^2_{step} diverge as $[(T_c - T)/T_c]^{-1}$. Therefore, due to the effect of the local structures of the small domains of flipped spins (Figure 6.9), both W_{step} and w_{step} diverge at T_c.

For a surface step, we expect a similar divergence. In this case, however, $\tilde{\gamma} \propto \exp[-A/\sqrt{T_R - T}]$, and then from Eqn (6.32), w^2_{step} diverges as $\exp[A/\sqrt{T_R - T}]$. Therefore, when $T = T_R$, steps roughen the surface, and, at the same time, the steps dissolve into the surface. The excited structures such as adatoms and islands contribute to the values of the step tension and step stiffness.

6.4.3.3 Preroughening Phenomena

Den Neijs and Rommels [113] predicted *preroughening phenomena* on surfaces where a short-range step–step repulsion exists.

When the temperature is near T_c for the 2D Ising model, the relatively large blobs are excited. Due to repulsion, the steps are kept apart from other steps. As a result, when the temperature is in the region $T_c < T < T_R$, the surface seems to be rough, but the steps are confined to two levels on the surface. This phase was named the "disordered flat phase"

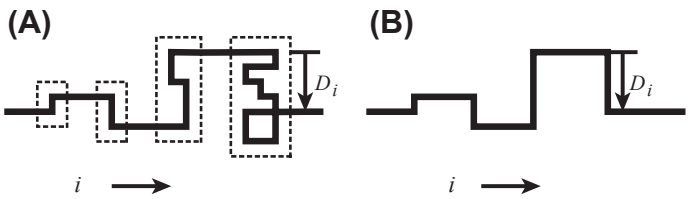


FIGURE 6.20 Schematic diagram of deformations. (A) An example of a 1D interface (phase separation line) in a 2D Ising model. The configurations in the dotted squares are "blobs" [4,111,112]. A blob at the site i is replaced by a kink, which is called a deformation with size D_i. (B) 1D converted interface with kinks.

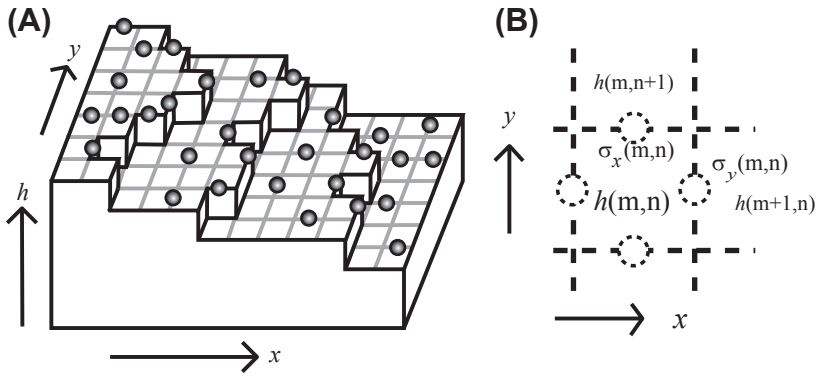


FIGURE 6.21 A "decorated" RSOS model for the Langmuir adsorption; the coverage is less than 1 [107]. (A) Perspective view. (B) Adsorption site. *From Ref. [107].*

(DFP) by den Neijs and Rommels. The DFP seems to be a candidate for the atomically rough phase.[20]

6.4.4 Changes in the Roughening Temperature on Complex Surfaces

6.4.4.1 Inverse Roughening

In 1994, Luijten et al. [114] predicted the "inverse roughening" phenomena on the (001) surface of a CsCl-type crystal structure. The model that Luijten et al. adopted is the staggered BCSOS model, which has a uniform field (stoichiometric chemical potential) that favors the Cs component on the surface of the BCSOS model. In the inverse roughening phenomena, the surface is rough when the temperature is zero; as the temperature increases, it becomes smooth; with further increases, the step tension begins to decrease; finally, at sufficiently high temperature, a second transition occurs and the surface again becomes rough [33,114].

Using this type of model with different statistical weights for the vertices, the disordered flat phase [115–117] and the reconstructed rough phase [116,117] can also be studied. Details of these models are not discussed in this article.

6.4.4.2 Surface Modified by Langmuir Adsorption

Adsorbed materials on a surface are empirically known to change the thermodynamic behavior of that surface [94]. Using a lattice model named the "decorated" RSOS model (Figure 6.21), Akutsu et al. [107] showed the change of surface tension in a case of Langmuir adsorption. They also showed that Langmuir-type adsorbed materials with coverage less than one changes the roughening transition temperature. In this subsection, we explain how the roughening transition temperature is changed by a small amount of adsorbed materials.

[20]The DFP can be translated to an $S = 1/2$ quantum spin system as the resonating valence bond (RVB) state relating to the Haldane gap.

The Hamiltonian of the decorated RSOS model consists of an RSOS model ($\mathcal{H}_{\mathrm{RSOS}}$), a lattice gas model for adsorption ($\mathcal{H}_L$), and their coupling parts, as follows:

$$\mathcal{H} = \mathcal{H}_{\mathrm{RSOS}} + \mathcal{H}_{\mathrm{L}} + \mathcal{H}_{\mathrm{int}}, \tag{6.49}$$

where

$$\begin{aligned} \mathcal{H}_{\mathrm{RSOS}} + \mathcal{H}_{\mathrm{int}} = \sum_{m,n} \{ & \varepsilon\big[1 - \alpha\sigma_y(m,n)\big] \cdot |h(m+1,n) - h(m,n)| \\ & + \varepsilon[1 - \alpha\sigma_x(m,n)] \cdot |h(m,n+1) - h(m,n)| \} \\ \mathcal{H}_L = -H \sum_{m,n} & \big[\sigma_x(m,n) + \sigma_y(m,n)\big]. \end{aligned} \tag{6.50}$$

$\mathcal{H}_L$ is expressed by the Ising spin variables $\sigma = \pm 1$.[21] The spin variables are put on the bridge site, as shown in Figure 6.21(B), because we consider that an adsorbed material on the RSOS ledge changes the microscopic ledge energy. The parameter α in $\mathcal{H}_{\mathrm{int}}$ describes the coupling between the RSOS model and the lattice gas model of the adsorbed materials. For $\alpha > 0$, the adsorbed materials favor the edge of the step, while for $\alpha < 0$, the adsorbed materials favor the terrace sites.

The partition function is calculated as $Z = \sum_{\{h_i\}} \sum_{\{\sigma_i\}} \exp[-\beta\mathcal{H}]$, where $\beta = 1/k_{\mathrm{B}}T$. After taking the partial sum with respect to $\{\sigma_i\}$, we obtain the exact RSOS Hamiltonian with the effective ledge energy $\varepsilon^{\mathrm{eff}}$, as follows:

$$\varepsilon^{\mathrm{eff}}(T,H) = \varepsilon - k_{\mathrm{B}}T \ln\left[\frac{\cosh(\beta\alpha\varepsilon + \beta H)}{\cosh \beta H}\right]. \tag{6.51}$$

The mean coverage of the materials of adsorption at the step ledge C_y is given by the following equation:

$$M_y = 2C_y - 1 \approx \tanh \beta H + \big[\tanh \beta\big(H + \varepsilon\alpha\big) - \tanh \beta H\big] p_x, \tag{6.52}$$

where M_y represents the mean magnetization of $\langle\sigma_y\rangle$. Equation (6.51) says that the effective ledge energy depends on the temperature and the chemical potential of the adsorbed materials in the ambient phase.

Since the roughening temperature is given by $k_{\mathrm{B}}T_{\mathrm{R}}/\varepsilon^{\mathrm{eff}} = 1/\zeta_R \approx 1.580$ for the RSOS model, where $\zeta_R = 0.6330$, we have

$$\frac{H}{k_{\mathrm{B}}T_{\mathrm{R}}} = \frac{1}{2} \ln \frac{\exp[\zeta_R - (1+\alpha)\varepsilon/(k_{\mathrm{B}}T_{\mathrm{R}})] - 1}{1 - \exp[\zeta_R - (1-\alpha)\varepsilon/(k_{\mathrm{B}}T_{\mathrm{R}})]}, \tag{6.53}$$

which allows us to draw the critical line in the $H - T$ plane, as shown in Figure 6.22. In the case of $\alpha \approx 1$, there is a region of H where the inverse roughening occurs. We can see the region of H where the inverse roughening occurs in Figure 6.22(B) and (C).

[21] To obtain the expression for the lattice gas variables, replace σ with $2C - 1$ as in section 3.1.2. $H = \mu/2 = (k_{\mathrm{B}}T/2)\ln(P/P_e)$ where P represents the vapor pressure of the adsorbed materials and P_e represents the vapor pressure of the adsorbed materials with the coverage being 50%.

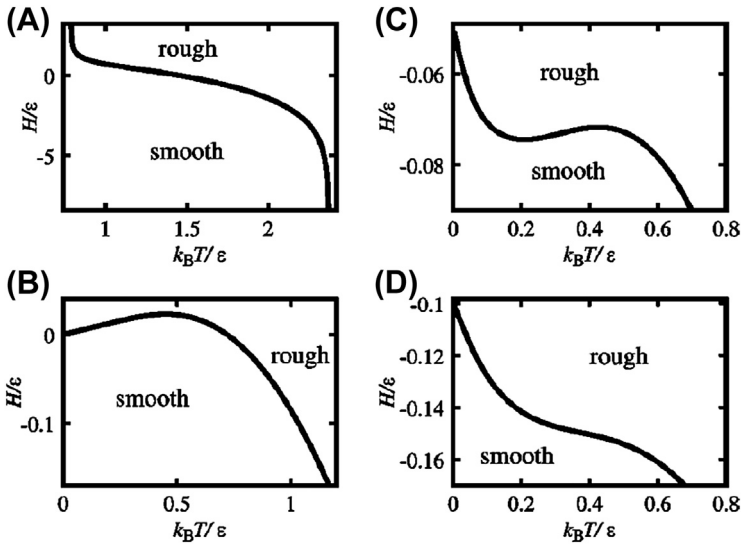


FIGURE 6.22 Change of the roughening transition temperature. (A) $\alpha = 0.1$. (B) $\alpha = 1.0$. (C) $\alpha = 1.1$. (D) $\alpha = 2.0$. *From Ref. [107].*

6.5 Vicinal Surface

6.5.1 Rough or Smooth? The Terrace-Step-Kink Picture

As mentioned in Section 1, studying the vicinal surface by using the BCSOS model [18] gave us a correct understanding of the faceting transition. In addition, the vicinal surface for $T < T_R$ is itself interesting because the system can be mapped to a 1D free fermion (FF) system by using a terrace-step-kink (TSK) picture of the vicinal surface. This characteristic of the 1D FF, as in Eqn (6.1), is the GMPT [13] type of universal character.

A vicinal surface is a slightly tilted surface from a plane with a low Miller index. Figure 6.23 shows a typical vicinal surface for which the temperature T is lower than T_R of the terrace surface. As mentioned in the previous section, the excited structures such as adatoms and islands contribute to changes in the step tension and the step interaction coefficients of the step.[22] Hence, in the mesoscopic scale, a vicinal surface is well described by terraces, steps, and kinks in a step. This is called the TSK picture.

Since the terrace is smooth, a vicinal surface seems to be a smooth surface. In fact, as mentioned in the work of BCF [87], the growth rate of a surface that has a regular train of steps is different from that of a surface with $T > T_R$. The linear dependence on the driving force $\Delta\mu$ concerning the growth rate, however, is similar to a rough surface. The vicinal surface grows in the "step flow growth" mode, and the growth rate of a vicinal surface is proportional to $\Delta\mu$. According to the definition of the surface roughness in Section 3.2.1,

[22]Recall that islands and negative islands on a terrace are irrelevant.

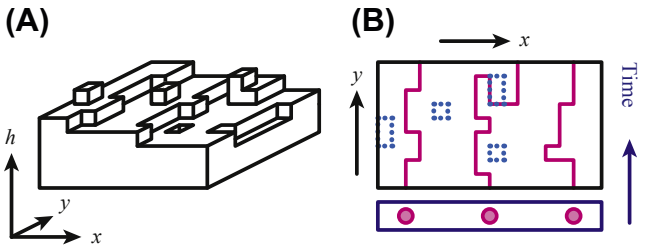


FIGURE 6.23 (A) A perspective view of a vicinal surface. (B) A terrace-step-kink picture of a vicinal surface. Solid line: steps on the vicinal surface shown in (A) (top view). Dotted rectangles: islands or negative islands on the surface.

the vicinal surface is rough, because the variance of the surface height diverges logarithmically as the area L^2 increases (Eqn (6.28)).

Therefore, we can say that a vicinal surface has characteristics of both a rough surface and a smooth surface. Statically, a vicinal surface has the characteristics of a rough surface; dynamically, it has the characteristics of a smooth surface.

6.5.2 1D Free-Fermion Universal Features—Gruber-Mullins-Pokrovsky-Talapov Behavior

In the TSK picture (Figures 6.4 and 6.23), the steps can be regarded as linear excitations buried in a 2D planar lattice. Since overhang structures are inhibited in SOS models, the linear excitations are impenetrable by adjacent excitations. Due to the impenetrability, the vicinal surface can be exactly described by a 1D spinless free fermion (FF) at temperature $T = 0$ [26]. The zigzag structure of a step corresponds to the quantum zero-point oscillations. Hence, as previously shown in Eqn (6.1), the free energy of the vicinal surface $f(\rho)$ can be obtained by the ground-state energy of the 1D FF, as follows:

$$f(\rho) = f(0) + \gamma_{\text{step}}\rho + B\rho^3 + \cdots, \tag{6.54}$$

where ρ represents the step density, $f(0) = \gamma_{\text{surf}}(0)$ represents the surface tension of the terrace plane, γ_{step} represents the step tension, and B represents the step interaction coefficient. In the case of the 1D FF [25,26],

$$B = \frac{\pi^2}{6} k_{\text{B}} T w_{\text{step}}^2. \tag{6.55}$$

Recalling Eqn (6.32), we have the following relationship:

$$\beta B = \frac{\pi^2}{6\beta\tilde{\gamma}}, \tag{6.56}$$

where $\beta = 1/k_{\text{B}}T$ and $\tilde{\gamma} = \gamma_{\text{step}} + \partial^2\gamma_{\text{step}}/\partial\phi^2$ represents the step stiffness. Substituting Eqn (6.56) into Eqn (6.54) with $|\boldsymbol{p}| = \rho a_z$, we have

$$f(\boldsymbol{p}) = \gamma_{\text{surf}}(0) + \gamma_{\text{step}}\frac{|\boldsymbol{p}|}{a_z} + \frac{\pi^2}{6}\frac{(k_{\text{B}}T)^2}{\tilde{\gamma}}\left(\frac{|\boldsymbol{p}|}{a_z}\right)^3 + \cdots. \tag{6.57}$$

The surface tension for the RSOS model is calculated from $f(\boldsymbol{p})$ as $\gamma_{\mathrm{surf}}(\boldsymbol{p}) = f(\boldsymbol{p})/\sqrt{1+|\boldsymbol{p}|^2}$ at $k_{\mathrm{B}}T/\varepsilon = 0.6$, and it is shown in Figure 6.18(B).

In addition, applying Eqn (6.57) to the stiffness tensor (Eqn (6.27)), we have another universal jump of $\det(f^{ij})$ at the facet edge, as follows [25,26]:

$$\det\left(f^{ij}\right) = \begin{cases} \left(\dfrac{2}{K_R}\right)^2 \left(\dfrac{k_{\mathrm{B}}T}{a_z^2}\right)^2 = \pi^2 \left(\dfrac{k_{\mathrm{B}}T}{a_z^2}\right)^2, & (|\boldsymbol{p}| \to 0) \\ \infty, & (|\boldsymbol{p}| = 0) \end{cases} \tag{6.58}$$

for small $|\boldsymbol{p}|$, where $K_R = 2/\pi$. Similarly, for the surface roughness and the Gaussian curvature on the ECS, we have a jump, as follows:

$$w_{\mathrm{surf}} = \begin{cases} \sqrt{\dfrac{K_R a_z^2}{4\pi}} = \dfrac{a_z}{\sqrt{2}\pi}, & (|\boldsymbol{p}| \to 0) \\ 0, & (|\boldsymbol{p}| = 0) \end{cases} \tag{6.59}$$

$$K_G = \begin{cases} \left(\dfrac{\lambda a_z^2}{k_{\mathrm{B}}T}\right)^2 \left(\dfrac{K_R}{2}\right)^2, & (|\boldsymbol{p}| \to 0) \\ 0, & (|\boldsymbol{p}| = 0) \end{cases} \tag{6.60}$$

Therefore, $\det(f^{ij})$, the Gaussian curvature, and the surface roughness are constant near a facet edge.

6.5.3 Logarithmic Behavior on the Width of a Single Step

6.5.3.1 Height–Height Correlation Function of a Single Step

As mentioned in Section 3.4, we have shown that the variance W^2_{step} of the "height" of a step on a surface is proportional to L, where L is the linear size of the system. In this subsubsection, we will show that the variance of the "height" of a single step on a vicinal surface is not proportional to L because the step collides with adjacent steps on the vicinal surface.

We now consider the height–height correlation function $\Delta^2(r)$ for a single step on a vicinal surface, as follows [103–105]:

$$\Delta^2(r) = \left\langle (x(y+r) - x(y))^2 \right\rangle, \tag{6.61}$$

where, without loss of generality, the y-axis can be chosen to lie along the mean running direction of the steps. Here we introduce the probability $Q(x,y)$ that two points (0,0) and (x,y) lie in the same terrace:

$$Q(x,y) = C\langle \delta(h(0,0) - h(x,y)) \rangle, \tag{6.62}$$

where C is a normalization factor. Using the capillary wave Hamiltonian (Eqn (6.26)), we obtain

$$Q(x,y) \propto \exp\left(\frac{-p^2x^2}{2G(x,y)}\right), \tag{6.63}$$

where $G(x,y)$ represents the height–height correlation function of the surface (Eqn (6.22)). Equation (6.63) shows that $Q(x,y)$ is identical to the probability that a step passes the two points (0,0) and (x,y). Using the probability in Eqn (6.63), Eqn (6.61) then becomes the following:

$$\Delta^2(r) = \int Q(x,y)x^2 dx = \frac{1}{p^2} G(0,r) \quad (r \to \infty). \tag{6.64}$$

Since $G(0,y) = [k_B T/(\pi\sqrt{\det(f^{ij})})]\ln r$ (Eqn (6.23)), we have the relation

$$\begin{aligned} \Delta^2(r) &= \left[k_B T \Big/ \left(\pi\rho^2 a_z^2 \sqrt{\det(f^{ij})}\right)\right] \ln r \\ &= \frac{1}{\pi\rho^2} \ln r \quad (\rho \to 0). \end{aligned} \tag{6.65}$$

The Eqn (6.65) on $\Delta^2(r)$ is also derived exactly in the limit $\rho \to 0$ for the TSK model on a lattice [103,104]. The variance of the height of a single step on the vicinal surface $W^2_{\text{step,vicinal}}$ becomes exactly [104].

$$\begin{aligned} W^2_{\text{step,vicinal}} &= \left[k_B T \Big/ \left(2\pi\rho^2 a_z^2 \sqrt{\det(f^{ij})}\right)\right] \ln L \\ &= \frac{1}{2\pi\rho^2} \ln L \quad (\rho \to 0). \end{aligned} \tag{6.66}$$

6.5.3.2 *Elastic Step–Step Repulsion*

On a crystal surface, elastic repulsion such as $g_0/(x_i - x_{i+1})^2$ [118,119] sometimes exists. Here, g_0 represents the coupling constant, and x_i represents the location of the i-th step (Figure 6.23(B)). By adding this term to the capillary wave Hamiltonian, we have the surface free energy for this system, as follows [103,120]:

$$f_{\text{elastic}}(p) = f(0) + \gamma(\phi)\frac{p}{a_z} + \frac{\pi^2}{6}\frac{(k_B T)^2}{\tilde{\gamma}(\phi)}\tilde{\lambda}^2(g)\frac{p^3}{a_z^3}, \tag{6.67}$$

where

$$g = 2\beta^2\tilde{\gamma}(\phi)g_0, \quad \tilde{\lambda}(g) = \left[1 + \sqrt{1+2g}\right]\Big/2. \tag{6.68}$$

Here, we used the exact result of Sutherland [121]. Therefore, we have this new relationship around the facet edge:

$$\det(f^{ij}) = \left[k_B T \pi \tilde{\lambda}(g)\right]^2, \quad G(0,r) = \frac{1}{\pi^2\tilde{\lambda}(g)}\ln r, \quad \Delta^2(r) = \frac{1}{\pi^2\rho^2\tilde{\lambda}(g)}\ln r. \tag{6.69}$$

The jump of the Gaussian curvature at the facet edge $\Delta K_G(g)$ becomes

$$\Delta K_G(g) = \Delta K_G(0)/\tilde{\lambda}^2(g) = \frac{1}{\tilde{\lambda}^2(g)}\left(\frac{\lambda a_z^2}{k_B T}\right)^2\left(\frac{K_R}{2}\right)^2. \tag{6.70}$$

6.6 Step Faceting

6.6.1 Stability of a Macrostep

In this section, we discuss the faceting of a macrostep.

Since faceted steps [122] are often observed on real surfaces, it seems natural to consider the roughening transition of an isolated step. Empirically, some impurities or materials of adsorption are known to induce faceted steps, and surface reconstructions are also known to do so [94,123]. As mentioned in Section 2.3, however, the drumhead wandering of a 1D surface (a step) activated by thermal fluctuations destabilizes a smooth surface. Hence, a step on a 2D surface is always rough when $T < T_R$, due to drumhead wandering. This seems to be contradicted by the observation of faceted steps in equilibrium.

Cabrera [62,63] presented the special polar angular dependence of the surface tension and its profile for the ECS[23] (Figure 6.24). Cabrera and Coleman [62] discussed the stability of a macrostep by assuming the surface free energy shown in Figure 6.24. If the surface free energy has the Type-II slope dependence, as shown on the left-hand side of Figure 6.24, a macrostep stabilizes as the two surfaces coexist under equilibrium.

The profile of the Type-II ECS, Figure 6.24, is called a first-order shape transition [20,34–36]. Rottman and Wortis [34] showed the first-order shape transition on the ECS for the 3D cubic Ising models with the negative second nearest-neighbor interaction using mean field approximation. Jayaprakash and Saam [20] showed the first-order shape transition on the ECS for the fcc lattice models with the negative second-nearest-neighbor interaction, using the mean-field approximation. Jayaprakash et al. [35] studied the vicinal surface between the (100) surface and the (110) surface by using a TSK model. Assuming an attractive step–step interaction, such as $g_0/(x_i - x_j)^2$, they obtained the first-order shape transition between the (100) surface and the (110) surface using the mean-field approximation. In the 1980s, the phase transition on the ECS attracted attention, but not much attention was paid to the stability of macrosteps.

Recently, various microscopic models that show an ECS similar to Type II in Figure 6.24 have been presented; these include the p-RSOS model [86,96,124,125], the RSOS-I model [126–128], and a modified RSOS-I model [129]. The p-RSOS model (Figure 6.25), explained in the following subsection, is the RSOS model with a point-contact type of step–step attraction. This type of step–step attraction works at the meeting point of adjacent steps. It was introduced to describe the energy gain that arises from the formation of a bonding state between steps due to the overlapping of orbitals. The RSOS-I model (the RSOS model coupled with the Ising system; see Appendix G.2) is used to study the effects of adsorption on the surface thermodynamic quantities. The

[23] In Figure 6.24, $\beta(\boldsymbol{p})$ represents a vicinal surface free energy $f(\boldsymbol{p})$ of this article. Later, the shape shown as type II in Figure 6.24 was called a "first-order shape transition" [22,36,57] because the surface slope jumps at the facet edge. The surface slope is given by the first derivative of the vicinal surface free energy with respect to η (Eqn (6.25)), and at the facet edge, it converges to p_1 from the left but converges to 0 from the right.

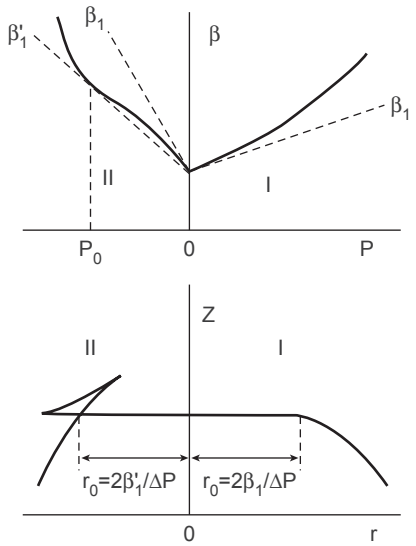


FIGURE 6.24 Top figure: the vicinal surface free energy denoted by β as a function of $p = p_x$. Type I (right-hand side of the figure): the case of $\partial^2\beta/\partial p^2 > 0$. Type II (left-hand side of the figure): the case of $\partial^2\beta/\partial p^2 < 0$. Bottom figure: The profile of the ECS. Type I (right-hand side) and Type II (left-hand side). *From Ref. [63].*

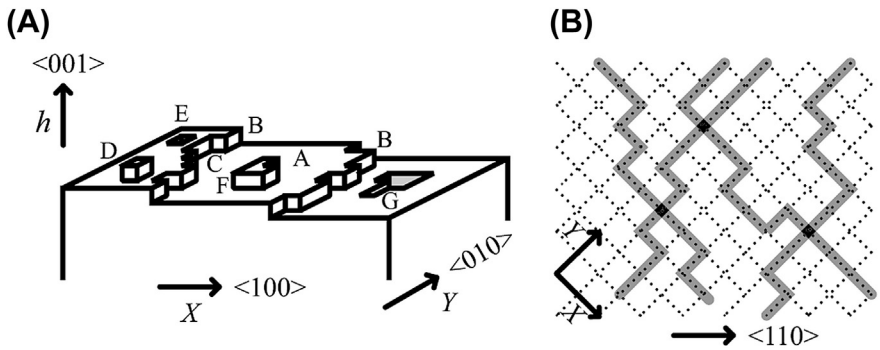


FIGURE 6.25 An example of a vicinal surface on the p-RSOS model. (A) Perspective view. (B) Top view. *From Ref. [86].*

modified RSOS-I model is the RSOS-I model in which a relaxation of stress at the step edge is taken into consideration.

The faceted steps are obtained in these models at low temperatures, where the two surfaces coexist at equilibrium. The key quantity for these phenomena was found to be the microscopic step–step attraction [86,96,124–128]. This attractive step–step interaction causes a discontinuity in the surface tension [96] at low temperatures. The temperature dependence of the phenomena is understood to be due to the competition

between the attractive energy between the steps and the entropy of the step wandering. These phenomena can also be understood as 1D Bose gas condensation at $T=0$ [86].

6.6.2 Discontinuous Surface Tension

6.6.2.1 p-RSOS Model

One of the microscopic models showing step faceting is the RSOS model with the point-contact type of step–step attraction (p-RSOS model) [86,96,124,125]. The Hamiltonian of the p-RSOS model is written as follows:

$$\begin{aligned}\mathcal{H}_{\mathrm{p-RSOS}} = &\sum_{i,j} \varepsilon[|h(i+1,j)-h(i,j)|+|h(i,j+1)-h(i,j)|] \\ &+\sum_{i,j} \varepsilon_{\mathrm{int}}[\delta(|h(i+1,j+1)-h(i,j)|,2) \\ &+\delta(|h(i+1,j-1)-h(i,j)|,2)],\end{aligned} \tag{6.71}$$

where ε is the microscopic ledge energy, $\varepsilon_{\mathrm{int}}$ is the microscopic step–step interaction energy, and $\delta(a,b)$ is the Kronecker delta. The summation with respect to (i,j) is performed over all of the sites on the square lattice. The RSOS restriction is required implicitly. In the case of $\varepsilon_{\mathrm{int}}<0$, the interaction between the steps becomes attractive.

For a vicinal surface, we add the terms of the Andreev field [57] $\boldsymbol{\eta}=(\eta_x,\eta_y)$ to the Hamiltonian in Eqn (6.71) as an external field. The model Hamiltonian given in Eqn (6.71) for the vicinal surface then becomes

$$\mathcal{H}_{\mathrm{vicinal}} = \mathcal{H}_{\mathrm{p-RSOS}} - \eta_x \sum_{i,j}[h(i+1,j)-h(i,j)] - \eta_y \sum_{i,j}[h(i,j+1)-h(i,j)]. \tag{6.72}$$

The partition function $\mathcal{Z}$ for the p-RSOS model is given by $\mathcal{Z}=\sum_{\{h(i,j)\}}\exp[-\beta\mathcal{H}_{\mathrm{vicinal}}]$. The Andreev surface free energy $\tilde{f}(\boldsymbol{\eta})$ is the thermodynamic potential calculated from the partition function $\mathcal{Z}$ by using the expression $\beta\tilde{f}(\boldsymbol{\eta})=-\lim_{\mathcal{N}\to\infty}(1/\mathcal{N})\ln\mathcal{Z}$, where $\mathcal{N}$ is the number of points on the square lattice.

6.6.2.2 Discontinuity in Surface Tension

In Figure 6.26, we show the Andreev free energy calculated by the PWFRG method (Appendix A) [86,96,124,125]. The model is found to have two transition temperatures $T_{f,1}$ and $T_{f,2}$,[24] where for $T<T_{f,1}$, the surface tension of the (111) surface becomes discontinuous, and when $T<T_{f,2}$, the surface tension of the (001) surface becomes discontinuous. The shape of the equilibrium (001) facet is shown in Figure 6.27. As seen from this figure, the (001) facet directly contacts the (111) facet at $T<T_{f,2}$.

From the Legendre transformation of $f(\boldsymbol{p})=\tilde{f}(\boldsymbol{\eta})+\boldsymbol{p}\cdot\boldsymbol{\eta}$, we obtain the vicinal surface free energy and the surface tension $\gamma_{\mathrm{surf}}(\boldsymbol{p})=f(\boldsymbol{p})/\sqrt{1+\left|\boldsymbol{p}\right|^2}$. In Figure 6.28, we show the vicinal surface free energy and the surface tension calculated by the PWFRG

[24] $k_{\mathrm{B}}T_{f,1}/\varepsilon=0.3610\pm0.0005$ and $k_{\mathrm{B}}T_{f,2}/\varepsilon=0.3585\pm0.0007$.

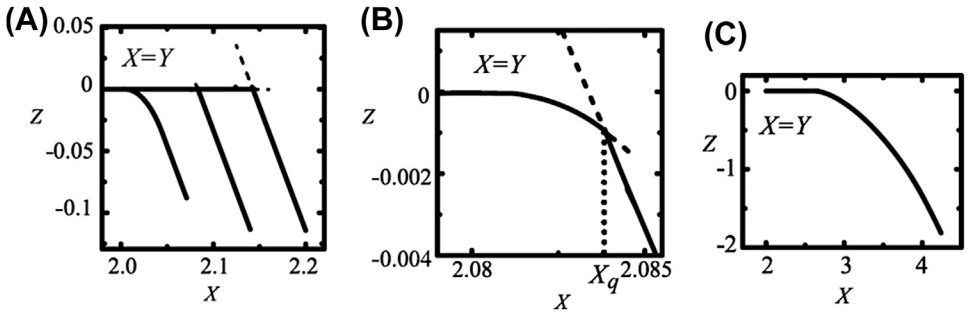


FIGURE 6.26 Profile of the reduced equilibrium crystal shape calculated by the PWFRG method. $X = \eta_x/k_BT = -\lambda x/k_BT$, $Y = \eta_y/k_BT = -\lambda y/k_BT$, and $Z = \tilde{f}(\boldsymbol{\eta})/k_BT = \lambda z(x,y)/k_BT$. $\varepsilon_{int}/\varepsilon = -0.5$. Broken lines represent metastable lines. (A) From right to left, $k_BT/\varepsilon = 0.35$, 0.36, and 0.37. (B) $k_BT/\varepsilon = 0.36$. The edge of the (111) facet is denoted by X_q. (C) Original RSOS model ($\varepsilon_{int} = 0$); $k_BT/\varepsilon = 0.3$. *From Ref. [96].*

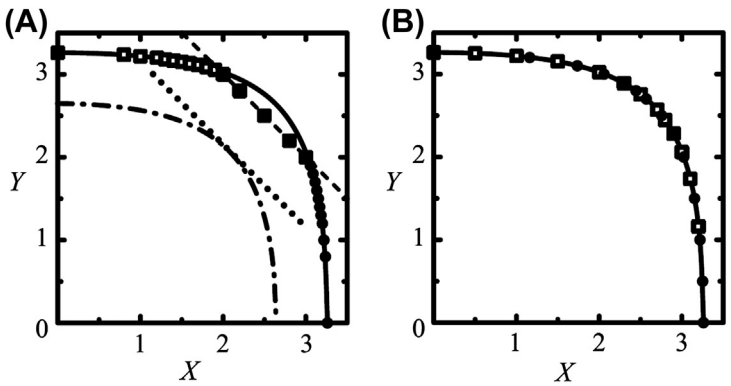


FIGURE 6.27 Equilibrium facet shape (EFS) for $X > 0$ and $Y > 0$. Filled circles: (X_c, Y_c) values calculated by the PWFRG method [44] (Appendix A) for $k_BT/\varepsilon = 0.3$. Open squares: (Y_c, X_c) values. Solid lines: EFS of the 2D square nn Ising model for $k_BT/\varepsilon = 0.3$. Dash-dotted lines: EFS of the 2D Ising model for $k_BT/\varepsilon = 0.361$. EFS of 2D Ising model is calculated by $\cosh(X_c) + \cosh(Y_c) = [\cosh^2(\varepsilon/(k_BT))]/[\sinh(\varepsilon/(k_BT))]$. Dashed lines: $Y = -X + 5.0$. Dotted lines: $Y = -X + 4.1551$ ($Y = -X + (2\varepsilon + \varepsilon_{int})/k_BT$). (A) $\varepsilon_{int}/\varepsilon = -0.5$; (B) $\varepsilon_{int} = 0$. *From Ref. [86].*

method. For $T_{f,2} < T < T_{f,1}$, we have obtained a non-GMPT form for the vicinal surface free energy, as follows [86]:

$$f_{eff}(\boldsymbol{p}) = f(0) + \gamma_{step}(\phi)\left|\frac{\boldsymbol{p}}{a_z}\right| + A_{eff}(\phi)\left|\frac{\boldsymbol{p}}{a_z}\right|^2 + B_{eff}(\phi)\left|\frac{\boldsymbol{p}}{a_z}\right|^3 + \mathcal{O}\left(\left|\frac{\boldsymbol{p}}{a_z}\right|^4\right). \tag{6.73}$$

Here, the quadratic term with respect to $|\boldsymbol{p}|$ has reappeared.

It should be noted that the metastable lines in Figure 6.26 converge to points when $T < T_{f,2}$. This means that there are no Gaussian-type capillary wave fluctuations (Figure 6.11(C)) with respect to the surface slope. This lack of slope fluctuations

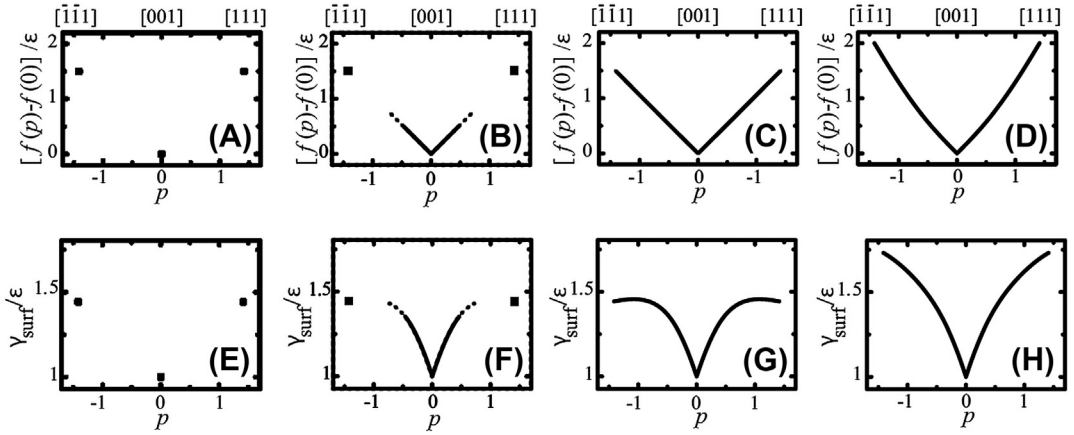


FIGURE 6.28 The surface tilted toward the (110) direction, as calculated by the PWFRG (DMRG) method. (A)–(D) Slope dependence of the vicinal surface free energy $[f(p_x, p_y) - f(0,0)]/\varepsilon$; and (E)–(H) surface tension $\gamma_{\mathrm{surf}}(p,p)/\varepsilon$. For (A)–(C) and (E)–(G), $\varepsilon_{\mathrm{int}}/\varepsilon = -0.5$; and for (D) and (H), $\varepsilon_{\mathrm{int}} = 0$, the original RSOS model. Temperatures: $k_{\mathrm{B}}T/\varepsilon = 0.35$ for (A), (D), (E), and (H); $k_{\mathrm{B}}T/\varepsilon = 0.36$ for (B) and (F); $k_{\mathrm{B}}T/\varepsilon = 0.37$ for (C) and (G). Closed squares: (A) and (B), (0,0) and $[f(1,1) - f(0,0)]/\varepsilon$; (E) and (F), $\gamma_{\mathrm{surf}}(0,0)/\varepsilon$ and $\gamma_{\mathrm{surf}}(1,1)/\varepsilon$. Broken lines: (B) and (F), the curves for the metastable states. *From Ref. [96].*

stabilizes the flat plane of the side surface of the macrostep. This faceted macrostep formation was confirmed by a Monte Carlo calculation [86].

In addition to the above properties, the movements of a macrostep are inhibited near equilibrium [96]. When a macrostep is faceted ($T < T_{f,2}$), the kink density at the side surface of the macrostep is extremely small because the side surface is flat and smooth. The macrostep moves by way of 2D nucleation, and thus there is intermittent motion of the macrostep near equilibrium. When the driving force is large enough to frequently create new 2D nuclei, kinetic roughening occurs on the side surfaces of the macrosteps. Hence, the macrostep dissolves into a homogeneous vicinal surface.

In the temperature region $T_{f,2} < T < T_{f,1}$, the steps merge locally, forming "step droplets" with a finite lifetime [86,96]. Due to the discontinuity in the surface tension around the (111) surface, the kink density decreases on the merged steps. The movements of the vicinal surface with step droplets, therefore, become smaller than what would be expected for the vicinal surface with regular train of steps.

6.7 Summary

Roughening and faceting transitions of surfaces and steps have been reviewed from the point of view of statistical mechanics. We paid attention to the crystal surfaces expressed by the distinct border in the microscopic scale between the crystal phase and the ambient phase such as in the case of vapor growth or solution growth. The definition of the surface free energy (or the interface free energy) was introduced. We have also shown how to define and calculate the surface free energy and step tension.

The microscopic surface shape is well described by the SOS model. Analysis of the statistical mechanics of the SOS model shows that the roughening transition is the Kosterlitz–Thouless type phase transition. In the rough phase, the variance of the surface heights becomes large and diverges logarithmically with the linear size of the surface due to the drumhead wandering. The surface size dependence of the variance is attributed to the lower *cutoff* of the wave number of the drumhead wandering. The universal features associated with the KT transition were reviewed concerning the roughening and faceting transitions. We have presented a brief description of other systems that have complex phase diagrams for the roughening transition temperature.

Universal features on the vicinal surface have been also discussed for temperatures lower than the roughening temperature of the terrace surface. The logarithmic behavior with respect to the length of the step has been explained for the variance of a single step height.

Finally, we discussed the faceting of a macrostep and its stability.

Appendix A. Transfer Matrix Method

PWFRG is an acronym for the product wave-function renormalization group method [44]. The PWFRG method is a transfer matrix [130] version of the White's density matrix renormalization group (DMRG) method [131,132] for 1D quantum spin systems. The 1D quantum system can be mapped to a 2D classical system through the transfer matrix with the Suzuki–Trotter formula [133].

In order to apply the PWFRG method to the RSOS systems, we construct the transfer matrix $\widehat{T}(t_1, t_2, \cdots, t_N; s'_1, s'_2, \cdots, s'_N)$ (Figure A.29(C)) by using the 19-vertex model [32,113] (Figure A.29(B)). The partition function $\mathcal{Z}$ is then rewritten by $\widehat{T}$ as follows:

$$\mathcal{Z} = \mathrm{Tr}\left[\widehat{T}\left(t_1, t_2, \cdots, t_N; s'_1, s'_2 \cdots, s'_N\right)^M\right], \tag{A.1}$$

where N is the number of the linked vertices, and M is the linear size of the system in the vertical-direction in Figure A.29(D). Then the transfer matrix is expressed by use of the statistical weight denoted by $V(s,t;s',t')$ as follows:

$$\widehat{T}\left(t_1, t_2, \cdots, t_N; s'_1, s'_2, \cdots, s'_N\right) = \sum_{\{q_1\},\{q_2\},\cdots} V\left(s_1, t_1; s'_1, q_1\right) V\left(q_1, t_2; s'_2, q_2\right) \cdots V\left(q_{N-1}, t_N; s'_N, t'_N\right). \tag{A.2}$$

In the limit of $M,N \to \infty$, only the largest eigenvalue of the transfer matrix $\Lambda(N)$ contributes to the partition function. The Andreev surface free energy, therefore, is obtained from the partition function (Eqn (A.1)) as

$$\beta\tilde{f}(\boldsymbol{\eta}) = -\lim_{M,N\to\infty} \frac{1}{NM} \ln \Lambda(N)^M, \tag{A.3}$$

The transfer matrix is diagonalized efficiently by the PWFRG method. In the PWFRG calculation, the number of the so-called "retained bases" m are set from 7 to 37. The iteration number for the diagonalization is set to be $1500 - 1 \times 10^5$.

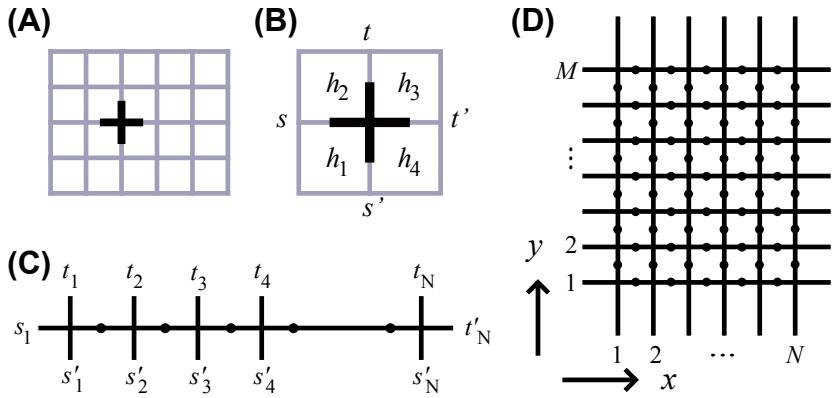


FIGURE A.29 (A) Top view of the (001) surface of a cubic lattice. (B) Quadruplet of squares surrounding a vertex. In the figure, $s = h_2 - h_1$, $t = h_3 - h_2$, $s' = h_4 - h_1$, $t' = h_2 - h_1$ with s, s', t, and T' being $\{0,\pm 1\}$. (C) The transfer matrix assembled by the vertices. (D) Products of the transfer matrices.

Appendix B. Driving Force for Crystal Growth

The driving force for crystal growth $\Delta\mu$ is defined as the difference of the bulk chemical potential between the crystal phase and the ambient phase. Examples of the driving force are shown in Table B.3 [94].

Appendix C. Example of the Anisotropy of the Entropy of a Step

Let us consider a step on the (001) surface of cubic lattice (Figure C.30). The ground-state structure of a (010) step is uniquely determined. In Figure C.30(A), the step has six broken bonds, and eight in (B). Hence, the structure in (A) is the ground state, and the structure in (B) is in an excited state. On the other hand, (C), (D), and (E) also have eight broken bonds. There are 67 other structures with the same number of broken bonds. These 70 structures are all in the ground state, but they are degenerated. Hence, the ground-state structure of the (110) step cannot be determined uniquely. Since the entropy S is given by $S = k_B \ln W$, where W represents the degeneracy, the entropy of a step depends on the azimuth of the step.

Appendix D. IPW Method

In this section, we explain how to derive the 1D interface tension from a rough phase separation line in the 2D Ising model [100]. The 1D interface as the phase separation line is made by an antiphase boundary condition, as shown in Figure 6.9. We regard the

Table B.3 Equation of Driving Force

Vapor growth	$\Delta\mu = k_B T \ln P/P_0$[a]
Solution growth	$\Delta\mu = k_B T \ln C/C_0$[b]
Melt growth	$\Delta\mu = \mu_{liquid} - \mu_{crystal} \approx = \Delta s_{melting}(T_m - T)$[c]
Electrocrystallization of metals	$\Delta\mu = ze\eta$[d]

[a]k_B represents the Boltzmann constant, T represents the temperature, P represents the pressure in the ambient phase, and P_0 represents the equilibrium pressure at some temperature T_0.[b]C and C_0 are the real and the equilibrium concentrations of the solute, respectively.[c]T_m represents the melting temperature, and $\Delta s_{melting}$ represents the entropy of melting.[d]z represents the valence of the neutralizing ions, e represents the elementary electric charge, and $\eta = E - E_0$ represents the over-voltage given by the difference of the equilibrium potential E_0.

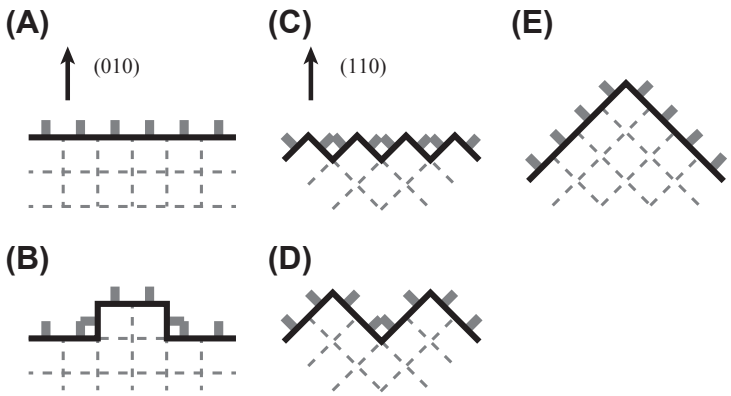


FIGURE C.30 Microscopic configuration of a step on a cubic lattice with both ends being fixed. Black solid line: the edge of a step. Gray line: a broken bond. (A) and (B): (010) steps. (C), (D), and (E): (110) steps.

phase separation line as a random walk connecting the ends O and P. The main idea for obtaining $\gamma_{1D}(\theta,T)$ is to use the "duality" between the interface tension and the correlation length [134]. The asymptotic form of the correlation function becomes

$$G(\boldsymbol{R}) \sim \exp[-|\boldsymbol{R}|/\xi(\theta, T^*)], \quad (|\boldsymbol{R}| \to \infty), \tag{D.1}$$

where T^* represents the temperature in the world of $\xi(\theta,T^*)$. From the duality relation, we have $\gamma_{1D}(\theta,T)/k_B T = 1/\xi(\theta,T^*)$, and

$$\mathcal{G}(\boldsymbol{R}+\boldsymbol{r}, \beta|\boldsymbol{R}, \alpha) \sim \exp[-\gamma_{1D}(\theta, T)|\boldsymbol{r}|/k_B T], \quad (|\boldsymbol{r}| \to \infty), \tag{D.2}$$

We will only give a brief review of the derivation of the final equations, Eqns (D.6)–(D.8) [100].

Let us label each elementary path by a pair $(\boldsymbol{R},\alpha)$, where $\boldsymbol{R}$ represents the starting position of the path and α represents its direction, such as $\{\rightarrow,\leftarrow,\uparrow,\downarrow\}$. We denote the weighted sum over all possible N-step walks by $\mathcal{G}_N(\boldsymbol{R}, \alpha|\boldsymbol{R}_0, \alpha_0)$, where the random walk

starts at $\boldsymbol{R}_0$ with direction α_0 and ends at $\boldsymbol{R}$ with direction α. Using the connectivity matrix $A(\boldsymbol{R},\alpha|\boldsymbol{R}',\beta)$, $\mathcal{G}_N$ satisfies the following recursion relation:

$$\mathcal{G}_{N+1}(\boldsymbol{R},\alpha|\boldsymbol{R}_0,\alpha_0)=\sum_{\boldsymbol{R}',\beta}A(\boldsymbol{R},\alpha|\boldsymbol{R}',\beta)\mathcal{G}_N(\boldsymbol{R}',\beta|\boldsymbol{R}_0,\alpha_0). \tag{D.3}$$

Writing in matrix form, Eqn (D.3) becomes $\mathcal{G}_{N+1}=A\mathcal{G}_N$. Let us introduce $\mathcal{G}$ as $\mathcal{G}=\sum_{N=0}^{\infty}\mathcal{G}_N$. Then with initial condition $\mathcal{G}_0=1$, we have $\mathcal{G}=[1-A]^{-1}$. Then, evaluating Z^{++} and $Z^{+-}(\theta)$ by Vdovichenko's method [135], we have

$$\frac{Z^{+-}(\theta)}{Z^{++}}\propto\mathcal{G}(\boldsymbol{R}+\boldsymbol{r},\beta|\boldsymbol{R},\alpha). \tag{D.4}$$

From Eqns (D.2)–(D.4),

$$\gamma_{1D}(\theta,T)=-k_BT\lim_{|\boldsymbol{r}|\to\infty}\frac{1}{|\boldsymbol{r}|}\ln\left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}dk^2\frac{e^{i\boldsymbol{k}\boldsymbol{r}}}{D(\boldsymbol{k})}\right], \tag{D.5}$$

where $D(\boldsymbol{k})$ represents $D(\boldsymbol{k})=\det[1-A(\boldsymbol{k})]$ and $A(\boldsymbol{k})$ represents the matrix of the Fourier components of $A(\boldsymbol{R},\alpha|\boldsymbol{R}',\beta)$.

We estimate the right-hand side of Eqn (D.5) by the saddle-point method, and denote the saddle-point $\boldsymbol{k}^*=(i\omega_1,i\omega_2)$. Then, we have a set of equations as follows [39,86,100]:

$$\gamma_{1D}(\theta,T)=k_BT(\omega_1\cos\theta+\omega_2\sin\theta) \tag{D.6}$$

$$D(\boldsymbol{\omega})=0 \tag{D.7}$$

$$[\partial D(\boldsymbol{\omega})/\partial\omega_2]/[\partial D(\boldsymbol{\omega})/\partial\omega_1]=\tan\theta. \tag{D.8}$$

From the thermodynamics of ECS, we have $-\eta=\lambda x=k_BT\omega_2$ and $\tilde{f}(\eta)=\lambda y=k_BT\omega_1$.

It should be noted that $D(\boldsymbol{k})$ also appears in the bulk free energy of the 2D Ising model.

Appendix E. Calculation of Surface Width

In this section, we derive an explicit equation for the variance of the surface height, Eqn (E.6) [23].

The variance of the surface height,

$$W^2=\left\langle[h(\boldsymbol{x})-\langle h(\boldsymbol{x})\rangle]^2\right\rangle, \tag{E.1}$$

can also be written as

$$W^2=\sum_k\langle h(\boldsymbol{k})h(-\boldsymbol{k})\rangle \tag{E.2}$$

where

$$h(\boldsymbol{x}) = \sum_{\boldsymbol{k}} h(\boldsymbol{k})\exp(i\boldsymbol{k}\boldsymbol{x}), \quad \boldsymbol{k} = (k_1, k_2),$$
$$k_1 = (2\pi m)/L, \quad k_2 = (2\pi m')/L. \quad (m, m' = \pm 1, \pm 2, \pm 3, \cdots) \tag{E.3}$$

Using Eqn (6.26) for the thermal average, we obtain the following equation:

$$W^2 = \frac{k_B T}{L^2} \sum_{\boldsymbol{k}} \left[\sum_{\alpha,\beta} f^{\alpha\beta} k_\alpha k_\beta \right]^{-1}, \tag{E.4}$$

where $f^{\alpha\beta}$ represents $f^{xx}(\boldsymbol{p}_e)$, $f^{yy}(\boldsymbol{p}_e)$, $f^{xy}(\boldsymbol{p}_e)$, and $f^{yx}(\boldsymbol{p}_e)$, and they are the components of the surface stiffness tensor defined by Eqn (6.27). Applying a continuous approximation to the r.h.s. of the above summation with respect to $\boldsymbol{k}$, we have

$$W^2 = \frac{k_B T}{2\pi\sqrt{\det(f^{\alpha\beta})}} \int_{k_0}^{\Lambda} dq \frac{1}{q}, \tag{E.5}$$

where $\Lambda = 2\pi/a$ represents the upper cutoff, and $k_0 = 2\pi/L$ represents the lower cutoff. In this way, we have in the limit of $L \to \infty$,

$$W^2 = \frac{k_B T}{2\pi\sqrt{\det(f^{\alpha\beta})}} \ln L. \tag{E.6}$$

Appendix F. Derivation of the Capillary Wave Hamiltonian

In this section, we derive Eqn (6.26).

The total surface free energy Φ of a crystal droplet surrounded by rough surfaces is described as follows:

$$\Phi = \int \gamma_{\text{surf}}(\boldsymbol{n}) dA = \int f(\boldsymbol{p}) dx dy, \quad f(\boldsymbol{p}) = \gamma_{\text{surf}}(\boldsymbol{n})\sqrt{g} \tag{F.1}$$

where $\gamma_{\text{surf}}(\boldsymbol{n})$ represents the surface tension, $\boldsymbol{n}$ represents the surface normal unit vector, dA represents a small surface area, and $f(\boldsymbol{p})$ represents the surface free energy per projected xy area. For an inclined surface of area L^2, a generalized free energy $\tilde{F}(\boldsymbol{p}, \boldsymbol{\eta})$ [23] is considered from Eqn (F.1) as follows:

$$\tilde{F}(\boldsymbol{p}, \boldsymbol{\eta}) = L^2 [f(\boldsymbol{p}) - \boldsymbol{p} \cdot \boldsymbol{\eta}], \tag{F.2}$$

where $\boldsymbol{\eta}$ and $\boldsymbol{p}$ are assumed to be independent variables. The equilibrium orientation of the surface is obtained by minimizing $\tilde{F}(\boldsymbol{p}, \boldsymbol{\eta})$ with respect to $\boldsymbol{p}$. After some calculations, we have the following equation near the equilibrium surface slope $\boldsymbol{p}_e$, Figure 6.11(C):

$$\tilde{F}(\boldsymbol{p}, \boldsymbol{\eta}) = L^2 \tilde{f}(\boldsymbol{\eta}) + \frac{1}{2} L^2 \left[f^{xx}(\boldsymbol{p}_e) \Delta p_x^2 + f^{yy}(\boldsymbol{p}_e) \Delta p_y^2 + \left(f^{xy}(\boldsymbol{p}_e) + f^{yx}(\boldsymbol{p}_e) \right) \Delta p_x \Delta p_y \right], \tag{F.3}$$

where $\tilde{f}(\boldsymbol{\eta})$ represents the Andreev free energy [57]; and $f^{xx}(\boldsymbol{p}_e)$, $f^{yy}(\boldsymbol{p}_e)$, $f^{xy}(\boldsymbol{p}_e)$, and $f^{yx}(\boldsymbol{p}_e)$ are the components of the surface stiffness tensor defined by Eqn (6.27).

Hence, the capillary wave Hamiltonian of the inclined crystal surface becomes Gaussian (Figure 6.11(C)) as follows:

$$\mathcal{H}_{CW} = \frac{1}{2}\int_0^L dx \int_0^L dy \Big[f^{xx}(\boldsymbol{p}_e)\Delta p_x^2 + f^{yy}(\boldsymbol{p}_e)\Delta p_y^2 + \big(f^{xy}(\boldsymbol{p}_e) + f^{yx}(\boldsymbol{p}_e)\big)\Delta p_x \Delta p_y\Big]. \tag{F.4}$$

Appendix G. Other Microscopic Models

Appendix G.1. Discrete Gaussian Model

The discrete Gaussian (DG) model introduced by Chui and Weeks [42] is one of the SOS models, and the Hamiltonian of the DG model is given as follows:

$$\mathcal{H}_{DG} = \frac{J}{2}\sum_{j,\boldsymbol{\delta}} \big(h_j - h_{j+\boldsymbol{\delta}}\big)^2, \tag{G.1}$$

where $\{h_j\}$ represents the surface height at a site j and J represents the energy cost to create the height difference on the surface between i and $i+\boldsymbol{\delta}$ as the nn lattice sites. The summation with respect to i runs over all lattice points, and the summation with respect to $\boldsymbol{\delta}$ runs over all the nn sites around i. Note that $\{h_j\}$ in the DG model is a continuous variable.

Since the DG model is a Gaussian-type model, it is easy to analyze theoretically. The model, therefore, contributes to establishing an essential connection between the roughening transition on the surface and the XY model in two dimensions.

Appendix G.2. RSOS-I Model

The image of the RSOS model and the materials of adsorption for the RSOS-I model are similar to Figure 6.21. The Hamiltonian for the RSOS-I model [126,128] is written as follows:

$$\begin{aligned}
&\mathcal{H}_{\text{RSOS-I}} = \mathcal{H}_{\text{RSOS}} + \mathcal{H}_{\text{Ising}} + \mathcal{H}_{\text{int}} \\
&= \sum_{m,n}\big[\varepsilon\big(1-\alpha\sigma_y(m,n)\big)|h(m+1,n)-h(m,n)| + \varepsilon(1-\alpha\sigma_x(m,n))|h(m,n+1)-h(m,n)|\big] \\
&\quad -J\sum_{m,n}[\sigma_x(m,n)\sigma_y(m,n) + \sigma_x(m,n)\sigma_y(m-1,n) \\
&\quad +\sigma_x\big(m,n-1\big)\sigma_y\big(m,n\big) + \sigma_x\big(m,n-1\big)\sigma_y\big(m-1,n\big)\big] \\
&\quad -H\sum_{m,n}\big[\sigma_x(m,n)+\sigma_y(m,n)\big].
\end{aligned} \tag{G.2}$$

In the case of $J=0$, the model reduces to the decorated RSOS model in Section 4.4.2 (Eqn (6.50)).

Acknowledgment

This work is supported by the Japan Society for Promotion of Science (JSPS) KAKENHI Grant Number 25400413.

References

[1] van Beijeren H. Phys Rev Lett 1977;38:993.

[2] Swendsen RH. Phys Rev B 1982;25:2019.

[3] den Nijs M. J Phys A Math Gen 1985;18:L549.

[4] Abraham DB. In: Domb C, Lebowitz JL, editors. Phase transition and critical phenomena, vol. 10, 1. London: Academic Press; 1986.

[5] Aizenman M, Fernandez R. J Stat Phys 1986;44:383.

[6] Akutsu Y, Akutsu N. J Phys A 1986;19:2813.

[7] Fisher MPA, Fisher DS, Weeks JD. Phys Rev Lett 1982;48:368. de Coninck J, Ruiz J. J Phys 1988;A21: L147. Abraham DB, Upton PJ. Phys Rev 1988;B37:3835.

[8] Rottman C, Wortis M. Phys Rev B 1981;24:6274. Avron JE, van Beijeren H, Schulman LS, Zia RKP. J Phys 1982;A15:L81.

[9] Müller-Krumbhaar H. In: Kaldis E, editor. Current topics in materials science, vol. 1. Amsterdam: North-Holland Publishing; 1978. Ch. 1, 1.

[10] Knops HJF. Phys Rev Lett 1977;39:766.

[11] Jose JV, Kadanoff LP, Kirkpatrick S, Nelson DR. Phys Rev B 1977;16:1217.

[12] Weeks JD. In: Riste T, editor. Ordering in strongly fluctuation condensed matter systems. Plenum, New York and London; 1980. p. 293.

[13] Gruber EE, Mullins WW. J Phys Chem Solids 1967;28:6549. Pokrovsky VL, Talapov AL. Phys Rev Lett 1979;42:65; Zh Eksp Teor Fiz 1980;78:269 [Sov Phys JETP 1980;51:134].

[14] Haldane FDM, Villain J. J Phys Paris 1981;42:1673.

[15] Villain J. In: Riste T, editor. Ordering in strongly fluctuation condensed matter systems. Plenum, New York and London; 1980. p. 222. Villain J, Bak P. J Phys Paris 1981;42:657.

[16] Cabrela N, Garcia N. Phys Rev B 1982;25:6057.

[17] Izuyama T, Akutsu Y. J Phys Soc Jpn 1982;51:50. Yamamoto T, Izuyama T. J Phys Soc Jpn 1987;56: 632.

[18] Jayaprakash C, Saam WF, Teitel S. Phys Rev Lett 1983;50:2017.

[19] Rottman C, Wortis M. Phys Rep 1984;103:59. Wortis M. Chemistry and physics of solid surface VII. In: Vanselow R, Howe R, editor. Berlin, Heidelberg: Springer-Verlag;1988. p. 367.

[20] Jayaprakash C, Saam WF. Phys Rev B 1984;30:3916.

[21] Schultz HJ. J Phys 1985;46:257.

[22] van Beijeren H, Nolden I. In: Schommers W, von Blanckenhagen P, editors. Structure and dynamics of surfaces II. Berlin, Heidelberg: Springer-Verlag; 1987. p. 259.

[23] Akutsu N, Akutsu Y. J Phys Soc Jpn 1987;56:1443.

[24] Akutsu N, Akutsu Y. J Phys Soc Jpn 1987;56:2248 [Uwaha M, unpublished].

[25] Akutsu Y, Akutsu N, Yamamoto T. Phys Rev Lett 1988;61:424.

[26] Yamamoto T, Akutsu Y, Akutsu N. J Phys Soc Jpn 1988;57:453.

[27] Mikheev LV, Pokrovsky VL. J Phys 1991;1:373.

[28] Saam WF. Phys Rev Lett 1989;62:2636.

[29] Akutsu Y, Akutsu N, Yamamoto T. Phys Rev Lett 1989;62:2637.

[30] Yamamoto T, Akutsu Y, Akutsu N. J Phys Soc Jpn 1989;58:3531. Yamamoto T, Akutsu N, Akutsu Y. J Phys Soc Jpn 1990;59:3831; J Phys Soc Jpn 1991;60:3600.

[31] Akutsu N, Akutsu Y. Surf Sci 1997;376:92.

[32] Akutsu N, Akutsu Y. Phys Rev B 1998;57:R4233.

[33] Akutsu N, Akutsu Y. Prog Theor Phys 2006;116:983.

[34] Rottman C, Wortis M. Phys Rev B 1984;29:328.

[35] Jayaprakash C, Rottman C, Saam WF. Phys Rev B 1984;30:6549.

[36] Nozières P. In: Godrèche C, editor. Solids far from equilibrium. Cambridge, New York, Port Chester, Mellbourne, Sydney; 1991. p. 1.

[37] Pimpinelli A, Villain J. Physics of crystal growth. Cambridge University Press; 1998.

[38] Akutsu N, Akutsu Y. J Phys Soc Jpn 1990;59:3041. J Phys 1990;A23:L217. Akutsu N. J Phys Soc Jpn 1992;61:477. Akutsu N, Akutsu Y. J Phys Soc Jpn 1995;64:736; J Phys Soc Jpn 1996;65:1233.

[39] Akutsu N, Akutsu Y. J Phys Condens Matter 1999;11:6653.

[40] Kosterlitz JM, Thouless DJ. J Phys C 1973;6:1181. Kosterlitz JM. J Phys C 1974;7:1046.

[41] Young AP. In: Riste T, editor. Ordering in strongly fluctuation condensed matter systems. Plenum Press, New York and London; 1980. p. 271. Wegner FJ. Phase transition and critical phenomena, vol. 6, Domb C, Green MS, editors. London: Academic Press; 1976. p. 7.

[42] Chui ST, Weeks JD. Phys Rev B 1976;14:4978.

[43] Bishop DJ, Reppy JD. Phys Rev Lett 1978;40:1727. Nelson DR. Phase transition and critical phenomena, vol. 7, p. 1. Domb C, Lebowitz JL, editors. London: Academic Press; 1983.

[44] Nishino T, Okunishi K. J Phys Soc Jpn 1995;64:4084. Hieida Y, Okunishi K, Akutsu Y. Phys Lett 1997; A233:464; New J Phys 1999;1:7.1.

[45] Bauer E, Telieps W. Ultramicroscopy 1985;17:57. Scan Microsc 1987;(suppl. 1):99; In: The study of surfaces and interfaces by electon optical techniques, Howie A, Valderé U, editors. Plenum, New York, 1988; p. 195.

[46] Bartelt NC, Tromp RM, Williams ED. Phys Rev Lett 1994;73:1656.

[47] Mermin ND, Wagner H. Phys Rev Lett 1966;17:1133.

[48] Wulff G, Kristallogr Z. Miner 1901;34:449.

[49] von Laue M, Kristallogr Z. Miner 1943;105:124.

[50] Herring C. Phys Rev 1951;82:87.

[51] Hartman P, Pedok WG. Acta Cryst 1955;8:49. Acta Cryst 1955;8:521; Acta Cryst 1995;8:525.

[52] MacKenzie JK, Moore AJW, Nicholas JF. J Chem Phys Solids 1962;23:185.

[53] Frank FC. Metal surfaces, ASM. 1963. 1.

[54] Toschev S. In: Hartman P, editor. Crystal growth, an introduction; 1973. p. 328. North-Holland.

[55] Ookawa A. Crystal growth. Syōkabō, Tokyo. 1977 [in Japanese].

[56] Landau LD, Lifshitz EM. Statistical physics. 2nd ed. Oxford, Pergamon; 1968. Landau LD, Lifshitz EM. Statistical physics, Part 1 (Course of theoretical physics vol. 5), 3rd ed., Elsevier, Butterworth-Heinemann, Amsterdam; 1980).

[57] Andreev AF. Zh Eksp Theor Fiz 1981;80:2042. Sov Phys JETP 1981;53:1063.

[58] Kern R, Le Ley G, Metois JJ. In: Kaldis E, editor. Current topics in materials science, vol. 3. Amsterdam: North-Holland Publ.; 1979. Ch. 3, 131. Kern R. Morphology of crystals. Sunagawa I, editor. Tokyo: Terra Scientific Publishing; 1987. p. 77.

[59] Kreyszig E. Introduction to differential geometry and Riemannian geometry. University of Tronto Press; 1968.

[60] Ehrenfest P. Ann Phys 1915;48:360.

[61] Yamada M. Phys Zeit 1923;24:364. 1924;25:52; 1924;25:289.

[62] Cabrera N, Coleman RV. In: Gilman JJ, editor. The art and science of growing crystals. New York, London: John Wiley & Sons; 1963. p. 3.

[63] Cabrera N. Surf Sci 1964;2:320.

[64] Balibar S, Edwards DO, Laroche C. Phys Rev Lett 1979;42:782.

[65] Landau J, Lipson SG, Maattanan LM, Balfour LS, Edwards DO. Phys Rev Lett 1980;45:31.

[66] Avron JE, Balfour LS, Kuper CG, Landau J, Lipson SG, Schulman LS. Phys Rev Lett 1980;45:814.

[67] Wolf PE, Gallet GF, Balibar S, Rolley E, Noziéres P. J Phys 1985;46:1987.

[68] Noziéres P, Gallet F. J Phys 1987;48:35. Gallet GF, Noziéres P, Balibar S, Rolley E. Europhys Lett 1986;2:701.

[69] Babkin AV, Keshishev KO, Kopeliovich DB, Ya Parshin A. Sov Phys JETP Lett 1984;39:633.

[70] Carmi Y, Lipson SG, Polturak E. Phys Rev B 1987;36:1894.

[71] Balibar S, Guthmann C, Rolley E. J Phys I 1993;3:1475.

[72] Balibar S, Alles H, Parshin AY. Rev Mod Phys 2005;77:318.

[73] Pavlovska A, Bauer E. Europhys Lett 1989;9:797. Pavlovska A, Faulian K, Bauer E. Surf Sci 1989; 221–233.

[74] Bonzel HP, Dückers K. In: Vanselow R, Howe RF, editors. Chemistry and physics of solid surfaces VII. Berlin, Heidelberg: Springer-Verlag; 1988. p. 429.

[75] Cheng W-C, Wynblatt P. Surf Sci 1996;364:417. Arenhold K, Surnev S, Bonzel HP, Wynblatt P. Surf Sci 1999;424:271.

[76] Nowicki M, Bombis C, Emundts A, Bonzel HP, Wynblatt P. Europhys Lett 2002;59:239. New J Phys 2002;4:60.

[77] Ohachi T, Taniguchi I. J Cryst Growth 1983;65:84.

[78] Rottman C, Wortis M, Heyraud JC, Métois J. Phys Rev Lett 1984;52:1009.

[79] Heyraud JC, Métois JJ. Surf Sci 1986;177:213. Métois JJ, Heyraud JC. Surf Sci 1987;180:647.

[80] Eaglesham DJ, Unterwald FC, Jacobson DC. Phys Rev Lett 1993;70:966.

[81] Bermond JM, Métois JJ, Egea X, Floret F. Surf Sci 1995;330:48.

[82] Suzuki T, Métois JJ, Yagi K. Surf Sci 1995;339:105. Suzuki T, Minoda H, Tanishiro Y, Yagi K. Surf Rev Lett 1999;6:985.

[83] Pavlovska A, Nenow D. J Cryst Growth 1971;8:209. Suf Sci 1971;27:211. Pavlovska A. J Cryst Growth 1979;46:551.

[84] Nenow D, Stoyanova V. J Cryst Growth 1979;46:779.

[85] Görnert P, Voigt F. In: Kaldis E, editor. Current topics in materials science, vol. 11. Amsterdam: North-Holland Publ; 1984. Ch. 1, 1.

[86] Akutsu N. J Phys Condens Matter 2011;23:485004.

[87] Burton WK, Cabrela N, Frank FC. Philos Trans R Soc Lond A 1951;243:299.

[88] Jackson KA. In: Doremus RH, editor. Growth and perfection of crystals: proceedings. New York, London: John Wiley & Sons; 1958. p. 319.

[89] Leamy HJ, Gilmer GH, Jackson KA. In: Blakely JM, editor. Surface physics of crystalline materials. New York: Academic Press; 1975. p. 121.

[90] Buff FP, Lovett RA, Stillinger FH. Phys Rev Lett 1965;15:621.

[91] Weeks JD. J Chem Phys 1977;67:3106.

[92] Jasnow D. Rep Prog Phys 1984;47:1059. Jasnow D. Phase transition and critical phenomena, vol. 10, Domb C, Lebowitz JL, editors. London: Academic Press; 1986. p. 269.

[93] Frank FC. Discuss Faraday Soc 1949:48. No.5.

[94] Markov IV. Crystal growth for beginners. 2nd ed. World Scientific, New Jersey, London, Singapore, Hong Kong; 2003.

[95] Müller-Krumbhaar H, Burkhardt TW, Kroll DM. J Cryst Growth 1977;38:13.

[96] Akutsu N. Phys Rev E 2012;86:061604. Phys Rev E, "Kaleidoscope," December 2012 (http://pre.aps.org/). Akutsu N, J Cryst Growth 2014;401:72. (http://dx.doi.org/10.1016/j.jcrysgro.2014.01.068).

[97] Saito Y. Statistical physics of crystal growth. World Scientific, New Jersey, London, Singapore, Hong Kong; 1996.

[98] Saito Y. In: Riste T, editor. Ordering in strongly fluctuation condensed matter systems. Plenum, New York and London; 1980. p. 319.

[99] Onsager L. Physical review, series II, vol. 65; 1944. 117.

[100] Akutsu Y, Akutsu N. Phys Rev Lett 1990;64:1189. Holzer M. Phys Rev Lett 1990;64:653; Phys Rev B 1990;42:10570.

[101] Bartelt NC, Goldberg JL, Einstein TL, Williams ED, Heyraud JC, Métois JJ. Phys Rev B 1993;48:15453.

[102] Barabasi A-L, Stanley HE. Fractal concepts in surface growth. Cambridge; 1994.

[103] Yamamoto T, Akutsu Y, Akutsu N. J Phys Soc Jpn 1994;63:915. Akutsu Y, Akutsu N, Yamamoto T. J Phys Soc Jpn 1994;63:2032.

[104] Yamamoto T, Phys J. Soc Jpn 1996;65:3810.

[105] Yamamoto T, Akutsu N, Akutsu Y. Advances in the understanding of crystal growth mechanisms, part I. In: Nishinaga T, Nishioka K, Harada J, Sasaki A, Takei H, editors. Crystal growth theory and simulations. Amsterdam, New York: Elsevier Science; 1997. p. 19.

[106] Syoji I. In: Domb C, Green MS, editors. Phase transitions and critical phenomena, vol. 1. London: Academic Press; 1972. p. 270.

[107] Akutsu N, Akutsu Y, Yamamoto T. Phys Rev B 2001;64:85415–21.

[108] Kossel W. Nach Ges Wiss Gottingen 1927:135. Naturewissenschaften 1930;18:901.

[109] Sogo K, Akutsu Y, Abe T. Prog Theor Phys 1983;70:739. Truong TT, den Nijs M. J Phys A 1986;19:L645.

[110] http://noriko-akutsu.com/data/.

[111] Bricmont J, Lebowitz J, Pfister Ch. J Stat Phys 1981;26:313.

[112] Akutsu Y, Akutsu N. J Phys A 1987;20:5981.

[113] den Nijs M, Rommelse K. Phys Rev B 1989;40:4709.

[114] Luijten E, van Beijeren H, Blöte HWJ. Phys Rev Lett 1994;73:456.

[115] Mazzeo G, Carlon E, van Beijeren H. Surf Sci 1996;352–354:960. Carlon E, Mazzeo G, van Beijeren H. Phys Rev B 1997;55:757.

[116] Davidson D, den Nijs M. Phys Rev E 1997;55:1331.

[117] Carlon E, van Beijeren H. Phys Rev E 2000;62:7646.

[118] Marchenko VI, Ya Parshin A. Sov Phys JETP 1980;52:129.

[119] Alerhand OL, Vanderbilt D, Meade RD, Jonnopoulos JD. Phys Rev Lett 1988;61:1973.

[120] Williams ED, Phaneuf RJ, Wei J, Bartelt NC, Einstein TL. Surf Sci 1993;294:219. Surf Sci 1994;310:451.

[121] Sutherland B. J Math Phys 1971;12:246.

[122] Mullins WW. Philos Mag 1961;6:1313.

[123] Bartelt NC, Einstein TL, Rottman C. Phys Rev Lett 1991;66:961.

[124] Akutsu N. Appl Surf Sci 2009;256:1205.

[125] Akutsu N. J Cryst Growth 2011;318:10.

[126] Akutsu N, Akutsu Y, Yamamoto T. Surf Sci 2001;493:475. Prog Theor Phys 2001;105:361.

[127] Akutsu N, Akutsu Y, Yamamoto T. J Cryst Growth 2002;237.

[128] Akutsu N, Akutsu Y, Yamamoto T. Phys Rev B 2003;67:125407–11.

[129] Akutsu N, Hibino H, Yamamoto T. e-J Surf Sci Nanotech 2009;7:39.

[130] Lieb E. In: Domb C, Green MS, editors. Phase transitions and critical phenomena, vol. 1. London, New York: Academic Press; 1972. p. 331. Baxter RJ. Exactly solved models in statistical mechanics, Section 2 and Section 8. London, New York: Academic Press; 1982).

[131] White SR. Phys Rev Lett 1992;69:2863. Nishino T. J Phys Soc Jpn 1995;64:3598.

[132] Peschel I, Wang X, Kaulke M, Hallberg K. In: Density-matrix renormalization, a new numerical method in physics. Berlin, Heidelberg: Springer-Verlag; 1999.

[133] Trotter HF. Proc Am Math Soc 1959;10:545. Suzuki M. Comm Math Phys 1976;51:183.

[134] Zia RKP. Phys Lett A 1978;64:345.

[135] Vdovichenko NV. Zh Eksp Theor Fiz 1964;47:715 [Sov Phys JETP 1965;20:477]. Feynman RP. Statistical mechanics (Benjamin/Cummings, Reading, Mass., 1972). Morita T. J Phys A 1986;19:1197.

7 Theory of Nucleation

Christo N. Nanev

ROSTISLAW KAISCHEW INSTITUTE OF PHYSICAL CHEMISTRY, BULGARIAN ACADEMY OF SCIENCES, ACAD. G. BONCHEV STR. BL. 11, 1113 SOFIA, BULGARIA

CHAPTER OUTLINE

7.1 Introduction 316
7.2 Classical (Capillary) Nucleation Theory and Nucleation in Vapors 317
7.2.1 Nucleation Driving Force 318
7.2.2 Thermodynamics of Homogeneous Nucleation, Energy Barriers for Homogeneous Nucleation, and the Critical Nucleus Size 319
7.2.3 Rate of Homogeneous Nucleation: Steady-State Nucleation Rate 322
7.2.4 Non-Steady-State Nucleation Rate 326
7.2.5 Induction Time for Nucleation 327
7.3 Nucleation in some other systems, Nucleation of Gas Bubbles from Superheated Liquids and Boiling 328
7.3.1 Crystal Nucleation in Melts 330
7.3.2 Crystal Nucleation in Solutions 331
7.4 Earlier Corrections of the CNT 331
7.4.1 Some Recent Nucleation Theories 333
7.5 Molecular-Kinetic Approach to Crystal Nucleation 335
7.6 Equilibrium Shape of Crystals 339
7.7 Two-Dimensional Crystal Nucleation 339
7.8 Heterogeneous (Substrate) Nucleation, the Equilibrium Shape of Crystals on Supports, and Energy Barriers for Heterogeneous Nucleation 342
7.8.1 Saturation Density of Nuclei during Mass Crystallization in Solutions and Melts 345
7.9 Nucleation Theorem 346
7.10 Probabilistic Features of the Nucleation Process 347
7.11 Use of Burst Nucleation for Producing Equally-Sized Nanoparticles 348
7.12 Nucleation of Protein Crystals 348
7.12.1 Bond Selection Mechanism 349
7.12.2 Shape of the Protein Crystal Nucleus 350
7.12.3 Energy Barrier for Protein Crystal Nucleation 351
7.12.4 Protein Cluster–Cluster Aggregation on Diffusional Encounter 352
7.13 Concluding Remarks 354

Handbook of Crystal Growth. http://dx.doi.org/10.1016/B978-0-444-56369-9.00007-1

7.14 Perspectives **355**
Frequently Used Abbreviations **355**
References **355**

7.1 Introduction

Any phase transformation starts with a new phase appearing inside the supersaturated old (mother) phase. Because the thermodynamic stability of the system requires continuity of the thermodynamic functions during the entire course of phase transformation, the change in system thermodynamic potential occurring infinitely close to the transition point has to be infinitely small. Two possibilities correspond to this postulate: either an infinitely small amount of the new phase would appear, with properties distinctly different from the properties of the old phase, or an "infinitesimal" new property would appear simultaneously in the entire phase volume. The first case is denoted as a *first-order phase transition* because first derivatives of the thermodynamic potential are changed, whereas the second case is denoted as a *second-order phase transition* because second derivatives of the thermodynamic potential are changed.

Nucleation is a first-order phase transition. It is a widespread phenomenon in both nature and technology. Rain, fog, ice and snow, salt crystallization by evaporation of sea water, and gas bubble formation in mineral water—just to mention some—begin with the nucleation of a new phase. Nucleation predetermines some basic properties of the new phases created in chemical technology processes (e.g., during evaporation, condensation, and crystallization), metallurgy, deposition of epitaxial layers in electronics, purification (including pharmaceutical substances) by crystallization, formation of nanocrystals, etc. Even in some processes of biological matter crystallization, such as in protein crystallization, nucleation plays a central role. The questions to be answered here are: Why is nucleation required for new phase formation in so many phenomena and processes? Why is it ubiquitous?

Qualitative consideration of the simplest case of a first-order phase transition, namely the transition from single molecules randomly scattered in vapor to a new condensed liquid phase, comes in evidence to nucleation inevitability. Vapor condensation starts with the formation of a sequence of molecule clusters: pairs, triplets, etc. (Figure 7.1). However, because they have a highly convex shape, such clusters tend to dissolve into the ambient mother phase rather than continue to grow. According to the Young–Laplace equation, the smaller the droplet is (i.e., the higher its surface curvature), the higher the droplet vapor pressure is. Although evaporation of a single molecule from a large (flat) liquid surface does not change its curvedness, the same process with very small droplets results in a noticeable curvature increase, and thus in vapor pressure augmentation and further droplet evaporation. In 1878, J.W. Gibbs [1] assumed that repeated density fluctuations in the mother phase were the only reason to oppose such a course of events. He stipulated that a series of density fluctuations were responsible for the formation of an entire set of differently sized undercritical molecule clusters (Figure 7.1(b)) and the critical nucleus itself.

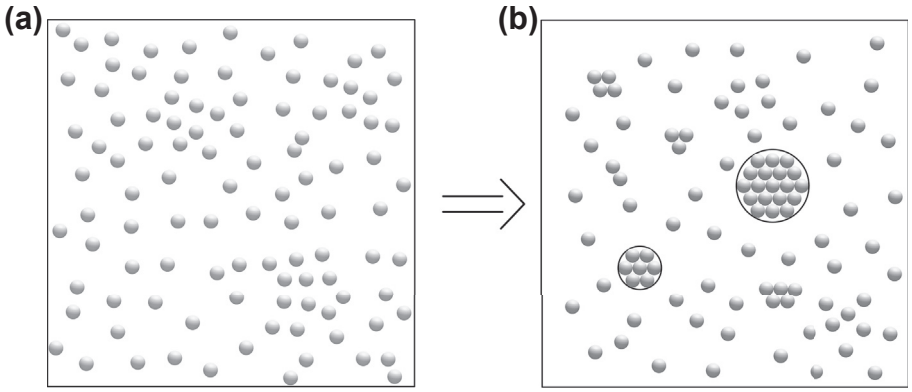


FIGURE 7.1 Snapshots of the dynamic process of new phase formation. (a) The initial state of the vapor phase. Although some very small molecule clusters (e.g., pairs) may appear randomly, they exist only temporarily, decay, and new pairs are born. (b) The start of the transition to the new phase formation.

As is very well known, fluctuations are not limited to metastable system conditions only. Under equilibrium conditions, there already exist locally different and temporally fluctuating numbers of variously sized molecule clusters (Figure 7.1(a)). Gibbs' notion suggests that the probability of sufficiently large fluctuations leading to a stable new phase is infinitesimal in view of the processes occurring near equilibrium. This suggestion is attributed to the large barrier to phase transition arising from the energy cost for creating an interface between the new born cluster and the original phase. Moreover, the larger the critical nucleus is, the larger density fluctuation that is needed for its formation; also, the larger the fluctuation, the less probable (and rarer) is its appearance. Therefore, the prerequisite for the occurrence of a noticeable nucleation process is the establishment of sufficiently high supersaturation, when small fluctuations are required for critical nuclei appearance. Conversely, a limit on metastability is set when, due to a supersaturation increase, fluctuations leading to the new phase become comparable in number to the equilibrium thermodynamic fluctuations in the original phase [2]. This means that the system needs to be brought sufficiently deep into the metastable region to reduce the phase transition barrier until it becomes of the same order as the thermal energy, $k_B T$ (where k_B is the Boltzmann constant and T is the absolute temperature).

7.2 Classical (Capillary) Nucleation Theory and Nucleation in Vapors

Despite nucleation inevitability, even direct observation of critical nuclei has proven elusive.[1] That is why theoretical explanations of nucleation processes have been developed. As already

[1]The protein crystal nuclei make no exception. Although formed by huge protein molecules, being still nanosized particles they remain invisible by optical microscopy. In addition, the critical nuclei are not labeled; it is impossible to distinguish them in the whole assembly of undercritical, critical, and supercritical molecule clusters. In addition, the number of the critical nuclei changes dynamically because of the constant growth/decay of differently sized clusters.

mentioned, Gibbs [1] has given a thermodynamic description of the condensation of supersaturated vapors into liquid droplets. His ideas represent the cornerstone of the classical nucleation theory (CNT), also called the capillary theory or fluctuation theory. Unfortunately, the significance of Gibbs' ideas on nucleation was largely ignored until 1926, when Volmer and Weber [3] acknowledged their importance.

Evidently, when the phase transition is thermodynamically favored, the molecules in the bulk of sufficiently large clusters of the new phase have to be of a lower free energy than the same molecules residing in the parent phase. However, when a nucleus of a new phase with distinctly different properties is formed within the original phase, the two have to be separated by an interface region with intermediate structure and properties; the interfacial width is determined by the structures of the two phases and the interactions between the molecules within these phases. In addition, the smaller the cluster of the new phase, the larger is the percentage of building units that reside in the interfacial region. In his fundamental work, Gibbs [1] introduced the notion of a dividing surface between the old and new phase, and he assumed it was sharp.

Interface matter bonding is less strong than the one in the bulk of the new phase cluster, which is why molecules in the interface region possess more energy compared to the one they would have in the macroscopic new phase. Therefore, the interface, appearing between small clusters and parent phase, is associated with definite interface energy. That is why the creation of small clusters of the new phase requires work to be done; a free energy cost has to be paid, which is the key barrier to nucleation (establishment of sufficiently high supersaturation being condicio sine qua non for creation of small clusters).

7.2.1 Nucleation Driving Force

The thermodynamic supersaturation, which is the nucleation driving force, is generally given as the difference in the chemical potentials of the parent and new phases, $\Delta\mu > 0$. Respectively, equilibrium is characterized by $\Delta\mu = 0$ and undersaturation (overheating, undervoltage) by $\Delta\mu < 0$. The thermodynamic supersaturation for vapor condensation is expressed as

$$\Delta\mu = k_B T \ln (p/p_\infty) \tag{7.1}$$

where p is the actual vapor pressure and p_∞ is the equilibrium pressure (where the infinitely large condensed phase stands in equilibrium). An analogous expression holds true for the bubble formation driving force (see Section 7.3). For the vapor deposition of thin films, $\Delta\mu$ is expressed by replacing the actual and equilibrium vapor pressures through the corresponding impingement rates of vapor atoms on the substrate.

In the case of crystallization from solutions,

$$\Delta\mu = k_B T \ln (a/a_\infty) \approx k_B T \ln (c/c_\infty) \tag{7.2}$$

where a and a_∞ are the corresponding solute activities. Because the activity coefficients are usually taken as equal to 1, $\Delta\mu$ is expressed in this case by the concentration ratio c/c_∞, with c being the actual concentration and c_∞, being the equilibrium concentration.

For crystallization from melts, $\Delta\mu$ is expressed as

$$\Delta\mu \approx Q_M \Delta T / T_M \tag{7.3}$$

where Q_M is the molar heat of fusion and the undercooling $\Delta T = T_M - T$, with T_M being the melting temperature and T being the actual temperature.

Most nucleation processes do not involve chemical transformations; rather, they are purely physical events. However, a few exceptions do exist, such as crystallization of chemical reaction products and electrochemical nucleation. In the case of electrocrystallization of metals, the nucleation driving force is

$$\Delta\mu = zF\eta \tag{7.4}$$

where z denotes the number of exchanged charges (e.g., the valence of the neutralizing ions); $F = 96{,}500$ C/mol, which is the so-called Faraday equivalent; and $\eta = e - e_o$ is the overpotential (overvoltage), which is given by the difference between the electrical potential e that is applied on the electrochemical system and the equilibrium potential e_o of the deposited ion in the solution [4].

7.2.2 Thermodynamics of Homogeneous Nucleation, Energy Barriers for Homogeneous Nucleation, and the Critical Nucleus Size

The spontaneous formation of nuclei in the bulk of a supersaturated system where the probability of the process is equal throughout the whole system, commonly known as *homogeneous nucleation*, usually marks the beginning of any consideration regarding the process. The reason is that the basic physics of nucleation is best illustrated with the help of this (simplest) theory.

The formation of molecule clusters of the new phase in an isothermal and isobaric system consisting of N molecules is to be considered for a one-component system. Gibbs [1] has defined the change in free energy (or thermodynamic potential, ΔG), required for the new phase formation as a sum of two terms: (1) the free energy gain resulting from the transfer of n molecules (atoms, ions) from the supersaturated phase to the new phase cluster, and (2) the free energy penalty Φ imposed due to the formation of the new interface.

If we denote the starting thermodynamic potential by G_{start}, and the final thermodynamic potential by G_{fin}, we can write $G_{start} = N\mu_{mother}$, and $G_{fin} = (N-n)\mu_{mother} + n\mu_{new} + \Phi$, where μ_{mother} and μ_{new} are the chemical potentials of the mother and the new phase. Thus,

$$\Delta G = G_{fin} - G_{start} = -n\Delta\mu + \Phi \tag{7.5}$$

where $\Delta\mu = \mu_{mother} - \mu_{new} > 0$. The first term in Eqn (7.5) is negative because the new phase is more stable than the old one. This term decreases the system thermodynamic potential and indicates a tendency toward a spontaneous phase transition.

Consideration of droplet nucleation in the vapor phase, with droplets assumed to be of spherical shape, is an important basis to analyze other cases. So, n depends on the third power of droplet radius r, and the second term in the sum is proportional to the

interface area (i.e., it depends on droplet radius in power 2). Initially, with very small cluster sizes, the surface-to-volume ratio is large and the second term prevails. That is why the smallest liquid clusters tend to evaporate. Rising cluster size leads to the volume term increasing faster than the surface term. Thus, the competition between these two terms determines the energy barrier for nucleus formation as the maximum in total free energy change, $\Delta G^* = \max$, which is reached at the *critical cluster size* (Figure 7.2).

In large and/or open systems, the critical nucleus stands in an unstable equilibrium with the surrounding parent phase.[2] The equilibrium is unstable because even an infinitesimal increase in the critical nucleus size leads to a decrease in thermodynamic potential (Figure 7.2), and nucleus growth becomes favorable. With further growth, droplet vapor pressure decreases continuously. Because the external pressure remains unchanged in such systems, the growth process becomes irreversible. On the contrary, if the critical nucleus radius decreases, its vapor pressure will augment and the droplet will evaporate. A mechanical analog of unstable equilibrium is depicted in Figure 7.3.

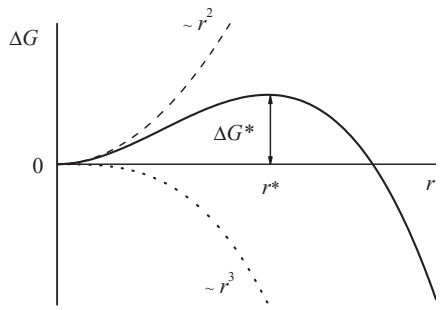


FIGURE 7.2 Plot of the free energy (Gibbs' thermodynamic potential), ΔG, versus droplet radius, r. $\Delta G^* = \max$ determines the radius of the critical nucleus r^*.

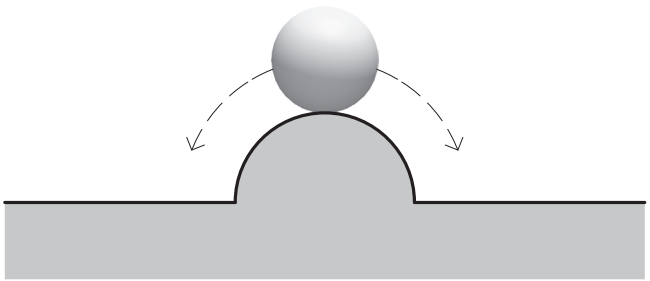


FIGURE 7.3 Mechanical analog of unstable equilibrium. Even an extremely small push exerted on the ball causes it to roll down, either to the right (which corresponds to growth) or to the left (which corresponds to dissolution).

[2]Stable equilibrium is achievable only in sufficiently small closed systems, when any droplet evaporation increases appreciably the vapor pressure and leads to back condensation. Vice versa, if the droplet grows above the critical size, the vapor pressure will decrease below the equilibrium value, which inevitably leads to some droplet back evaporation.

A unified consideration of liquid droplet and crystal nucleation is presented here. For homogeneous crystal nucleation, the total surface free energy (of various shapes) is $\Phi = \Sigma(S^{hkl}\,\gamma^{hkl})$, where S^{hkl} denotes the surface area of the crystal face with Miller indices hkl and γ^{hkl} are the corresponding specific interface energies. Basically, Φ also includes the binding energy of crystal edges and apexes. However, edge energy is negligibly small as compared to surface free energy (see Section 7.7), whereas apex number is constant; it changes only with the change in crystal shape. Then, we have for crystals of any shape

$$\Delta G = -n\Delta\mu + \Sigma\left(S^{hkl}\gamma^{hkl}\right) \tag{7.6}$$

From the postulate that the critical crystal nucleus is determined by the maximum of the total free energy, $d(\Delta G)/dn = 0$, with $\Delta\mu$ and γ^{hkl} being independent on n, one obtains

$$\Delta\mu = \Sigma\left[\gamma^{hkl}\left(dS^{hkl}/dn\right)\right], \tag{7.7}$$

which is a general expression of the Gibbs–Thomson equation for crystals, in which edge and apex energies are neglected.

For combinations of crystallographically equivalent faces, Eqn (7.6) simplifies to

$$\Delta G = -n\Delta\mu + S\gamma, \tag{7.8}$$

where S is the total surface of the new phase and γ is the specific interphase energy. The graphical plot of Eqn (7.8) for spherical crystals is the same one as the presented in Figure 7.2.

It should be emphasized here that Eqns (7.5), (7.6), and (7.8), which also hold true for the energy barrier for nucleus formation, are meaningful for $n \gg 1$ only; otherwise, these equations would assign nonphysical, nonzero work for formation of the monomer, $n = 1$.

Again from the condition for a maximum of the thermodynamic potential, we have

$$dS^*/dn = \Delta\mu/\gamma \tag{7.9}$$

The two simplest cases are usually considered as examples. Firstly, the same Gibbs–Thomson equation is obtained for liquid droplets and spherical crystals:

$$r^* = 2\Omega\gamma/\Delta\mu \tag{7.10}$$

where r^* is the critical nucleus radius and Ω is the volume of a crystal building block (CBB).

Secondly, a convenient model is the so-called Kossel crystal, which is a crystal built by tiny cubic building blocks that are held together by equal forces in a cubic primitive lattice. With an edge length δ of a building block in the crystal lattice, the surface S^* of the critical nucleus constituted of n^* molecules is $S^* = 6\delta^2 n^{*2/3}$, and $\Omega = \delta^3$. Then, Eqn (7.9) yields

$$n^* = 64\Omega^2\gamma^3/\Delta\mu^3 \tag{7.11}$$

The radius r_3^* of the sphere inscribed in the crystal nucleus is $r_3^* = (n^*\Omega)^{1/3}/2 = r^*$.

Combining Eqns (7.8) and (7.9), one obtains the energy barrier ΔG^*_{homo} for homogeneous crystal nucleation:

$$\Delta G^*_{homo} = \gamma[S^* - n^*(dS/dn)] \tag{7.12}$$

Using Eqn (7.10), ΔG^*_{homo} is calculated as

$$\Delta G^*_{\text{homo}} = K\Omega^2\gamma^3/\Delta\mu^2 \tag{7.13}$$

with a coefficient $K = 16\pi/3$ for a sphere and $K = 32$ for the Kossel crystal. (Consideration of fully completed crystals is an implicit assumption in the latter case.) Comparing Eqns (7.13) and (7.11), one sees that for the Kossel crystal

$$\Delta G^*_{\text{homo}} = n^*\Delta\mu/2 \tag{7.14}$$

Gibbs found that the energy barrier for nucleus formation ΔG^*_{homo} amounts to one third of the surface free energy. Now, substituting $\Delta\mu$ from Eqn (7.10) in Eqn (7.8), and with $n^* = 4\pi r^{*3}/3\Omega$, one obtains the energy for reversible isothermal-isobaric formation of a spherical nucleus:

$$\Delta G^*_{\text{homo}} = 4\pi r^{*2}\gamma/3 = S^*\gamma/3 \tag{7.15}$$

Two other expressions can be obtained by substituting $\Delta\mu$ from Eqn (7.10) in Eqn (7.8):

1. For spheres,

$$\Delta G(r) = \Delta G^*\left[3(r/r^*)^2 - 2(r/r^*)^3\right], \tag{7.16}$$

2. For the Kossel crystal, substituting $n\Omega = 4\pi r^3/3$

$$\Delta G(n) = \Delta G^*\left[3(n/n^*)^{2/3} - 2(n/n^*)\right] \tag{7.17}$$

Equation (7.17) will be used for deriving expressions for the nucleation rate in the next subsection.

7.2.3 Rate of Homogeneous Nucleation: Steady-State Nucleation Rate

One of nucleation theory's main purposes is to provide expressions for the nucleation rate, J, which is the number of nuclei that appear in unit volume, 1 cm^3 per unit time, $t = 1$ s. Using a statistical-thermodynamic approach, Volmer [5] further developed the Gibbs thermodynamic nucleation theory. He introduced and popularized the notion of two- and three-dimensional nuclei. By treating the nucleation barrier as activation energy, Volmer derived expressions for the rates of both kinds of nuclei formation. On the basis of Boltzmann's relationship between entropy and probability, taken in the form used by Einstein, Volmer wrote [5]

$$J = A\exp(-\Delta G^*/k_{\text{B}}T) \tag{7.18}$$

where A is a pre-exponential factor. Although A remained unknown in Volmer's statistical-thermodynamic derivation, Eqn (7.18) shows why the nucleation rate J is extremely sensitive to $\Delta\mu$ and γ (compare Eqn (7.13)).

The pre-exponential factor A is revealed by the kinetic derivation of Eqn (7.18). After Volmer and Weber [3] formulated the core kinetic ideas, Frakas [6] looked into the details of the process mechanism. Using the chain reaction idea of Leo Szilard, he attempted to derive the nucleation rate. This basic notion is the root of theoretical considerations made later by Kaischew and Stranski [7], Becker and Döring [8], and Volmer [5].

For the sake of simplicity, steady-state liquid droplet formation from supersaturated vapors is examined further. Evidently, no steady-state process can be realized in a closed supersaturated system because irreversible growth of supercritical nuclei exhausts the monomers. To overcome this obstacle, a feasible physical model has been developed [5,8]. According to this model, all droplets that become larger than the critical size are removed from the system and an equivalent amount of vapors is added immediately, so that the number N of the molecules in the vapor phase is maintained constant. This secures constant supersaturation at all times. The assumption is valid in some practical cases because the number of molecules involved in forming the nuclei is sufficiently small, and the single-molecule depletion has a negligible effect in the earliest stages of nucleation.

The model excludes the possibility for molecule clusters to coalesce and produce larger size aggregates and/or disintegrate in smaller clusters. The clusters can grow or decay by attachment/detachment of single molecules only. In other words, the classical kinetic model describes cluster formation through a succession of steps, each leading to a cluster of slightly (merely by one molecule) larger size. Then, a stationary flux of rising size clusters will flow throughout the system:

$$1 \underset{\nu_2^-}{\overset{\nu_1^+}{\rightleftarrows}} 2 \underset{\nu_3^-}{\overset{\nu_2^+}{\rightleftarrows}} 3 \ldots (n-1) \underset{\nu_n^-}{\overset{\nu_{n-1}^+}{\rightleftarrows}} n \underset{\nu_{n+1}^-}{\overset{\nu_n^+}{\rightleftarrows}} (n+1)$$

where ν^+ and ν^- are the corresponding probabilities (rate constants) for attachment and detachment of single molecules. Taking into account the birth and decay processes of clusters of size n, one obtains the time dependence of the concentration c_n of clusters constituted by $n = 2, 3\ldots$ molecules:

$$\mathrm{d}c_n/\mathrm{d}t = \left(\nu_{n-1}^+ c_{n-1} + \nu_{n+1}^- c_{n+1}\right) - \left(\nu_n^+ c_n + \nu_n^- c_n\right) \tag{7.19}$$

Evidently, the net flux J_n of clusters through the size n is

$$J_n = \nu_{n-1}^+ c_{n-1} - \nu_n^- c_n \tag{7.20}$$

Thus,

$$\mathrm{d}c_n/\mathrm{d}t = J_n - J_{n+1} \tag{7.21}$$

Considering the steady-state process, $\mathrm{d}c_n/\mathrm{d}t = 0$, $J_n = J_{n+1} = J_{\mathrm{st}}$, where J_{st} denotes the steady-state rate (that is frequency) of formation of clusters, which is independent on the cluster size, and thus, includes the formation of critically sized nuclei as well. Therefore, the rate of nucleation J is defined by the flux J^* through the critical size, and the steady state is characterized by

$$\begin{aligned}
J_{\mathrm{st}} &= \nu_1^+ c_1 - \nu_2^- c_2, \\
J_{\mathrm{st}} &= \nu_2^+ c_2 - \nu_3^- c_3, \\
&\ldots\ldots \\
J_{\mathrm{st}} &= \nu_n^+ c_n - \nu_{n+1}^- c_{n+1} \\
&\ldots\ldots \\
J_{\mathrm{st}} &= \nu_{\Lambda-1}^+ c_{\Lambda-1}.
\end{aligned} \tag{7.22}$$

To totally exclude any probability of cluster evaporation, the chain is cut off at some upper limit $\Lambda \gg n^*$, where all supercritical nuclei leave the system, $c_\Lambda = 0$. Becker and Döring [8] multiplied each equation of this system by an appropriate rate constant ratio: the first equation is multiplied by $1/\upsilon_1^+$, the second one by $\upsilon_2^-/\upsilon_1^+\upsilon_2^+$, and the n-th by $\upsilon_2^-\upsilon_3^-\ldots\upsilon_n^-/\upsilon_1^+\upsilon_2^+\ldots\upsilon_n^+$. Then, they summed up the equation system. In doing so, Becker and Döring got rid of the intermediate terms on the right-hand side of the equation. Because $c_\Lambda = 0$, the right-hand site of the sum becomes equal to c_1 (molecules per cm^3), and

$$J_{\mathrm{st}} = c_1 \Big/ \left[\sum_{n=1}^{\Lambda-1} \left(\upsilon_2^-\upsilon_3^-\ldots\upsilon_n^-\right)/\upsilon_n^+\left(\upsilon_1^+\upsilon_2^+\ldots\upsilon_{n-1}^+\right) \right] \tag{7.23}$$

It is well known that the probability υ_{n-1}^+ for the attachment of a molecule from the vapor phase to the surface of a liquid droplet of size $(n-1)$ is determined by the number of collisions per 1 cm^2, times the surface area S_{n-1}:

$$\upsilon_{n-1}^+ = pS_{n-1}/(2\pi m k_B T)^{1/2} \tag{7.24}$$

where m is the atomic (molecular) mass.

Taking into account the fact that the molecules have a nonzero size, Volmer scrutinizes the crossing of a molecule through the spherical intermolecular-interaction boundary that surrounds the droplet [5]. Evidently, a molecule is finally detached from the droplet only after its mass center crosses the said boundary; just then, the liquid surface of a cluster consisting of n molecules shrinks to S_{n-1}:

$$\upsilon_n^- = p_n S_{n-1}/(2\pi m k_B T)^{1/2} \tag{7.25}$$

Using the Gibbs–Thomson equation,

$$p/p_\infty = \exp(2\upsilon\gamma/k_{\mathrm{B}}Tr) \tag{7.26}$$

where υ is the specific volume of a liquid molecule, we have for the critical nucleus, $n^* = n - 1$,

$$\upsilon_n^-/\upsilon_{n-1}^+ = p_n/p = \exp\{(2\upsilon\gamma/k_BT)[(1/r_n) - (1/r^*)]\} \tag{7.27}$$

Keeping in mind that $n\upsilon = 4\pi r^3/3$, we obtain

$$\left(\upsilon_2^-\upsilon_3^-\ldots\upsilon_n^-\right)/\left(\upsilon_1^+\upsilon_2^+\ldots\upsilon_{n-1}^+\right) = \exp\left\{ (2\gamma/k_BT)\left(4\pi\upsilon^2/3\right)^{1/3} \sum_1^n \left[\left(1/n^{1/3}\right) - \left(1/n^{*1/3}\right) \right] \right\} \tag{7.28}$$

In the framework of the CNT, $n^* \gg 1$, and the sum in the right-hand side of this equation can be replaced by an integral. Then, integrating from 0 to n, we obtain

$$\left(\upsilon_2^-\upsilon_3^-\ldots\upsilon_n^-\right)/\left(\upsilon_1^+\upsilon_2^+\ldots\upsilon_{n-1}^+\right) = \exp\left\{ (\gamma/k_BT)\left(4\pi\upsilon^2 n^{*2}/3\right)^{1/3} \left[3(n/n^*)^{2/3} - 2(n/n^*) \right] \right\} \tag{7.29}$$

The comparison with Eqns (7.15) and (7.17) leads to

$$\left(\upsilon_2^-\upsilon_3^-\ldots\upsilon_n^-\right)/\left(\upsilon_1^+\upsilon_2^+\ldots\upsilon_{n-1}^+\right) = \exp[\Delta G(n)/k_BT] \tag{7.30}$$

Let us now note that the equilibrium distribution of heterophase fluctuations can be calculated considering the metastable equilibrium in a slightly supersaturated system,

where no critical nuclei can arise, $J = 0$. Evidently, the condition for (both steady-state process and) equilibrium is a time-independent cluster-size distribution. However, the concentration of clusters of category n is constant only provided the appearance and disappearance rates are equal. Thus, the equilibrium concentration C_n of clusters in the absence of molecular flux in the system is

$$v_{n-1}^{+} C_{n-1} = v_n^{-} C_n \tag{7.31}$$

Equation (7.31) represents the so-called *detailed balance*, according to which at equilibrium, each elementary process should be equilibrated by its reverse process. Rewriting Eqn (7.31) as $C_n/C_{n-1} = v_{n-1}^{+}/v_n^{-}$ and multiplying the ratios C_n/C_{n-1} from $n = 2$ to n gives

$$C_n/C_1 = \prod_{i=2}^{n} \left(v_{i-1}^{+}/v_i^{-}\right) = 1/\left[\left(v_2^{-} v_3^{-} \ldots v_n^{-}\right)/\left(v_1^{+} v_2^{+} \ldots v_{n-1}^{+}\right)\right] \tag{7.32}$$

Thus, Eqns (7.30) and (7.32) tell us that the metastable equilibrium concentration of droplets consisting of n molecules is

$$C_n = C_1 \exp[-\Delta G(n)/k_B T] \tag{7.33}$$

where C_1 is the number of single atoms when the system is in equilibrium.

Bearing in mind that the nucleation rate J_{st} can be roughly estimated from the number density of critical clusters multiplied by the attachment probability (frequency) v^{*+} of a molecule to the critical cluster, we can write (with $C_1 \cong c_1$)

$$J_{st} = C_n^{*} v^{*+} = c_1 v^{*+} \exp[-\Delta G^{*}/k_B T] \tag{7.34}$$

The next step is to specify the expression. With this end in view, we replace the sum in the denominator of Eqn (7.23) with an integral:

$$\left[\sum_{n=1}^{\Lambda-1} \left(v_2^{-} v_3^{-} \ldots v_n^{-}\right)/v_n^{+}\left(v_1^{+} v_2^{+} \ldots v_{n-1}^{+}\right)\right] \approx \int_1^{\Lambda} \left(1/v_n^{+}\right) \exp\left\{\left(\Delta G^{*}/k_B T\right)\left[3(n/n^{*})^{2/3} - 2(n/n^{*})\right]\right\} \mathrm{d}n \tag{7.35}$$

The function $(\Delta G^{*}/k_B T)[3(n/n^{*})^{2/3} - 2(n/n^{*})]$ possesses a sharp maximum in the vicinity of n^{*} and can be expanded in Taylor series. Thus, following the known procedure (e.g., [9,10]) and after some approximations (including the assumption that v_n^{+} is constant, equal to the attachment probability to the nucleus, and thus it can be placed in front of the integral), the integration from $-\infty$ to $+\infty$ (instead from 1 to n) yields

$$J_{st} = c_1 v^{*+} Z \exp(-\Delta G^{*}/k_B T) \tag{7.36}$$

where

$$Z = \left(1/n^{*}\right)\left(\Delta G^{*}/3\pi k_B T\right)^{1/2} \tag{7.37}$$

In the literature, Z is known as the Zeldovich factor, which accounts for the difference between the equilibrium and the actual steady-state numbers of critical nuclei. Zeldovich [11] assumed that only variation by $k_B T$ in both sides around the maximum ΔG^{*} are of interest. The width of the energy barrier maximum is of special interest

because, evidently, only the near-critical clusters contribute the most to the nucleation rate. (For a more rigorous treatment of the problem, see Ref. [12].) In fact, Z is not a large correction; its value is on the order of 10^{-2} for water (see Table 1.1 in Toshew's work [9]).

To conclude, it should be noted that vapor droplet formation is one of the oldest nucleation study topics. The literature on it is extensive and there is not space to review it all here. I will note only that condensation of water droplets is of considerable practical interest as well in view of its technological importance; for instance, to prevent energy efficiency losses, it is essential to calculate the onset of water droplet nucleation in steam turbines [13].

7.2.4 Non-Steady-State Nucleation Rate

Basically, steady-state nucleation can occur only for a short period of time in sufficiently large systems and under the condition that concentration and/or temperature are altered by supercritical nuclei growth in nuclei nearest proximity only, while remaining unchanged in the bulk. Following that short period of time, supersaturation and nucleation rate constantly drop until they finally approach zero—the reason being the growth of supercritical nuclei. Zeldovich [11], Frenkel [14], Turnbull and Fisher [15], Kashchiev [12] and many others have explored the transition periods of the non-steady-state nucleation process, $dc_n/dt \neq 0$. For such a process, Zeldovich and Frenkel regarded the number n of molecules constituting clusters as a continuous variable and replaced the discrete cluster concentration dependence, expressed by Eqn (7.21), with a new differential equation to satisfy the continuity condition:

$$dc(n,t)/dt = -dJ(n,t)/dn \tag{7.38}$$

where $c(n, t)$ and $J(n, t)$ are the size- and time-dependent cluster concentration and rate of cluster formation, respectively.

Correspondingly, Eqn (7.20) is replaced in this case by

$$J(n,t) = v_{n-1}^{+}c(n-1,t) - v_{n}^{-}c(n,t) \tag{7.39}$$

Again for $J = 0$, the detailed balance equation is changed for the non-steady-state case:

$$v_{n-1}^{+}C(n-1) = v_{n}^{-}C(n) \tag{7.40}$$

where $C(n)$ denotes the equilibrium concentration of clusters of size n [14].

Thus, we can replace the detachment probability v_n^- to obtain

$$J(n,t) = v_{n-1}^{+}C(n-1)[c(n-1,t)/C(n-1) - c(n,t)/C(n)] \tag{7.41}$$

Using the approximation $v_{n-1}^{+}C(n-1) \cong v_{n}^{+}C(n)$, we can write for the continuous case

$$J(n,t) \cong -v_{n}^{+}C(n)\partial\,[c(n,t)/C(n)]/\partial\, n \tag{7.42}$$

Combining Eqns (7.38) and (7.42), we obtain the partial differential equation that describes the non-steady-state nucleation process:

$$\partial\, c(n,t)/\partial\, t = \partial\left\{v_{n}^{+}C(n)\partial\,[c(n,t)/C(n)]/\partial\, n\right\}/\partial\, n \tag{7.43}$$

Many authors attempted to solve this equation, but most of them made more or less physically acceptable assumptions and approximations. Perhaps one of the most rigorous solutions was given by Kashchiev [12]. Conducting a profound mathematical examination of the non-steady-state nucleation problem, he concluded that only within the critical region introduced by Zeldovich (of k_BT in both sides around the maximum ΔG^*) does the non-steady-state distribution $c(n, t)$ differ essentially from the steady-state distribution, C_n. Thus, he wrote:

$$J(n,t) = J_{st}\left[1 + 2\sum_{i=1}^{\infty}(-1)^i \exp\left(-i^2 t/\tau_N\right)\right] \tag{7.44}$$

where the parameter τ_N is called *induction time* or *nonstationary time lag*. It characterizes the ability of the system to reorganize itself until producing a steady flow of nuclei

$$\tau_N = 8k_BT/\pi^2\lambda v^{*+} \tag{7.45}$$

where $\lambda = -[d^2\Delta G(n)/dn^2]_{n=n^*} > 0$. After $t > 5\tau_N$, the sum in Eqn (7.44) can be neglected and the steady-state nucleation rate is achieved [12].

7.2.5 Induction Time for Nucleation

The nucleation process cannot become stationary immediately after setting the conditions rendering a steady state. The reason is that when supersaturation is established in the system, cluster size distribution changes first from the equilibrium distribution to that corresponding to the metastable state. The larger the cluster, the longer it takes for the cluster to emerge; hence, an equilibrium distribution of clusters smaller than the critical size has to be attained before the appearance of the first critical nucleus. Furthermore, any supersaturation change provokes an entirely new set of stationary populations of subcritical and critical clusters; the inherited cluster size distribution has to accommodate the new steady state.

Keeping in mind that the total number of nuclei $C_n^*(t) = \int_0^t J\,dt$ and substituting $J(n, t)$ from Eqn (7.44), Kashchiev [12] showed that the integral renders

$$C_n^*(t) = J_{st}\left\{t - \left(\pi^2\tau_N/6\right) - 2\tau_N\sum_{i=1}^{\infty}\left[(-1)^i/i^2\right]\exp\left(-i^2 t/\tau_N\right)\right\} \tag{7.46}$$

Again, the sum in the above equation can be neglected after $t > 5\tau_N$ [12], and C_n^* augments further linearly with time:

$$C_n^*(t) = J_{st}\left[t - \left(\pi^2\tau_N/6\right)\right] \tag{7.47}$$

When τ_N is smaller than the observation time t, that can be neglected:

$$C_n^*(t) = J_{st}t \tag{7.48}$$

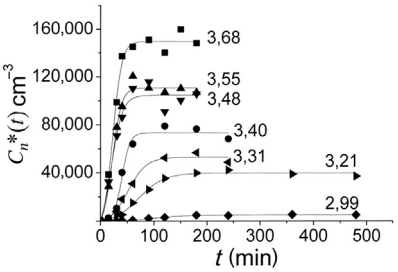


FIGURE 7.4 Number densities $C_n^*(t)$ of nucleated insulin crystals versus nucleation time t (Nanev, unpublished work). Because the nucleation rate J cannot be determined directly, the figure plots experimentally determined number densities of critical nuclei formed per the corresponding nucleation time. The numbers for the curves give concentration ratios, $\ln(c/c_\infty) = \Delta\mu/k_BT$. The measurements plotted were carried out by separation of crystal nucleation and growth stages (e.g., [16]).

Induction time τ_N depends mainly on the attachment probability ν^{*+} to the clusters of critical size (Eqn (7.45))—that is, on the mechanism of molecule supply from the mother phase. Therefore, the system type strongly influences τ_N; induction time can be a small fraction of a second in vapor condensation, but can take many years in a process such as glass crystallization. Moreover, τ_N is inversely proportional to the square of supersaturation or undercooling [9]; that is, the higher the supersaturation, the shorter is the induction time. This is confirmed experimentally (see the intersections of the curves in Figure 7.4 with the x-axis).

Induction time determination is of special interest because τ_N represents the inverse of the frequency of critical nuclei formation, $\tau_N \sim 1/J$. The temptation to measure τ_N might be so irresistible that some authors tend to use the time for observing the first new phase particles as a substitute for the true nucleation induction time. However, the time for observing the first new phase particles is longer than the authentic induction time for nucleation. Induction time per se includes only the time needed to achieve a stationary size distribution of undercritical clusters and the time required for forming nuclei of critical size. In contrast, the time for observing the first new phase particles includes additionally the time for their growth to visible sizes. However loosely defined, the latter depends on the growth rate and on the resolution capability of the observation method used. Only by capturing nuclei at their birth by means of, for instance, laser light scattering, can one render reliable data. Another possibility to directly measure the supersaturation dependence of τ_N has been proposed [16,17]. Nucleation in some other systems will be considered in the following sections.

7.3 Nucleation in some other systems, Nucleation of Gas Bubbles from Superheated Liquids and Boiling

Although evaporation is taking place from any free liquid surface, boiling is a common property of liquids. Therefore, nucleation of gas bubbles is of considerable technological interest. It appears also in aerated waters (including carbonated beverages) and magma.

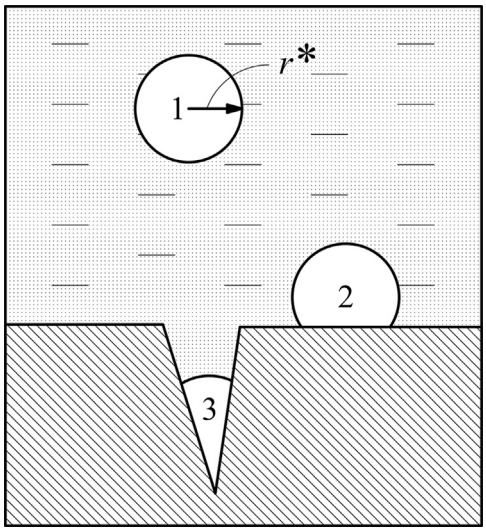


FIGURE 7.5 Homogeneous (1), heterogeneous (2), and nucleation at preexisting gas microcavities (3) are depicted. A finite nucleation energy barrier must be overcome for each nucleus, but significantly less dissolved gas supersaturation is required for the third kind of bubble formation; increasing the supersaturation, a point is reached at which the radius of curvature of each meniscus equals the critical radius, *r**. Due to the smaller volumes of the heterogeneous nuclei, the nucleation threshold for heterogeneous nucleation is significantly lower than the one in the homogeneous case; this is most pronounced with preexisting gas cavities. *Courtesy of Elsevier; license Nr 3377621150350.*

If bubbles of nitrogen form in divers' blood while surfacing too fast, they suffer decompression sickness and may die.

Liquids can be superheated easily. This fact points unequivocally to the existence of a barrier to the formation of a new vapor phase inside the liquids. Although heterogeneous bubble nucleation occurs more easily than the homogeneous one, the basic principles of the homogeneous process are discussed here because they represent a general scientific platform for the phenomenon. For the impact of preexisting gas cavities in surfaces (Figure 7.5), of container walls, of suspended particles, or in the form of metastable microbubbles—called *heterogeneous nucleation*—interested readers may refer to Ref. [18].

In contrast to a liquid droplet, a gas bubble in a bulk liquid is compressible. Therefore, the thermodynamically derived Gibbs–Thomson equation for the vapor pressure of bubbles is expressed again with the liquid molecule volume υ:

$$k_B T \ln(p_b/p_\infty) = -2\upsilon\gamma/r_b, \tag{7.49}$$

where p_b is the gas pressure inside the bubble and r_b is bubble radius. The *significant* difference with respect to droplet formation is constituted in the negative sign of equation's right-hand side.

It is seen that p_b is lower than the equilibrium pressure p_∞ of the plane vapor–liquid interface. The evident equilibrium condition refers to p_b withstanding the sum of

external (usually barometric) pressure p_{ex}, capillary pressure $2\gamma/r_b$, and hydrostatic liquid pressure $\hat{h}\varsigma g$:

$$p_b^* = p_{ex} + 2\gamma/r_b + \hat{h}dg \approx p_{ex} + 2\gamma/r_b \tag{7.50}$$

where $\hat{h}$ is the height of the liquid column above the bubble, d is liquid density, and g is the gravity acceleration of Earth.

According to Eqn (7.49), $p_b^* < p_\infty$. Hence, gas bubble formation inside the liquid may proceed merely under the condition $p_{ex} < p_\infty$, but still positive: $p_{ex} > 0$.[3] Therefore, the bubble nucleus stands in equilibrium with a supersaturated liquid, with the equilibrium being also unstable. With bubble formation driving force $\Delta\mu = (p_b^* - p_{ex})\upsilon$, we obtain the expression

$$\Delta G_{homo}^* = 16\pi\gamma^3\upsilon^2/3\Delta\mu^2 \tag{7.51}$$

which is analogous to Eqn (7.13).

Noting that CNT describes fairly well the gas bubble formation during boiling, Hirth and Pound [20] supposed that the contradictions appearing in the case of droplet formation in vapors may be attributed to the incomplete molecule accommodation on the surface of the arising cluster of the new liquid phase.

7.3.1 Crystal Nucleation in Melts

Crystal nucleation in melts is important from a technological point of view because the nucleation rate predetermines grain size and, hence, the important mechanical properties of cast metal wares. The liquid-to-solid transition involves the appearance of new periodic structure and density changes. Once periodic order appears, the molecules become differently packed (in most cases, more closely) than in the liquid.

It is convenient to use Eqn (7.36) as a basis for the calculation of nucleation rates of melt crystallization. Evidently, the attachment probability ν^{*+} of a molecule from the melt to the (spherical) critical crystal nucleus does not depend on molecule transport, but only on molecule rearrangement in the crystal lattice

$$\nu^{*+} = c_1 4\pi r^{*2}\zeta \exp[-\Delta E_{re}/k_B T] \tag{7.52}$$

where ζ is the product of a frequency factor times the molecule's mean free path in the melt, ΔE_{re} is usually identified with the activation energy for viscous flow, and

$$J_{st} = c_1^2 4\pi r^{*2}\zeta Z \exp[-\Delta E_{re}/k_B T]\exp[-\Delta G^*/k_B T] \tag{7.53}$$

In this case, temperature plays a more significant role than in vapor condensation. When T decreases, it causes simultaneous augmentation of undercooling (Eqn (7.3)) and melt viscosity, and vice versa. The difficulties in the comparison of theoretical and experimental results stem from the presence of impurity particles and from the temperature history of the crystallizing system (*memory effects* [2] and presence of *athermal nuclei* [9]).

[3]Bubble formation under negative external pressure exerted on the liquid is called *cavitation*. This is a mechanical rupture of a liquid; similar to solids, liquid cavitation occurs abruptly (for more details, see Ref. [19]).

7.3.2 Crystal Nucleation in Solutions

Crystal nucleation in solutions is another case of new phase formation for which it is convenient to use Eqn (7.36). In this case, the attachment probability ν^{*+} of a molecule from the solution to the spherical critical crystal nucleus depends on both molecule transport and rearrangement in the crystal lattice. Being a relatively slow process, diffusion is usually the rate-limiting stage in the process. However, before attachments to the nucleus, the species (ions or molecules) also have to become free of their immediate solvent surroundings

$$\nu^{*+} = 4\pi r^{*2} c \varsigma \exp[-\Delta E_{\text{dis}}/k_{\text{B}}T] \tag{7.54}$$

where ΔE_{dis} is the dissolvation energy and ς is the product of vibration frequency times the molecule's mean free path in the solution. Thus,

$$J_{\text{st}} = c^2 4\pi r^{*2} \varsigma Z \exp[-\Delta E_{\text{dis}}/k_{\text{B}}T]\exp[-\Delta G^*/k_{\text{B}}T] \tag{7.55}$$

7.4 Earlier Corrections of the CNT

Operating with lucid and easily understandable ideas, and being relatively simple and easy to use, CNT properly explains the origins of the nucleation barrier and nucleation rate. However, the difficulties with CNT arise mainly because it assumes that data derived from property measurements on macroscopic phases, such as surface tension, structure, and density, can be applied to microscopic clusters containing a limited number of molecules; CNT treats such clusters as small phases cut out of large macroscopic phases. Capillarity—that is, the assumption of the interface between the arising small cluster of the new phase and the parent phase being sharp—is another drawback of CNT. The interface is described simply by a surface with a specific (per unit area) free energy, γ. However, γ is usually not available from direct measurements; for nanosizes, it is experimentally quantifiable only to a very limited extent (e.g., [21,22]). Besides, the γ-value for a curved boundary is relatively independent on the position chosen for the dividing surface, only when its radius of curvature is much larger than the width of the interface transition region. In contrast, the size of the smallest critical clusters is similar to the thickness of the molecularly diffuse interface, so that the dividing surface cannot be sharp. The interfacial width increases toward much larger values near the thermodynamic critical point, which finally brings about spinodal transformation in the unstable regions of the free energy; then, the work of formation for the fluctuation vanishes [2].

Tolman was the first to realize the extent by which the surface tension of a small liquid droplet, γ_{R}, deviates from its planar value, γ_∞. Using thermodynamic considerations to account for this circumstance, he introduced a curvature correction δ_{T} [23]:

$$\gamma_{\text{R}} = \gamma_\infty/(1 + 2\delta_{\text{T}}/R_{\text{e}}) \tag{7.56}$$

where δ_{T} is the difference in the radii of the equimolecular dividing surface [1] and the so-called surface of tension [24]; R_{e} is the equimolar radius of the liquid droplet. δ_{T} is also

known as the *Tolman length*. Tolman's correction represents the first term of the expansion in power series of the surface energy, Eqn (7.56), in terms of surface curvature:

$$\gamma_R \approx \gamma_\infty (1 - 2\delta_T / R_e) \tag{7.57}$$

Equations (7.56) and (7.57) show that the surface energy starts to deviate appreciably from its planar value when the droplet radius is on the order of δ_T. It should be noted here that, according to Eqns (7.13) and (7.18), a variation of only 10% in γ can alter the nucleation rate by many orders of magnitude. Due to this tremendous effect, the issue is still of enormous interest; using the keyword *Tolman length*, many dozens of papers devoted to the topic can be found online. Despite the large number of works concerned with Tolman's correction, there is hardly a generally accepted view on the problem; the topic is still debated [25] and a plethora of experimental techniques have been employed to perform nucleation rate measurements (e.g., see the first 12 references listed in Ref. [26]). Basically, they show that although CNT is valid near equilibrium where the critical fluctuation and the work required for it are large, a different behavior is observed moving further away from equilibrium.

Mutaftschiev [19] also studied the surface energy size dependence and the limits of capillary approximation applicability. Taking into account the corrections needed for this dependence in view of the discrete (atomic) constitution of extremely small clusters, he added a constant in the classical expression for the thermodynamic potential of the new condensed phase [19]. It is remarkable that, being a constant, this correction term does not change the derivation of the Thomson–Gibbs equation (Section 7.2.2).

Another disadvantage of CNT is its intrinsic feature to allow for a miscalculation of the degree of freedom of nucleating clusters. Because for homogeneous nucleation the nucleus can form around any one of the molecules present, it can appear anywhere in the homogeneous system with the same probability. Moreover, a free cluster is not static but translates and rotates freely in the entire volume of the parent phase, thus contributing to a higher system disorder. It has higher entropy than the one ascribed by CNT and, therefore, lower free energy than a small phase cut out of a large condensed phase. Thus, classical thermodynamics overestimates the cluster's excess free energy. With this in mind, Lothe and Pound [27] introduced the so-called Lothe and Pound (correcting) factor, Γ. In the case of a homogeneous water droplet nucleation from vapors, the calculations yield $\Gamma \approx 10^{17}$, but it turns out that it affects only the pre-exponential factor in the nucleation rate equation [28].

A typical assumption used in the kinetic models of CNT is that cluster–cluster interactions can be neglected because of their rarity. An alternative to this assumption is the so-called chemical approach presented by Frenkel [14]. He considered the molecule clusters as polymer molecules that obey chemical reaction laws. His approach allows aggregation of clusters of any size, but not only attachment/detachment of single molecules to/from the clusters. Mutaftschiev [29] discussed the merits and inconsistencies of the chemical approach. With this approach, it becomes clear why the critical nucleus plays the role of an "activated complex" in the theory of the chemical

reactions; the critical nucleus is the "molecule" of maximum energy and minimum concentration.

However, CNT is not an anachronism. Firstly, a quite surprising validity of the capillary approximation has been noticed by Bonissent and Mutaftschiev, even for particles of only a few atoms [30]. In the authors' opinion, this seems to be due to a kind of compensation effect between energy and entropy (for closer look at the topic, see Ref. [30]). As far as the cluster coagulation effect on the nucleation rate is concerned, it should be noted that it is not only too infrequent to be concurrent with single molecule attachment, but at least for one case—namely, the diffusion controlled aggregation of protein crystalline clusters—it is highly improbable (see Section 7.12.4 and Ref. [31]).

7.4.1 Some Recent Nucleation Theories

Due to space limitations, only some novel ideas will be mentioned here in addition to the corrections mentioned above. For instance, contrary to the CNT assumption that crystal embryo structure and properties are the same as that of the bulk macroscopic crystal phase, some authors assumed that the embryo's structure differs significantly (e.g., Ref. [32]). A fractal structure of the heterogeneous nucleus has been suggested as well [33]. However, polyhedral nuclei of apoferritin crystals have been observed by atomic force microscopy [34]; the authors have shown that crystals of near-nucleus size exhibit the same *fcc* structure as the bulk apoferritin crystals.

Numerous investigations have been carried out to solve specific nucleation problems, the recent methods of choice being predominantly molecular dynamic and Monte Carlo computer simulations (e.g., [26,35]), just to mention a few. McGraw and Laaksonen [36] considered a diffusive liquid droplet of changing density. Using the Gibbs surface dividing method [24] and density functional theory, the authors found a temperature-dependent correction to the nucleation barrier ΔG^*, which, however, is independent of the nucleus size (the latter is assumed to be the same as the one found in CNT). The surface energy γ_d correction introduced by the authors [36,37] is considered to be an alternative to the Tolman correction (Eqn (7.56)):

$$\gamma_d = \gamma_\infty + f(T)/S^* \tag{7.58}$$

where f is an arbitrary function.

Walton and Rhodin pointed out that although the CNT is capable of describing the nucleation phenomenon at low supersaturations, an atomistic approach is more common at high supersaturations [38–40]. They developed a statistical nucleation theory for vapor deposition of thin films, according to which, at high impingement rates of atoms on the substrate, even a single atom can be a critical nucleus. Moreover, point defects on the crystal surface can trap adatoms, which then appear as supercritical centers of irreversible growth [41].

Milchev and Stoyanov [42,43] (see also Ref. [44]) have adopted the atomistic approach for the case of electrochemical nucleation, where high overvoltages are applied as well. Scrutinizing the discrete character of the cluster size alteration for single-digit molecule

numbers, they pointed out that the n^* versus $\Delta\mu$ relationship is not continuous, but a stepwise one (Figure 7.6). It was shown, too, that a supersaturation interval, not a fixed supersaturation, corresponds to each critical nucleus; the smaller the nuclei, the wider the intervals and vice versa. Also, the supersaturation dependence of the steady-state nucleation rate is altered for extremely small nuclei. A broken line is observed in the atomistic case with high $\Delta\mu$, while according to CNT $\ln J_{st}$ should change smoothly (Figure 7.7).

Using numerical simulations, ten Wolde and Frenkel [45] have shown that density fluctuations in a protein solution located near the liquid–liquid boundary may induce the formation of a high-density protein drops surrounded by low-concentration solution. On this basis arises the so-called two-stage nucleation mechanism of protein crystal nucleation. The first step in this mechanism consists of the separation of dense protein liquid drops from the bulk of the solution of the nucleating substance. The second step is crystal nuclei formation inside the high-concentrated regions,

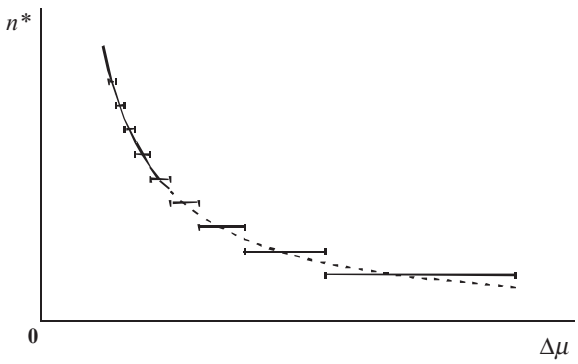


FIGURE 7.6 Critical nucleus size n^* versus $\Delta\mu$. The smooth line is the plot of the Gibbs–Thomson equation. *Courtesy of Elsevier; license Nr 3375810205283.*

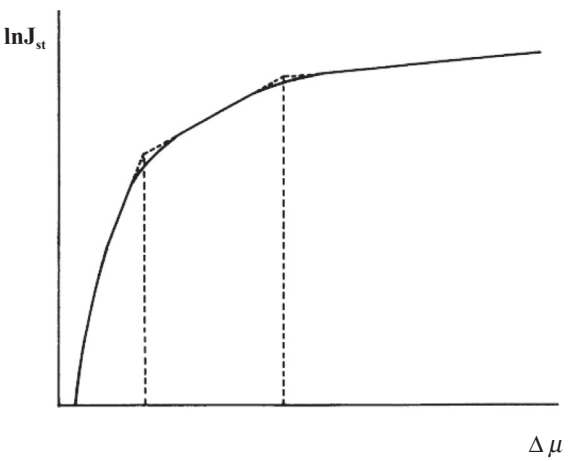


FIGURE 7.7 Schematic representation of the $\ln J_{st}$ vs. $\Delta\mu$ relationship. The slope of each straight line section of the curve gives the number of atoms in the critical nuclei. *Courtesy of Elsevier; license Nr 3375800275481.*

which proceeds at a reduced energy barrier. Because of the slow ordering kinetics of the crystalline phase, this step determines the nucleation process rate [46]. Based also on experimental observations, the two-step nucleation mechanism remarkably resembles Ostwald's rule of stages (e.g., [19]). The topic is considered in full detail by Vekilov in Chapter 19 of Volume 1B.

7.5 Molecular-Kinetic Approach to Crystal Nucleation

A significant advancement in CNT has been marked by the molecular-kinetic approach to crystal nucleation, equilibrium, growth, and dissolution developed by Stranski and Kaischew [47–49]. First, Kossel [50] and Stranski [51] recognized, simultaneously and independently from each other, the importance of the unique position on the crystal surface known nowadays as a kink (kink site). Kossel called it a *repeatable step* because no change in surface geometry and energy takes place if an individual CBB is added to/or detached from the kink on an "infinitely" large crystal face. For the same position, Stranski used the notion of *half-crystal position* (Figure 7.8). This position on the crystal face determines the thermodynamic equilibrium with the ambience of a sufficiently large crystal (i.e., its chemical potential) because, evidently, under equilibrium the statistical probabilities for its occupation or nonoccupation are equal. Thus, with a sufficiently large crystal—so large that the energy contributions of the particles situated at the ends of its faces can be ignored—any crystal can be grown or dissolved reversibly by repetitively attaching/detaching CBB at the kink site; this proceeds with the same attachment or detachment energy, respectively, which is measured as the work of separation $\varphi_{1/2}$. Indeed, $\varphi_{1/2}$ depends on crystal lattice structure. For instance,

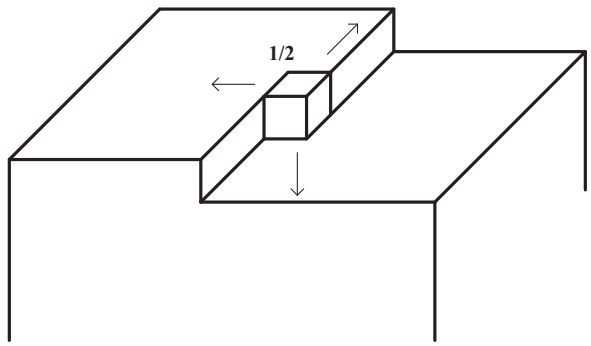


FIGURE 7.8 Schematic presentation of a half-crystal position (1/2) on a face of "infinitely" large crystal. CBB in this position is as follows (see the arrows): (1) at the end of a half CBB row (starting backwards); (2) connected to the ledge of a half-crystal lattice plane (situated on the left); and (3) standing on the half of the "infinite" crystal (beneath). To complete the crystal, one has to add the three missing crystal half-parts. Then, the atom (ion, molecule) under consideration would stand in crystal bulk. Hence, CBB in the half-crystal position is bound exactly two times less strongly than the atom (ion, molecule) standing in the crystal bulk and has an equal number of saturated and dangling bonds.

CBB at the kink site in the Kossel crystal is connected to three first-nearest neighbors (Figure 7.8). Taking into account only this kind of interaction, its bond energy is $\varphi_{1/2} = 3\psi$, where ψ denotes the energy of a single bond between two CBBs.

Kaischew [52] revealed the quantitative relationship between chemical potential, vapor pressure of the "infinitely" large crystal face, p_∞, and $\varphi_{1/2}$ for monoatomic vapors and for temperatures for which crystal energy is negligible:

$$\mu_\infty^{\mathrm{c}} = \mu_{\mathrm{o}} + k_{\mathrm{B}}T \ln p_\infty = -\varphi_{1/2} + k_{\mathrm{B}}T \ln\left[(2\pi m)^{3/2}(k_{\mathrm{B}}T)^{5/2}/\hbar^3\right] \tag{7.59}$$

where μ_∞^{c} is the chemical potential of an "infinitely" large crystal, μ_{o} is the standard chemical potential, and $\hbar$ is Planck's constant. Thus, at $T = 0$, $\mu_\infty^{\mathrm{c}} = -\varphi_{1/2}$.

The situation with small crystals is completely different. The smaller the crystal, the greater are the deviations in the work of separation at the beginning and the end of the crystal face as compared with the work of separation from the half-crystal position. Taking into account this fact (and the principle of detailed balance), Stranski and Kaischew [47–49] suggested that, under equilibrium conditions, the probability of attachment of a whole new lattice plane on a crystal face should be equal to the probability of its detachment. Consequently, they calculated the energetic parameter that determines the equilibrium of small crystals with the surrounding media as the mean value of the work of separation (MWS), φ, averaged per building block of the corresponding crystal face. In doing so, Stranski and Kaischew divided the total work of separation involved in disintegrating the uppermost lattice layer of any face belonging to the crystal equilibrium form by the total number of blocks in that layer. Being equal for every face of the crystal, MWS determines its equilibrium shape (see Section 7.6).

Figure 7.9 depicts the procedure for calculating MWS with a (100) face of a Kossel crystal. CBBs are removed successively, one by one $(n_3 - 1)^2$ CBB (namely all white cubes in Figure 7.9(a)), starting with the atom situated on the right-hand front crystal apex. Detaching each of them from its three neighboring cubic CBB, a separation work amounting 3ψ is performed per each single cube. Then, the two rows of $2(n_3 - 1)$ dashed cubes in Figure 7.9(b) are separated, performing work 2ψ for each. Finally, the last remaining cube (the black one at the left-behind crystal apex in Figure 7.9(c)) is

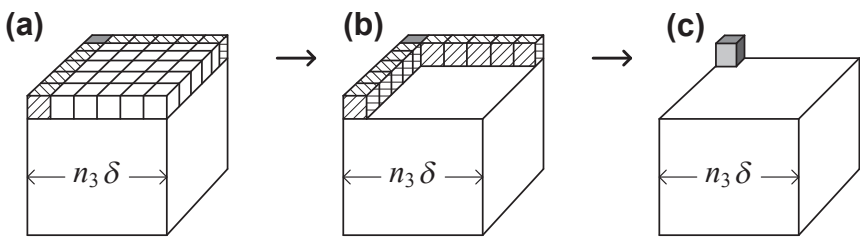


FIGURE 7.9 Three stages of detachment from a Kossel crystal of a whole (100) lattice plane, constituted of n_3^2 building blocks.

separated, performing work ψ. Now, dividing the sum of those three works by the total number of n_3^2 particles, we obtain the MWS for that crystal plane:

$$\varphi = \left[3\psi(n_3-1)^2 + 2\psi 2(n_3-1) + \psi\right] \Big/ n_3^2 = 3\psi - (2\psi/n_3) = \varphi_{1/2} - (2\psi/n_3) \tag{7.60}$$

It is seen that the smaller the crystal, the lower the MWS. In contrast, φ approaches $\varphi_{1/2}$ with a crystal size increase.

Keeping in mind the analogy with Eqn (7.58), Kaischew [52] has shown that in the case of finite size crystals, MWS appears instead of $\varphi_{1/2}$. Thus, MWS determines the vapor pressure, p, of such crystals (see also [10]):

$$\mu_{\mathrm{f}}^{\mathrm{c}} = \mu_{\mathrm{o}} + k_{\mathrm{B}}T \ln p = -\varphi + k_{\mathrm{B}}T \ln\left[(2\pi m)^{3/2}(k_{\mathrm{B}}T)^{5/2}/\hbar^3\right] \tag{7.61}$$

The supersaturation is

$$\Delta\mu = \mu_{\mathrm{f}}^{\mathrm{c}} - \mu_{\infty}^{\mathrm{c}} = k_{\mathrm{B}}T \ln(p/p_{\infty}) = \varphi_{1/2} - \varphi = 2\psi/n_3 \tag{7.62}$$

Stranski and Kaischew [47–49] conceived the concept of MWS to establish a link between the molecular kinetic theory of crystal nucleation and growth and the thermodynamic-statistical treatment of the problems, given by Gibbs [1] and Volmer [5]. To demonstrate this link, the specific surface energy γ, limited to energetic interaction between first-nearest neighbors in the Kossel crystal, is defined as

$$\gamma = \psi/2\delta^2 \tag{7.63}$$

Thus, $\psi = 2\gamma\delta^2$, and

$$\Delta\mu = 4\delta^2\gamma/n_3 \tag{7.64}$$

Keeping in mind that the diameter of the sphere inscribed in the Kossel crystal is $2r^* = n_3^*\delta$, and $\Omega = \delta^3$, Eqn (7.64) represents the Gibbs–Thomson equation (compare Eqn (7.10)).

The molecular-kinetic approach also enables the calculation of the energy barrier for crystal nucleation, $\Delta G_{\mathrm{homo}}^*$. Stranski [53] has defined the free energy Φ as the difference between the total binding energy of a cluster as if all its building blocks are in the bulk of the infinitely large crystal, expressed by the term $n_3^3\varphi_{1/2}$, and the energy of the bonds in the real small cluster, *U:*

$$\Phi = n_3^3\varphi_{1/2} - U, \tag{7.65}$$

where $U = \sum_{i=1}^{n}\varphi_i$ is the disintegration energy of the entire crystalline cluster into $n = n_3^3$ individual building blocks. Evidently, this difference gives the number of unsaturated dangling bonds on the outside of the cluster, multiplied by the energy necessary to break a bond. Note that Stranski's relationship is universal, in the sense that it applies equally well to large and one-digit molecule crystal clusters. Replacing Eqn (7.65) with Eqn (7.5) and using the Gibbs–Thomson equation presented in the form of Eqn (7.62), $\Delta\mu = \varphi_{1/2} - \varphi$, one obtains the energy barrier for crystal (including nucleus) formation:

$$\Delta G_{\mathrm{homo}}^* = {n_3}^{*3}\varphi - \sum_{i=1}^{n}\varphi_i. \tag{7.66}$$

Taking into account only the first-nearest neighbor interactions in a Kossel crystal, one can see that the separation work of a CBB situated on the eight cube vertices is $3\psi = \varphi_{1/2}$ (Figure 7.9(a)); that is, the cubic shape is stable [54]. The total binding energy of a Kossel crystal possessing n_3 CBB in the edge is

$$\sum_{i=1}^{n} \varphi_i = 3n_3^2(n_3 - 1)\psi \tag{7.67}$$

Using Eqns (7.66) and (7.60), Kaishew [54] obtained

$$\Delta G^*_{\text{homo}} = n_3^3[3\psi - (2\psi/n_3)] - 3n_3^2(\mathrm{n}_3 - 1)\psi = n_3^2\psi \tag{7.68}$$

Replacing n_3 from Eqn (7.62), $n_3 = 2\psi/\Delta\mu$, yields

$$\Delta G^*_{\text{homo}} = 4\psi^3/\Delta\mu^2 \tag{7.69}$$

Thus, one finally obtains [54]:

$$\Delta G^*_{\text{homo}} = 32\Omega^2\gamma^3/\Delta\mu^2 \tag{7.70}$$

Generally, applying the MWS method, Stranski and Kaischew arrived at the same conclusions for crystal nucleation that have already been known from CNT, and they have been able to develop further kinetic notions for the process. Moreover, operating with interaction energies in the crystal lattice only (thus avoiding the use of macroscopic surface tension values for small clusters), the MWS approach overcomes one of the deficiencies of CNT. Indeed, the absolute values of the intralattice interactions are known merely for limited crystal types (e.g., for ionic crystals [55]), but Stranski and Kaischew required to have their relative strength only. Hence, they considered the problems of crystal nucleation from a more realistic standpoint than CNT.

Perhaps the single shortcoming of the Stranski–Kaischew's theory is that it is limited to complete and highly symmetrical crystals. Excluded from the considerations remain vicinal surfaces on crystals. However, these limitations are not of principle importance. Firstly, the incomplete clusters would have an increased number of dangling bonds, which would be attractive sites for subsequent attachments until a complete shape would be attained. Secondly, even using the simplest models, such as the Kossel crystal, Stranski–Kaischew's considerations reveal the general features and trends of the nucleation phenomenon (and crystal growth and evaporation as well). This is shown here taking into account only the first-nearest neighbor interactions, but the same results have been obtained using first-, second- and third-neighbor interactions.

Although rarely used nowadays, as seen from the brief introduction provided here (see also Ref. [56]), the MWS approach is not obsolete. Based on this approach, a consideration of protein crystals equilibrium shapes is presented in Section 7.12.2 (see also Ref. [57]). Another example for an MWS application is the calculation of the energy barrier for protein crystal nucleation (Section 7.12.3). As seen in the next three sections, the predictive and explanatory power of the MWS approach should not be underestimated. However, the original works of Stranski and Kaischew are not very accessible (they are in Bulgarian and German). Therefore, the brief introduction provided above may be useful to the reader.

7.6 Equilibrium Shape of Crystals

Gibbs [1] showed thermodynamically that, at a constant crystal volume, the equilibrium crystal shape is established under a minimum in the total surface free energy, $\Phi = \Sigma(S^{hkl}\ \gamma^{hkl}) = \text{minimum}$. His idea has been elaborated further in a geometric construction, now known as the Gibbs–Curie–Wulff theorem. According to this theorem, the equilibrium shape of a crystal can be depicted by putting crystal faces at distances h_i from a central crystal (so-called Wullf's) point, with h_i being strictly proportional to the specific surface free energies γ_i of the corresponding crystal face. Then, the innermost body obtained is the equilibrium crystal shape. Wullf's theorem is expressed as

$$\gamma_i/h_i = \text{const}, \text{ or } \gamma_1{:}\gamma_2{:}\gamma_3\ldots = h_1{:}h_2{:}h_3\ldots \tag{7.71}$$

The derivation is simple (e.g., see Ref. [10]) but lengthy and is omitted here.

Chapter 5 (this volume) by Einstein is devoted to a thorough consideration of the issue. Therefore, it will not be reviewed here in any detail, but the underlying physical ideas of Stranski and Kaischew's determination of the equilibrium shapes of crystals will be discussed; the reason is their simplicity [58]. The principle is that *no* CBB whose bonding with the crystal is looser than the MWS can belong to the equilibrium shape because the vapor pressure of such crystals would be higher than the equilibrium pressure. From this standpoint, Stranski and Kaischew started with an arbitrary crystal shape and successively removed all CBB—the separation work of which is smaller than MWS. In doing so, they revealed the equilibrium shape faces. However, depending on the starting crystal shape, some nonequilibrium faces may still remain. Therefore, the areas of all faces have been varied until the same MWS is reached for all of them. Due to the disappearance of all nonequilibrium faces, this procedure leads to the true equilibrium crystal shape (see also Ref. [59]). In addition, Stranski and Kaischew showed [58] that, with an increase in the supersaturation, the equilibrium shape becomes simpler because some faces disappear from it; due to the diminishing nucleus size, the latter shrinks to a CBB size.

The equilibrium shapes of crystals nucleated on a foreign substrate are considered in Section 7.8.

7.7 Two-Dimensional Crystal Nucleation

Although he did not systematically consider nucleation kinetic problems, Gibbs noticed that the growth of a face on a crystal must be a periodic process, which is accomplished by a successive formation and spreading of crystal layers. Moreover, he pointed out that the formation of every new lattice layer is associated with surmounting an energetic barrier (although it is not high). In 1927, Brandes reconsidered the possibility of two-dimensional (2D) nuclei formation and found that the energy cost, ΔG_2^* was precisely half of the total edge energy:

$$\Delta G_2^* = (1/2)\Sigma(\chi_i \ell_i) \tag{7.72}$$

where χ_i is the specific edge energy of the i-th edge and ℓ_i is its length.

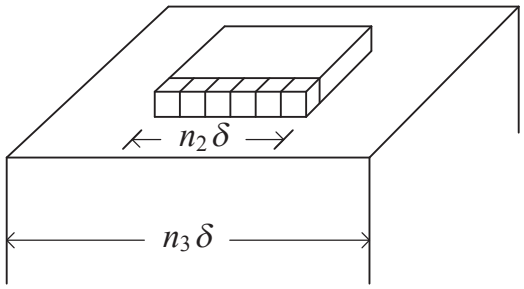


FIGURE 7.10 2D crystal on the (100) face of the Kossel crystal. n_2 is the number of CBBs in the edge row of the 2D crystalline cluster; n_3 is the number of CBBs in the Kossel crystal edge.

There are several different ways to obtain this result [5,8]. To demonstrate how fruitful the MWS method can be, we use it here. In analogy with deriving the MWS method for three-dimensional (3D) crystals, Stranski and Kaischew [48] suggested that, under equilibrium conditions, the probability for evaporation of all CBB that form the edge of a 2D crystalline cluster has to be equal to the probability for their deposition. Using as a model a 2D crystalline cluster of length $\ell_2 = n_2\delta$, formed onto (100) face of the Kossel crystal (Figure 7.10) and denoting the corresponding MWS as φ_2, Stranski and Kaischew determined it to be

$$\varphi_2 = 3\psi - \psi/n_2 = \varphi_{1/2} - \psi/n_2 \tag{7.73}$$

Analogically to Eqn (7.62) for the 2D case,

$$\Delta\mu = k_B T \ln(p/p_\infty) = \varphi_{1/2} - \varphi_2 = \psi/n_2 \tag{7.74}$$

Stranski and Kaischew [60] realized that the condition for a simultaneous equilibrium between the 3D-nucleus, 2D-nucleus, and vapor phase is the equality of the two MWS, $\varphi = \varphi_2$, or for the Kossel crystal $n_2 = n_3/2$ (compare Eqns (7.62) and (7.74)). From this standpoint, the work for the formation of a 2D nucleus upon a completed (100) face of a Kossel crystal is calculated by a conceivable separation of the process in three stages (Figure 7.11), with the first step being the attachment of a single atom from the vapor phase (Figure 7.11(a)). From an energy perspective, it is the most difficult step because the required separation work, equal to ψ, is substantially lower than φ_2, and the single atom should be removed. Therefore, a free energy fluctuation amounting $\varphi_2 - \psi = 2\psi - \psi/n_2$ is required for the deposition of the first CBB on the

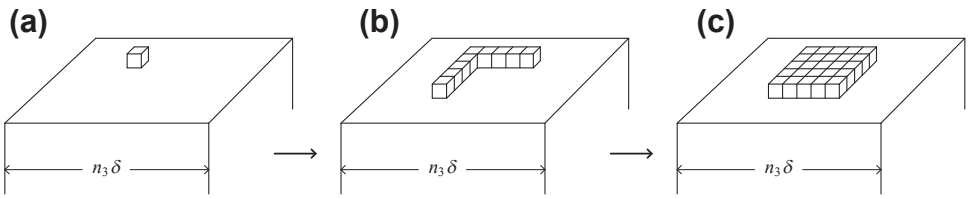


FIGURE 7.11 Formation of a 2D nucleus upon a completed (100) face of a Kossel crystal.

completed (100) face of a Kossel crystal (Figure 7.11(a)). The next step is the formation of two edge rows of the arising 2D nucleus (Figure 7.11(b)). It requires binding energy of 2ψ per CBB, so that the necessary energy fluctuation is $\varphi_2 - 2\psi = \psi - \psi/n_2$. Multiplying by the number $2(n_2 - 1)$ of such particles, one obtains the total energy fluctuation as $2(n_2 - 1)(\psi - \psi/n_2)$. Finally, the attachment of $(n_2 - 1)^2$ CBB that complete the 2D nucleus (Figure 7.11(c)) requires binding energy of 3ψ per CBB, and the necessary energy fluctuation is $\varphi_2 - 3\psi = -\psi/n_2$. Thus, the total energy fluctuation for this stage is minus $(n_2 - 1)^2(\psi/n_2)$. Summing all that up, Stranski and Kaischew show that the total energy fluctuation needed for the formation of the complete 2D nucleus is $\Delta G_2^* = n_2\psi$. Using the definition for the specific edge energy χ as $\chi = \psi/2\delta$ and with $\ell_2 = n_2\delta$, the authors [60] obtained:

$$\Delta G_2^* = 2\chi\ell_2 \tag{7.75}$$

This is Brandes' Eqn (7.72) for the Kossel crystal. Note that exactly the same result is obtained considering the formation of the 2D nucleus in an alternative way, such as dissolving its surrounding crystal lattice plain [60]. It should be emphasized that combining $\chi = \psi/2\delta$ with $\gamma = \psi/2\delta^2$, from Eqn (7.62) yields $\chi = \delta\gamma$. Because δ is on the order of 10^{-7} cm, we see that $\chi \ll \gamma$.

Burton, Cabrera, and Frank [61] have pointed out that in calculating ΔG_2^* Volmer [5] and Becker and Döring [8] neglected the conformational entropy. This is equivalent to an assumption that the nucleus shapes are independent of temperature, and that all shapes are like the nucleus shape at $T = 0$ K. A more rigorous treatment of the problem [61] shows that the inclusion of the conformational entropy rendered only a small correction of ΔG_2^* (multiplication by a factor of 0.8), and that the equilibrium nucleus shape depends on the temperature only, but not on the supersaturation. The shape is completely polygonized at $T = 0$ K and becomes increasingly rounded with temperature increase [61]. As a matter of fact, this circumstance was already known to Gibbs, who noticed that the apexes of a crystal at equilibrium with the surrounding media should be slightly rounded on a molecular scale.

Before the knowledge that real crystals are imperfect, Volmer suggested that 2D nucleation is a mandatory requirement for the growth process [5]. Frank's idea that the self-perpetuating steps originating from screw-dislocation emergence points make the nucleation mechanism unnecessary for crystal growth [62] sparked a new era in modern crystal growth theory, in which the interest in 2D nucleation declined significantly. However, the physical reality of 2D nucleation has been proven by experiments performed with free of screw-dislocation emergence points silver (100) and (111) crystal faces [63] and the (0001) face of cadmium crystals [64]; such faces grow by means of the so-called capillary technique for electrocrystallization.

Volmer also considered the formation of negative 2D nuclei (*Lochkeime* in German), which have to appear on sufficiently large close-packed crystal faces under considerable undersaturation. His suggestion has been confirmed experimentally with perfect faces of *p*-toluidine crystals; local undersaturation has been evoked using a directed sharp

air-blow, and the evaporation depressions created on the crystal faces were attributed to negative 2D nucleation [65]. However, even easier than on a perfect crystal faces, 2D nuclei arise also on emergence points of edge dislocations. The reason is that, due to the accumulated dislocation strain energy E_d, the chemical potential increases locally on such places. To elucidate the problem, Cabrera [66] described the change in the thermodynamic potential as

$$\Delta G = -\pi r_e^2 đ\Delta\mu/\Omega + 2\pi r_e đ\gamma' - hE_d \tag{7.76}$$

where r_e is the radius of the empty disc that arises on the emergence point of the dislocation, $đ$ is the hollow depth, and γ' is the energy of hollow periphery, which is usually set as $\gamma' \approx \gamma$. With the elastic strain energy, E_{el} and dislocation core energy, E_{core}, the total edge dislocation energy per unit length is $E_d = E_{el} + E_{core} = \breve{G}b_B^2 \ln(r_e/\rho_o)/4\pi(1-\xi) + E_{core}$, where $\breve{G}$ is the shear modulus (modulus of rigidity), b_B is the Burgers vector of the dislocation, ρ_o is the radius of the so-called core of the dislocation, and $\xi \approx 0.3$, being Poisson's cross-contraction ratio. From the condition for a maximum of ΔG and keeping in mind that E_{core} is a constant, two equilibrium radii arise [66]:

$$r^*_{min,\,max} = \left[1 \mp \left(1 - 4\rho_F/r_2^{-*}\right)^{1/2}\right] r_2^{-*}/2 \tag{7.77}$$

where $\rho_F = \breve{G}b_B^2/8\pi^2\gamma'(1-\xi)$ is Frank's hollow core radius and r_2^{-*} is the radius of the negative 2D nucleus on the perfect crystal face. Because r^*_{min} rises and r^*_{max} diminishes with an undersaturation increase (the latter diminishing r_2^{-*}), a critical undersaturation can be reached when $r^*_{min} = r^*_{max}$, and $r_2^{-*} = 4\rho_F$. The energy barrier for nucleation disappears and a spontaneous formation of hollow disks on the edge dislocation emergence points proceeds with further undersaturation augmentation.

7.8 Heterogeneous (Substrate) Nucleation, the Equilibrium Shape of Crystals on Supports, and Energy Barriers for Heterogeneous Nucleation

Homogeneous nucleation occurs very rarely. Much more frequent is heterogeneous nucleation, occurring on foreign surfaces of different origins, such as impurity particles, container walls, etc. (Ions also activate nucleation, the most famous example being droplet formation in the so-called Wilson camera [5].) The reason for the highly predominating heterogeneous nucleation is that ΔG^*_{heter} can be very much less than ΔG^*_{homo}.

Again, Gibbs [1] laid the foundations of the theory describing how liquid droplets nucleate on interfaces between two bulk liquids and on a solid support. The theory has been further developed by Volmer [5]. The expression for the heterogeneous nucleation work is

$$\Delta G^*_{heter} = \Delta G^*_{homo}\left(0.5 - 0.5\cos\beta - 0.25\sin^2\beta\cos\beta\right) \tag{7.78}$$

where ΔG^*_{homo} is given for spherical drops from Eqn (7.13) and β is the wetting angle. For $\beta = 0$, meaning complete wetting, $\Delta G^*_{\text{heter}} = 0$ and vapor condensation requires only a relatively small barrier for 2D liquid nucleation. For $\beta = \pi$, meaning complete non-wetting, $\Delta G^*_{\text{heter}} = \Delta G^*_{\text{homo}}$—that is, the substrate has no effect on droplet nucleus formation. For any β value between 0 and π, nucleation of liquid droplets proceeds heterogeneously in an easier manner than in a bulk vapor phase.

Due to its enormous practical significance, theoretical consideration of heterogeneous crystal nucleation has been provided by many authors (e.g., [67–69], just to mention some). Using the MWS method (Section 7.5) and the Kossel crystal model, Kaishew calculated equilibrium crystal shapes and energy barriers for crystal nucleation on foreign substrates [54,70]. Later, his significant contribution was called the Wulf–Kaischew theorem [68]. Here, Kaischew's work is presented in brief because his original papers [54,70] are in Bulgarian and are not accessible to the general scientific audience.

Because any molecule on a nucleus surface has to occupy an equilibrium position, at a given supersaturation the kind and size of nucleus faces contacting the ambient phase only have to remain the same (Figure 7.12). As for the four faces in Figure 7.13(a) contacting with the support, each built by $n'n_3$ CBB (with n' being the number of CBBs in the normal to the support crystal edges) Kaischew calculated the MWS, φ', as

$$\psi' = [n'n_3\psi + n'(n_3 - 1)\psi + n_3(n' - 1)\psi + n_3\psi']/n'n_3 = 3\psi - [(\psi - \psi')/n' + \psi/n_3] \tag{7.79}$$

where ψ' is the detachment energy of a single CBB from the support. Because under equilibrium the MWS of these faces must be equal to the MWS of the topmost crystal face contacting with the vapor phase only, it follows according to Eqns (7.62) and (7.79) that

$$\Delta\mu = 2\psi/n_3 = (\psi - \psi')/n' + \psi/n_3 \tag{7.80}$$

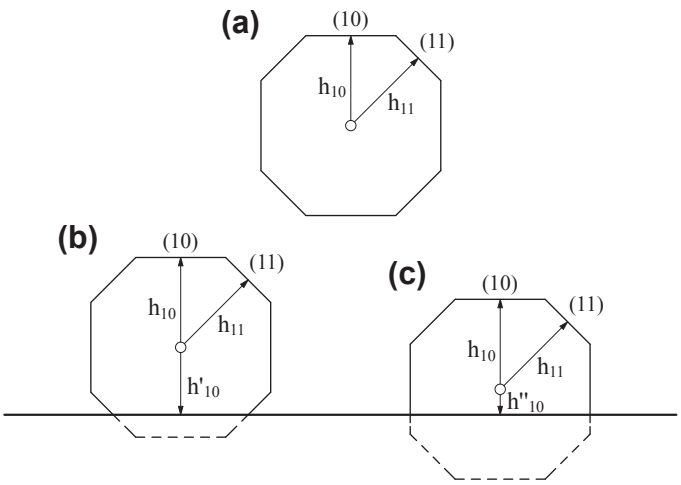


FIGURE 7.12 Three combined crystal shapes: a homogeneous nucleus (a) and heterogeneously formed nuclei (b) and (c). Wullf's points are shown by small circles, while the support is represented by the straight line. *From Ref. [54].*

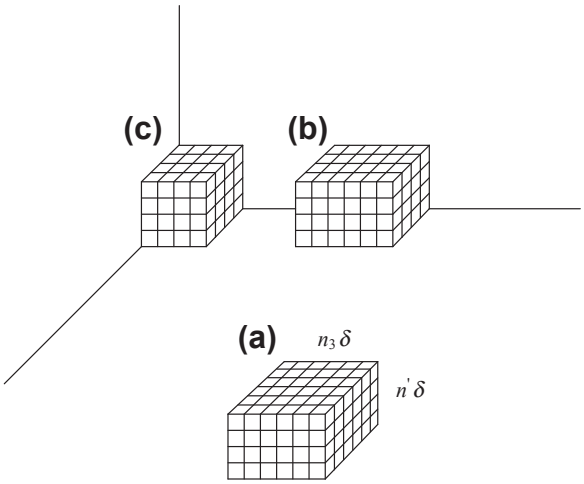


FIGURE 7.13 3D models of Kossel-crystals nucleated on (a) one substrate, (b) two equal substrates, and (c) three equal substrates.

Thus, the equilibrium crystal shape (Figure 7.13(a)) is determined by

$$n'/n_3 = 1 - \psi'/\psi. \tag{7.81}$$

To calculate the energy barrier $\Delta G^*_{\text{heter}}$, Kaischew used Eqn (7.63), which relates ψ with γ; analogically, he expressed ψ' by the specific adhesion energy, $\omega = \psi'/\delta^2$ (this expression reflects the fact that only one crystal surface is appearing by detachment of a CBB from the support). In the case under consideration,

$$\sum_{i=1}^{n} \varphi_i = 2n_3 n'(n_3 - 1)\psi + n_3^2(n' - 1)\psi + n_3^2\psi', \tag{7.82}$$

Again, Eqn (7.66)] is used to establish the energy barrier for heterogeneous crystal nucleation, $\Delta G^*_{\text{heter}}$. With Eqn (7.60), Kaischew obtained

$$\begin{aligned}\Delta G^*_{\text{heter}} &= n_3^2 n'[3\psi - (2\psi/n_3)] - \sum_{i=1}^{n} \varphi_i = n_3^2(\psi - \psi') = n_3^2\psi(1 - \psi'/\psi) = \Delta G^*_{\text{homo}}(1 - \psi'/\psi) \\ &= \Delta G^*_{\text{homo}}(1 - \omega/2\gamma).\end{aligned} \tag{7.83}$$

In a similar manner, he considered nucleation in concave edges (Figure 7.13(b)), where the crystal reclines on two support plains, and in concave vertices (Figure 7.13(c)), where the crystal sits on three support plains [54]. In doing so, Kaishew showed that in the former case

$$\Delta G^*_{\text{heter}} = \Delta G^*_{\text{homo}}(1 - \omega/2\gamma)^2, \tag{7.84}$$

In the second case,

$$\Delta G^*_{\text{heter}} = \Delta G^*_{\text{homo}}(1 - \omega/2\gamma)^3, \tag{7.85}$$

It is seen that the equilibrium crystal shapes (Figure 7.13(a)), energy barriers, and thus rates of crystal nucleation on substrates depend on the differing nucleation activity of

foreign surfaces, expressed by the corresponding specific adhesion energies, ω. Furthermore, Kaischew also calculated thermodynamically the ratio $\Delta G^*_{\text{heter}}/\Delta G^*_{\text{homo}}$ (for the same supersaturation) [70]:

$$\Delta G^*_{\text{heter}}/\Delta G^*_{\text{homo}} = V^*_{\text{heter}}/V^*_{\text{homo}}. \tag{7.86}$$

The physical explanation of this result is simple: depending on the substrate nucleation activity, the nucleus volume is decreased (see Figure 7.12), and the nucleation energy barrier is proportionally decreased.

Kaishew considered only the simplest case of nucleation on structureless (or isomorphous) substrate. Clearly, on a lattice-mismatched substrate, the deposited crystal as well as some part of the underlying substrate are strained. In such cases, the elastic strain energy arising due to the misfit with the support has to be added to the change in thermodynamic potential. In doing so, Mueller and Kern [68] concluded that, owing to the strain, the equilibrium crystal shape changes; some facets decrease while some others increase in size.

7.8.1 Saturation Density of Nuclei during Mass Crystallization in Solutions and Melts

Particularly active sites for nucleation are present on the substrates themselves as well. Assuming equal activity of the sites, the rate $\mathrm{d}\tilde{N}/\mathrm{d}t$ at which they are consumed has been calculated by Robins and Rhodin [41]. Evidently, this rate depends on the number $(\tilde{N}_{\text{a}} - \tilde{N}^*)$ of the still unoccupied active sites:

$$\mathrm{d}\tilde{N}^*/\mathrm{d}t = J'_{\text{st}}(\tilde{N}_{\text{a}} - \tilde{N}^*) \tag{7.87}$$

where J'_{st} (s^{-1}) is the nucleation frequency per active site and $\tilde{N}_{\text{a}}$ and $\tilde{N}^*$ denote the density of the active sites and nuclei numbers, respectively. Taking into account the initial condition that for $t = 0$, $\tilde{N}^* = 0$, the integration from $\tilde{N}_{\text{a}}$ to $\tilde{N}_{\text{a}} - \tilde{N}^*$ yields

$$\tilde{N}^* = \tilde{N}_{\text{a}}\left[1 - \exp\left(-J'_{\text{st}}t\right)\right] \tag{7.88}$$

As seen, at $t \to \infty$, $\tilde{N}^* \to \tilde{N}_{\text{a}} = \text{const}$. Due to the exponential dependence, the ratio $\tilde{N}^*/\tilde{N}_{\text{a}} = 0.98$ to 0.99 is attained at $t > 4/J'_{\text{st}}$ to $5/J'_{\text{st}}$, $1/J'_{\text{st}}$ being the time constant of the process.

However, the active sites on the support can possess different activity in respect to the nucleation phenomenon [63]. Even when the substratum is a single crystal, its surface is not homogeneous; there are always emergence points of (both edge and screw) dislocations, tilt and twist boundaries, foreign inclusions embedded in the crystal lattice, point defects, and surface steps, which, due to the locally increased chemical potential or/and relief change, can become the preferred nucleation sites. Evidently, any different kind of site can become active in the nucleation process, provided the system's supersaturation rises higher than some characteristic threshold. Correspondingly, these kinds of sites can be classified with respect to the said critical supersaturation. In such a case, the maximal nuclei density does not exceed the

number of corresponding kind of active sites. However, with the increase in the system's supersaturation, less active sites will be involved in the nucleation process. The limit to this stepwise activation process is set by the critical supersaturation that is sufficient for homogeneous nucleation. This issue has been quantitatively looked into by Kaischew and Mutaftschiev [71].

Similar to CNT, matter and heat transport in the mother phase are neglected by Stranski and Kaischew's theory. However, they are of importance in condensed systems. As already noted (Section 7.2.4), depending on the matter and heat transport in the parent phase, nuclei growth may substantially change the concentration and/or temperature in their immediate surroundings. The so-called excluded nucleation zones appear when supersaturation falls below the critical nucleation limit (e.g., see Ref. [10]). Then, the exhaustion of the active particles/centers and/or the overlap of the excluded nucleation zones sets the upper limit of the nuclei number.

7.9 Nucleation Theorem

Gibbs [1] was the first to notice that the thermodynamic nucleation theory proposed by him leads to a simple relationship between the critical work and critical size; the relationship between ΔG^*, n^*, and $\Delta\mu$ is given by Eqn (7.14) (see also Ref. [72]). Perhaps the famous nucleation theorem, given by Kashchiev and Oxtoby [73–76], was a result of this argument. This nucleation theorem is very valuable for experimenters because it reveals how the critical nucleus size is related to the supersaturation dependence of nucleation rate. Using experimental data for the nucleation rate J_{st}, the number of molecules n^* constituting critically sized nuclei[4] can be calculated by means of the following equation:

$$n^* \approx k_B T \mathrm{d}(\ln J_{st})/\mathrm{d}(\Delta\mu) + \alpha_1 \tag{7.89}$$

where α_1 is a small correction taking values between 0 and 1 [74].

It should be emphasized that nucleation theorem validity is not restricted by nucleus shape or size [73,74]. Moreover, nucleus size can be established without knowing the molecular mechanism details of its formation. An in-depth theoretical analysis [74] has shown that Eqn (7.89) provides a high degree of certainty; it "can be used to give n^* with an accuracy of 1–2 molecules" [73]. A statistical mechanical–kinetic derivation of the nucleation theorem, which takes into account the kinetic prefactor in the rate expression as well as the exponential term, has been proposed by Ford [77]. The nucleation theorem has also been checked thoroughly by Schmelzer [78], who expanded its application.

Kashchiev [75] gave strict phenomenological and thermodynamic proof of the nucleation theorem and provided some generalizations. Although oversimplified, a

[4]In concentrated systems, this is the excess number of molecules in the critical nucleus, which is given by the difference between the number of molecules in the nucleus and the number of molecules there would be in the same volume without the nucleus.

simple derivation of Eqn (7.89) is presented here, starting from the logarithmic form of Volmer's equation (Eqn (7.18)) for the nucleation rate:

$$\ln J = \ln A - \Delta G^*/k_B T \tag{7.90}$$

Then, replacing the ΔG^* value from Eqn (7.8), Eqn (7.90) is derived with respect to $\Delta\mu$, under the assumption that A and γ are constants. In doing so, an equivalent form of Eqn (7.89) is obtained:

$$n^* \approx k_B T \mathrm{d}(\ln J)/\mathrm{d}(\Delta\mu) - [\mathrm{d}n^*/\mathrm{d}(\ln\Delta\mu) - \gamma(\mathrm{d}S^*/\mathrm{d}(\Delta\mu))] \tag{7.91}$$

Then, with Eqn (7.10) for spherical crystals, one obtains

$$n^* \approx k_B T \mathrm{d}(\ln J)/\mathrm{d}(\Delta\mu) \tag{7.92}$$

Using Eqn (7.9) and with $\mathrm{d}S^*/\mathrm{d}(\Delta\mu) = (\mathrm{d}S^*/\mathrm{d}n^*)[\mathrm{d}n^*/\mathrm{d}(\Delta\mu)]$, Eqn (7.92) is yielded for the Kossel crystal as well.

7.10 Probabilistic Features of the Nucleation Process

Both molecule attachment and detachment are random processes, which means that any cluster smaller than the critically sized one performs a random size-walk forth and back on the time axis, until it eventually reaches the critical size [9]. Therefore, the nucleation rate itself is a random quantity as well. Thus, the formal probability laws govern the statistical distribution of the nucleation rates, whereas the kinetic nucleation theory predicts their average values only. This fact is reflected in the inherent data scatter of the measured nucleation rates.

The probabilistic features of the nucleation process have been discussed thoroughly by Toshev and co-workers [9,79]. Following these authors, only the simplest case of steady-state nucleation will be considered here because it is more lucid than the nonsteady case and renders a clear result.

Toshev [9] considered nucleation as a sequence of independent random events occurring during a fixed time interval. Using the Poisson expression, he calculated the probability P_M of finding M events within the time interval from 0 to t:

$$P_M = \left[\check{N}^M \exp(-\check{N})\right]/M! \tag{7.93}$$

where $\check{N}$ is the average number of expected nucleation events, which should appear in the chosen time interval.

Following the calculus of probability, Toshev [9] derivates Poisson's formula to obtain a relationship between the probability and nucleation rate $\mathrm{d}\check{N}/\mathrm{d}t$:

$$\mathrm{d}P_{\geq M} = (\mathrm{d}\check{N}/\mathrm{d}t)\left[(\check{N}^{M-1})/(M-1)!\right]\exp(-\check{N})\mathrm{d}t \tag{7.94}$$

Integrating Eqn (7.94), he obtained the probability of formation of minimum M nuclei before time t is elapsed. Introducing the total number of nuclei according to Eqn (7.48), he considered $\check{N}$ and $J = \mathrm{d}\check{N}/\mathrm{d}t$ as representative for the whole system and assumed that the supersaturation remains constant. To obtain an expression for appearance of at least

one nucleus, $M = 1$, Toshev assumed as a limiting case the rate equation for steady-state nucleation and reached

$$dP_{\geq 1} = J_{st} \exp(-J_{st}t)dt \tag{7.95}$$

This formula expresses the probability of occurrence of the first nucleus within the time interval between t and $t + dt$. By integrating, the probability $P_{\geq 1}$ (that at least one nucleus has arisen before the time t is elapsed) was obtained:

$$P_{\geq 1} = 1 - \exp(-J_{st}t) \tag{7.96}$$

According to the theorem for average values of a function, the mean expectation time T_1 for the appearance of at least one nucleus is

$$T_1 = \int_0^\infty t dP_{\geq 1} = 1/J_{st} \tag{7.97}$$

It is seen that the mean expectation time T_1 needed to form at least one nucleus is reciprocal to the steady-state nucleation rate. Although the time of first nucleus formation in the system is also a random quantity, Toshev and co-workers had shown that it yields valuable information concerning the kinetics of the process [79].

Some novel applications of the nucleation theory will be considered in the following sections.

7.11 Use of Burst Nucleation for Producing Equally-Sized Nanoparticles

A detailed understanding of the mechanisms responsible for formation of metallic particles with carefully tailored properties is indispensable to many contemporary technology areas, such as photovoltaics, catalysis, electronics, and medicine. In this respect, burst nucleation of metal colloids and nanoparticles in solutions, followed by diffusional growth and aggregation, or their combinations, represents an interesting development [80]. Burst nucleation is initiated by chemically generating (or introducing) monomers. By burst nucleation and further growth, molecule clusters reach sizes up to a couple of tens of nanometers. In many cases, nanocrystals of ZnS, CdS, Fe_2O_3, Au, Ag (and other metals) can begin to aggregate, becoming "monomers" for the formation of (crystalline) colloids. A modification of CNT has been used to explain the burst nucleation [80].

7.12 Nucleation of Protein Crystals

Protein crystal nucleation is a special case of spontaneous highly precise self-assembly of biological macromolecules into stable clusters, formed as a result of selective and appropriately directed interactions. It is a mandatory requirement that biological macromolecules have to be arranged in a geometry that is appropriate for creating

crystallographically symmetrical molecular structures. On the basis of reliable statistic data, Wukovitz and Yeates [81] revealed a tendency for proteins to crystallize in a small number of preferred space groups, where it is the easiest to achieve connectivity. The specificity of protein crystal nucleation will be considered in the following section.

7.12.1 Bond Selection Mechanism

It is known that only the structure of the protein molecule surface dictates a molecule's ability to bind to partners during protein crystallization. Evidence comes from the entirely analogous crystallization behavior of apo- and holoferritin, observed to occur due to the same molecule surface structure, regardless of the dramatically different molecule core. Recall that apoferritin is an empty shell, while a mineral core is present in the holoferritin. Nevertheless, when forming under the same conditions, the crystals of both proteins have exactly the same shape; the crystals differ only in their color: apoferritin crystals are yellowish, whereas holoferritin is reddish-brown [82]. Taking that into account and the experimental observations that protein crystal nucleation is rather slow, a bond selection mechanism (BSM) has been suggested [83,84]. It represents an attempt to describe the most important features of the extremely complex molecular-kinetic mechanism of protein crystal nucleation.

As is well known, there is a fundamental difference between small inorganic and large protein molecules. Small molecules possess spherical interaction fields with constant interaction potential. In supersaturated media, every hit between them, independently of molecules' spatial orientation, has the potential to contribute to a crystal bond formation. In contrast, the surface of the protein molecule is highly patchy and heterogeneous. The rationale behind BSM rests on the concept that patch–patch recognition is mandatory for the formation of lattice contacts in protein crystals.

Although it is impossible to observe the elementary acts of protein crystal bond formation, knowledge about lattice contacts (e.g., see Protein Data Bank, (PDB) data in Refs [85–88]), resulting in protein crystallization, lay a sound basis to suggest a mechanism of protein crystal bonding. It has been evidenced by PDB-structure statistics [85] that protein lattice contacts are not random; they occur through strict selection of amino-acid residues situated on the protein molecule surface. Only two (arginine and glutamine) predominate in crystal lattice contacts, of the 20 proteinogenic amino-acid residues that the vast multitude of living organisms rely on. In contrast, the least likely residues to be found in the crystal lattice contacts are lysine and glutamate residues [85].

Such a bond selection imposes severe steric restriction to protein molecule association, leading to crystal nucleus formation. In addition, it should be emphasized that only amino-acid residues situated in proper positions on the molecule surface are able to participate in producing crystal lattices. Even though potentially active, some residues that are out of crystallographically symmetrical positions (if, for instance, they are situated too close) can remain unused. Therefore, over their surface, protein molecules exhibit a highly limited number of discrete patches, which are the authentic bonding

sites under the actual crystallization conditions. Because such a patch occupies only a small fraction of the total surface of the protein molecule [82], BSM dramatically decreases the chance for a crystalline lattice bond formation and postpones the nucleation process significantly.

Indeed, the steric restriction effect is somewhat mitigated by rotational diffusion, which involves multiple collisions. The latter increases the chance for fine-tuning of the proper spatial positioning of crystallization patches on the two meeting protein molecules [89]. However, the rotational diffusion is highly effective only for protein pair formation. Its impact decreases strongly with larger complexes, following the order of pair $\gg$ trimer $\gg \ldots n$-mer $\ldots \rightarrow 0$. The reason is that random rotation slows down very fast with larger complexes; for spheres, it is inversely proportional to their volume [89,90]. Therefore, it is highly improbable that the rotational diffusion may effectively assist a reasonably frequent formation of critical clusters constituted of large number of protein molecules (see also Section 7.12.4).

The reasons for BSM and its impact on protein crystal nucleation are considered in full detail in a review paper [91].

7.12.2 Shape of the Protein Crystal Nucleus

Crystal nucleation kinetics is studied intensively with globular proteins, but the experimental determination of the shape of the crystal nucleus still remains a challenge. The evident reason is the principal impossibility to see the critical nuclei directly. As already mentioned (see Section 7.4.1), the shape of the critical crystal nucleus has been deduced from the observed nearly critical apoferritin crystallites [34]. Surprisingly, the crystallites had a raft-like form. This observation is rather puzzling [92,93] because the apoferritin molecule is quite symmetrical in shape, almost spherical, and crystallizes in *fcc* lattice. Using the classical approach of the MWS method, the observation of Yau and Vekilov has been explained on the basis of the hypothesis for different crystal lattice bond strengths [57].

A Kossel-like crystal nucleus model was considered for globular proteins [83,84], in which spheres replaced the cubic building blocks of the said crystal (Figure 7.14). Keeping in mind the nature of the lattice binding forces between the huge biomolecules,

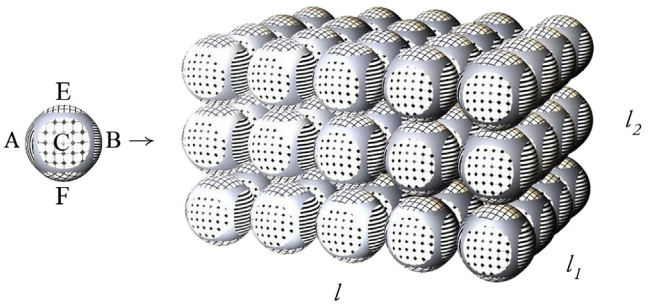


FIGURE 7.14 Cubic primitive lattice model of a 3D crystal, formed due to six sticky patches (not to scale): A, B, C, E, F, and the patch just behind C that is not seen. l, l_1 and l_2 are the numbers of molecules in the corresponding rows.

only the first-nearest neighbors were taken into account. A diametrical opposition of the sticky patches was assumed because it simplified substantially the quantitative consideration of the nucleus shape (Figure 7.14).

The crystal model shown in Figure 7.14 is constructed from three rows of molecules. The bond strength of the most stable contact between two molecules, the one in the horizontal front row in Figure 7.14, is denoted as ψ_1, $\psi_1 = \max$. Then, a second type of weaker bond of strength ψ_2 is added horizontally at the back. And finally, a third type of bond ψ_3 is necessary to construct the three-dimensional cluster; those are the weakest bonds. Using this model, it is easy to calculate the corresponding MWS, φ, for the three types of faces:

$$\varphi_1 = \psi_1 + \psi_2 + \psi_3 - \psi_1/l - \psi_2/l_1 \quad - \text{ for the top (or bottom) face in Figure 5.14.} \tag{7.98}$$

$$\varphi_2 = \psi_1 + \psi_2 + \psi_3 - \psi_1/l - \psi_3/l_2 \quad - \text{ for the front (or back) face in Figure 5.14.} \tag{7.99}$$

$$\varphi_3 = \psi_1 + \psi_2 + \psi_3 - \psi_2/l_1 - \psi_3/l_2 \quad - \text{ for the side – faces in Figure 5.14.} \tag{7.100}$$

At equilibrium, all three MWSs have to be equal. Hence:

$$l/l_1 = \psi_1/\psi_2,\ l_1/l_2 = \psi_2/\psi_3 \text{ and } l/l_2 = \psi_1/\psi_3 \tag{7.101}$$

Because $\psi_1 > \psi_2 > \psi_3$, the three-dimensional crystal is a parallelepiped, but not cube.

It is worth showing here that this result obeys the Gibbs–Curie–Wulff law. According to Eqn (7.101) we have: $l/l_1 > 1$, $l/l_2 > 1$ and $l_1/l_2 > 1$, and hence:

$$ll_2 > l_1 l_2,\ ll_1 > l_1 l_2, \text{ and } ll_1 > ll_2 \tag{7.102}$$

Thus, the smallest are the left- and right-hand side faces in Figure 7.14, which are situated normally to the strongest connecting force; the largest (top and bottom) faces are normal to the weakest bonding force. Keeping in mind Eqn (7.63) for the specific surface free energy γ_i, one sees that the equilibrium crystal face with the largest specific surface free energy $\gamma_1 \approx \psi_1/2s$ is the smallest in size, and the one possessing the lowest energy $\gamma_3 \approx \psi_3/2s$ (having the weakest dangling bonds) is the largest in size. Thus, the model obeys the Gibbs–Curie–Wulff law.

Closely packed 3D-crystal nuclei models built of spherical protein molecules, including those similar to the raft-like apoferritin crystal shapes of Yau and Vekilov, have been considered as well [57]. It should be noted that the formation of nonequilibrium shaped clusters is feasible but its energy cost is higher [94].

7.12.3 Energy Barrier for Protein Crystal Nucleation

Applying the MWS method to calculate the energy barrier for homogeneous protein crystal nucleation [84], using Eqn (7.66), gives

$$\Delta G^*_{\text{homo}} = ll_1 l_2 \varphi - \sum_1^{n^*} \varphi_i \tag{7.103}$$

The total binding energy of the crystal, built of $n^* = ll_1 l_2$ individual molecules, is

$$\sum_1^{n^*} \varphi_i = (l-1) l_1 l_2 \psi_1 + (l_1 - 1) l l_2 \psi_2 + (l_2 - 1) l l_1 \psi_3, \tag{7.104}$$

and thus,

$$\Delta G^*_{\text{homo}} = ll_1\psi_3 \left(= ll_2\psi_2 = l_1l_2\psi_1\right) \tag{7.105}$$

From Eqn (7.62), l, l_1, and l_2 are expressed by ψ_1, ψ_2, and ψ_3 and the supersaturation $\Delta\mu$: $l = 2\psi_1/\Delta\mu$, $l_1 = 2\psi_2/\Delta\mu$ and $l_2 = 2\psi_3/\Delta\mu$. Then, an analog to Eqn (7.13) is obtained expressing ψ_1, ψ_2, and ψ_3 by the corresponding surface free energies γ_1, γ_2, γ_3, according to Eqn (7.63):

$$\Delta G^*_{\text{homo}} = 32\gamma_1\gamma_2\gamma_3\Omega^2/(\Delta\mu)^2 \tag{7.106}$$

The numerical estimation, performed by means of Eqn (7.106) and with $\gamma \approx 1$ erg/cm^2 [95], $\Omega \approx 3 \times 10^{-20}$ cm^3, and $\Delta\mu \approx 3k_BT$, yields $\Delta G^*_{\text{homo}} = 2 \times 10^{-12}$erg. In view of Eqn (7.83), and taking into account the experimental data obtained for heterogeneous insulin crystal nucleation, $\Delta G^*_{\text{heter}} = (3.8 - 6.8) \times 10^{-13}$erg [84], the result is encouraging.

7.12.4 Protein Cluster–Cluster Aggregation on Diffusional Encounter

Despite the unusually high supersaturations applied in protein crystal nucleation studies, the nucleation kinetics is rather slow [84,31]. One possible explanation of this observation may be that successive series of multiple coalescences of critical and/or near-critical clusters could reduce the observed crystal yield. Because there is no technique to track the destiny of the individual undercritically, critically, and supercritically sized clusters, it is particularly difficult to give any confident statement regarding the intermediate processes that may occur before and after the onset of crystal nucleation. In addition to the complexity of the issue is the fact that, while coalescence of clusters equal or larger than the critical size would diminish the crystal nucleation rate, coalescence of smaller sized clusters would augment it.

Firstly, due to the patchy and highly inhomogeneous surface of the protein molecules constituting the clusters, it is reasonable to expect that BSM should affect protein molecule clusters as well. In other words, when such clusters meet, their proper spatial orientation provides an inevitable coalesce condition. Because crystalline protein clusters are polyhedrons in shape [34], the simplest model is coalescence of two clusters with a Kossel-like lattice that consists of eight molecules each (i.e., two dice; Figure 7.15). Evidently, the facets on the two dice that correspond to the strongest intermolecular binding are most prone to bind together [31].

The probability for a proper meeting of two such dice in space is equal to the probability for the simultaneous appearance of just the same number on both dice by crap-shooting—that is, ~0.028. This probability is about 3 times smaller than that one for formation of protein molecule dimmers, which has been estimated to be 0.09–0.13 [84].

Note that the dice-meeting probability also includes partial coincidence, when merely one or two molecules on each dice become associated. Such a construction would hardly have sufficient stability; complete stability could only be achieved by

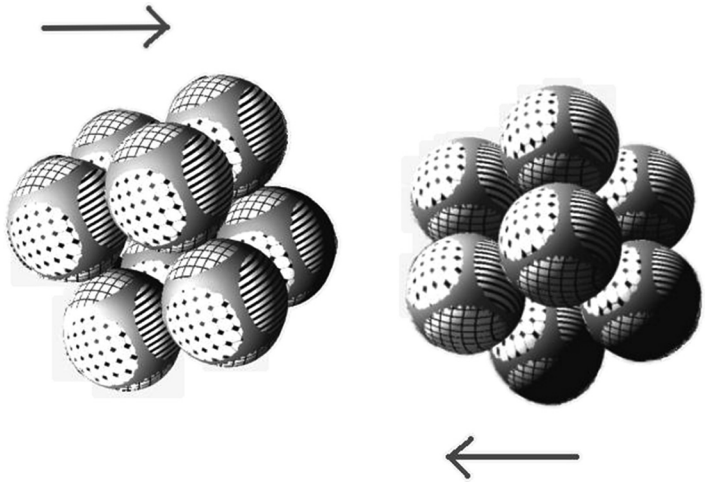


FIGURE 7.15 Accidental meeting in space of two clusters (two dice) that consist of eight protein molecules each.

bonding of all eight molecules on both preferred dice faces. Thus, an additional spatial adjustment is required, which may be provided by rotational diffusion. However, because of the Brownian motion, the dice also may be separated in the meantime, which additionally decreases the coalescence probability.

Coalescence of clusters larger than eight molecules has also been discussed [31]. Obviously, the attachment of a ninth (and also tenth and eleventh) protein molecule to the dice in Figure 7.15 will create bulges. Bulges appear also on clusters consisting of 13–15 protein molecules, etc. However, flat faces bind together much stronger than bulges do, thus reducing coalescence probability even further. As already mentioned (see Section 7.12.2), in complete parallelepiped clusters, the side faces normal to the strongest connecting forces are of the smallest area (this holds true also for completed clusters consisting of 12, 60, etc. molecules), which decreases their coalescence probability.

Only coalescence of equally sized protein clusters has been considered so far. Indeed, differently sized clusters coalesce as well, but their coalescence does not change the general trend [31]. The reason is that smaller clusters rotate diffusionally and accommodate faster than larger particles. This is especially pronounced with clusters of highly differing sizes. In such a case, only the smaller cluster rolls about the larger one, like a ball on a floor. The larger cluster is almost immovable.

A general conclusion can be drawn that, due to the spatial adjustment via rotational diffusion, coalescence of critical and/or supercritically sized crystalline clusters of protein molecules is less probable than that for clusters smaller than the critical nucleus size. Moreover, due to the decreasing effect of the rotational diffusion, any successive (multiple) cluster coalescence resulting in a large-scale conglomeration is improbable.

The protein cluster coalescence problem has been given some quantitative considerations as well [31]. Using Smoluchowski's approach to rapid coagulation of spherical colloids, the dynamic process of the single coalescence of clusters composed of n protein molecules, which are consumed to produce larger size of clusters, is described by the rate equation for second-order bimolecular reactions written for every category of equally sized clusters:

$$dZ_n/dt = -k_n Z_n^2, \tag{7.107}$$

where k_n is the coalescence rate constant and Z_n is the concentration of clusters in the category n. Our implicit assumption is based on the predominating process of monomer attachment/detachment to/from the clusters; thus, no cluster size is missing. In other words, a quasi-equilibrium cluster distribution (Szillard chain) is established sufficiently fast.

Coalescence balance is estimated by integrating Eqn (7.107) from Z_n^o at $t = 0$ to Z_n:

$$Z_n^o/Z_n = 1 + Z_n^o k_n t = R_n, \tag{7.108}$$

Note that $Z_n^o k_n t$ is dimensionless, likely R_n; $Z_n^o k_n$ has the meaning of reciprocal half-life time, and larger R_n means a faster coagulation for the particles in the category n.

It has been suggested [31] that the cluster coalescence rate constant k_n represents the coalescence probability. Whether the change in k_n is monotonous or not is rather uncertain; that is why the following logical scheme has been applied: denoting $j = n^*/2$ for even n^* numbers and $j = (n^* + 1)/2$ for odd n^* numbers, it follows from the qualitative consideration presented above that the coalescence probability P_j for clusters in the category j, is much larger than the coalescence probability P_{n^*} for the clusters in the category n^*. Respectively, $P_{j+1} > P_{n^*+1}$, $P_{j+2} > P_{n^*+2}$, … , $P_{n^*-1} > P_{n^*+j-1}$. Thus, the rate constants $k_{n<n^*}$ have to be systematically larger than the rate constants $k_{n\geq n^*}$. In addition, it is logical to assume that the initial concentrations of the clusters $Z^o_{n<n^*}$, for every cluster category $n = 2, 3, \ldots, n^* - 1$, is at least equal to $Z^o_{n\geq n^*}$; for instance, the critical nuclei are of minimum concentration (Section 7.4). Finally, summing up the system of equations of the type of Eqn (7.108) for the different cluster categories and for the same time t, and knowing that $k_{n<n^*} Z^o_{n<n^*} > k_{n\geq n^*} Z^o_{n\geq n^*}$, we see that the coalescence probability for clusters in the range from j to $n^* - 1$, which may produce critical and supercritical nuclei, should be larger as compared with the coalescence of clusters of critical and supercritical size. Taking into account the rotational diffusion effect as well, the general conclusion is that being too slow, cluster coalescence does not play any role in the process of bulk protein crystal nucleation.

7.13 Concluding Remarks

The study of nucleation is the subject of a huge number of papers. The literature also comprises more than 70 books and reviews. (For an extended list of books and reviews, see the preface of Kashchiev's book [75].) A number of excellent books (e.g., [2]) and

papers have been published subsequently (e.g., [46,80,96–99]), and the scientific and technological interest in the topic is not declining. Due to the large body of literature on the subject, this chapter is not intended to be exhaustive. I outlined some of the basic physics that is believed to underlie this phenomenon. With the intention of making the topic more comprehensible to a broader readership, heavy mathematics was omitted. In addition, the limited space permits consideration of nucleation in a one-component system only. (Readers interested in nucleation in multicomponent systems may refer to Ref. [2].) In view of the author's background and preferences, the chapter focuses on some cases that seem underestimated in contemporary literature, such as the almost forgotten but fruitful MWS method.

7.14 Perspectives

In a constant quest to improve the nucleation theory, many corrections have been suggested, some of which were mentioned in Sections 7.4 and 7.4.1. Except for the most basic notions of CNT, such as the necessity of fluctuative nucleus formation under a sufficiently high driving force, almost all remaining postulates have been questioned in recent decades. Nevertheless, CNT remains the essence of nucleation theory. Although it is particularly difficult to give any confident statement as to how the theory will develop in the coming years, it is possible to predict that improvement will most likely be centered on explanations of specific cases, such as nanoparticle and protein crystal nucleation.

Frequently Used Abbreviations

BSM	bond selection mechanism
CBB	crystal building block
CNT	classical nucleation theory
MWS	mean work of separation

References

[1] Gibbs JW. Trans Connect Acad 1876;3:108. 16(1878) 343 (Russian translation, Nauka, Moscow, 1982, 61).

[2] Kelton KF, Greer AL. Nucleation in condensed matter, applications in materials and biology. Elsevier; 2010.

[3] Volmer M, Weber A. Z phys Chem 1926;119:277.

[4] Kaischew R. Ann Sofia Univ Fiz Math (Chem) 1946/1947;63:53.

[5] Volmer M. Kinetik der Phasenbildung. Dresden-Leipzig: Theodor Steinkopff Verlag; 1939.

[6] Farkas L. Z Phys Chem 1927;125:236.

[7] Kaischew R, Stranski I. Z Phys Chem 1934;B26:317.

[8] Becker R, Döring W. Ann Phys 1935;24:719.

[9] Toshev S. Homogeneous nucleation. In: Hartman P, editor. Crystal growth: an introduction. North-Holland Publ. Co.; 1973.

[10] Markov I. Crystal growth for beginners, fundamentals of nucleation, crystal growth and epitaxy. 2nd ed. New Jersey, London, Singapore, Hong Kong: World Scientific; 2003.

[11] Zeldovich JB. Acta Physicochim URSS 1943;18:1.

[12] Kashchiev D. Surf Sci 1969;14:209.

[13] Wölk J, Strey R, Heath CH, Wuslouzil BE. J Chem Phys 2002;117:4954.

[14] Frenkel JI. Kinetik theory of liquids. Oxford: Oxford University Press; 1946.

[15] Turnbull D, Fisher JC. J Chem Phys 1949;17:71.

[16] Penkova A, Chayen N, Saridakis E, Nanev CN. Acta Crystallogr D 2002;D58:1606.

[17] Penkova A, Dimitrov I, Nanev Chr N. Ann N Y Acad Sci 2004;1027:56.

[18] Jones SF, Evans GM, Galvin KP. Adv Colloid Interface Sci 1999;80:27.

[19] Mutaftschiev B. Nucleation theory. In: Hurle DTJ, editor. Handbook of crystal growth, vol. 1. Amsterdam: Elsevier; 1993. p. 187–247 [Chapter 4].

[20] Hirth JP, Pound GM. Condensation and evaporation; nucleation and growth kinetics. In: Chalmers B, editor. Progress in materials science, vol. 11. Oxford: Pergamon Press; 1963.

[21] Mutaftschiev B, Zell J. Surf Sci 1968;12:317.

[22] Chernov SF, Fedorov YV, Zakharov VN. J Phys Chem Solid 1993;54:963.

[23] Tolman RC. J Chem Phys 1949;17:333–7.

[24] Ono S, Kondo S. In: Flügge S, editor. Handbuch der Physik, vol. 10. Berlin-Göttingen-Heidelberg: Springer Verlag; 1960.

[25] Xiao W, Ru-Zeng Z. Chin Phys B 2013;22:1–3. 036801.

[26] Nellas RB, Keasler SJ, Siepmann JI, Chen B. J Chem Phys 2010;132:1–9. 164517.

[27] Lothe J, Pound GM. J Chem Phys 1962;36:2080.

[28] Abraham FF. J Atmos Sci 1968;25:47.

[29] Mutaftschiev B. The atomistic nature of crystal growth. In: Springer series in materials science, vol. 43. Berlin, Heidelberg, New York: Springer-Verlag; 2001.

[30] Bonissent A, Mutaftschiev B. J Phys Chem 1973;56:3727.

[31] Nanev CN. Cryst Res Technol 2009;44:7.

[32] ten Wolde PR, Frenkel D. Phys Chem Chem Phys 1999;1:2191.

[33] Liu X-Y. J Cryst Growth 2002;237–239:106.

[34] Yau S-T, Vekilov PG. Nature 2000;406:494.

[35] Merikanto J, Zapadinsky E, Laury A, Vehkamäki H. Phys Rev Lett 2007;98:1–4. 145702.

[36] McGraw R, Laaksonen A. J Chem Phys 1997;106:5284.

[37] McGraw R, Laaksonen A. Phys Rev Lett 1996;76:2754.

[38] Walton D. J Chem Phys 1962;37:2182.

[39] Walton D. Phil Mag 1962;7:1671.

[40] Walton D, Rhodin TN, Rollins R. J Chem Phys 1963;38:2695.

[41] Robins JL, Rhodin TN. Surf Sci 1964;2:346.

[42] Milchev A. Contemp Phys 1991;32:321.

[43] Stoyanov S. Current topics in mat. sci.. Amsterdam: North-Holland; 1974.

[44] Milchev A, Malinovski J. Surf Sci 1985;156:36.

[45] ten Wolde PR, Frenkel D. Science 1997;277:1975.

[46] Vekilov PG. Cryst Growth Des 2010;10:5007–19 [and the references therein].

[47] Stranski I, Kaischew R. Z Phys Chem A 1934;170:295.

[48] Stranski I, Kaischew R. Z Phys Chem B 1934;26:100.

[49] Stranski I, Kaischew R. Z Phys Chem B 1934;26:114.

[50] [a] Kossel W. Göttinger Nachrichten 1927:135;
[b] Debye P, editor. Leipziger Voträge, p.1; Probleme der modernen Physik, Leipzig 1928.

[51] Stranski IN. Z Phys Chem 1928;136:259.

[52] Kaischew R. Z Phys 1936;102:684 [and the reference of O. Stern therein].

[53] Stranski IN. Ann Sofia Univ 1936/37;30:367.

[54] Kaischew R. Commn Bulg Acad Sci Phys Ser Bulg 1950;1:100.

[55] Madelung E. Phys Zs 1918;19:524.

[56] Tassev VL, Bliss DF. J Cryst Growth 2008;310:4209.

[57] Nanev CN. J Cryst Growth 2012;361:171.

[58] Stranski IN, Kaischew R. Ann Phys 1935;23:330.

[59] Honnigmann B. Gleichgewichts- und Wachstumsformen von Kristallen. Darmstadt: Theodor Steinkopff Verlag; 1958.

[60] Stranski IN, Kaischew R. Ann Sofia Univ, Fiz-Mat Faculty, Chem (Bulgarian) 1935/36;32:171.

[61] Burton W, Cabrera N, Frank FC. Phil Trans 1951;A243:299.

[62] Frank FC. Disc Farday Soc 1949;5:48.

[63] Kaischew R, Budevski E. Contemp Phys 1967;8:489.

[64] Bostanov V, Naneva R. J Electroanal Chem Interfacial Electroch 1986;208:153.

[65] Sears G. J Chem Phys 1956;24:868.

[66] Cabrara N. J Chem Phys 1956;53:675.

[67] Fletcher NH. J Chem Phys 1958;29:572–6.

[68] Mueller P, Kern R. Surf Sci 2000;457:229.

[69] Liu X-Y. In: Sato K, Nakajima K, Furukawa Y, editors. Advances in crystal growth research. Amsterdam: Elsevier Science; 2001. p. 42–61.

[70] Kaischew R. Commn Bulg Acad Sci Phys Ser Bulg 1951;2:191.

[71] Kaischew R, Mutaftschiev B. El Acta 1965;10:643.

[72] Nielsen AE. Kinetics of precipitation. International series of monographs on analytical chemistry, vol. 18. New York: Macmillan; 1964.

[73] Kashchiev D. J Chem Phys 1982;76:5098.

[74] Oxtoby DW, Kashchiev D. J Chem Phys 1994;100:7665.

[75] Kashchiev D. Nucleation, basic theory with application. Oxford: Butterworth-Heinemann; 2003.

[76] Kashchiev D. J Chem Phys 2006;125:014502.

[77] Ford IJ. Phys Rev E 1997;56:5615.

[78] Schmelzer J. J Colloid Interface Sci 2001;242:354.

[79] Toshev S, Milchev A, Stoyanov S. J Cryst Growth 1972;13/14:123.
[80] Sevonkaev I, Privman V, Goia D. J Solid State Electrochem 2013;17:279 [and references therein].
[81] Wukovitz SW, Yeates TO. Nat Struct Biol 1995;2:1062.
[82] Nanev CN. Cryst Res Technol 2008;43:914.
[83] Nanev CN. Cryst Res Technol 2007;42:4.
[84] Nanev CN. Cryst Growth Des 2007;7:1533.
[85] Dasgupta S, Iyer GH, Bryant SH, Lawrence CE, Bell JA. Proteins Struct Funct Genet 1997;28:494.
[86] Dale GE, Oefner C, D'Arcy AJ. J Struct Biol 2003;142:88.
[87] Bahadur RP, Chakrabarti P, Rodier F, Janin J. J Mol Biol 2004;336:943.
[88] Janin J, Rodier F, Chakrabarti P, Bahadur RP. Acta Crystallogr D Biol Crystallogr 2007;63:1.
[89] Northrup SH, Erickson HP. Proc Natl Acad Sci USA 1992;89:3338 [and references therein].
[90] Shoup D, Lipary G, Szabo A. Biophys J 1981;36:697.
[91] Nanev CN. Prog Cryst Growth Character Mater 2013;59:133.
[92] Oxtoby DW. Nature 2000;406:464–5.
[93] Gorman J. Sci News Week 2000;158:84.
[94] Kaischew R. Growth Cryst Russ 1965;5:45.
[95] Chernov AA. J Struct Biol 2003;142:3.
[96] Revalor E, Hammadi Z, Astier J-P, Grossier R, Garcia E, Hoff C, et al. J Cryst Growth 2010;312:939.
[97] Garcia-Ruiz JM. J Struct Biol 2003;142:22.
[98] Sear RP. J Phys Condens. Matter 2007;19:033101.
[99] Sleutel M, Van Driessche AES. PNAS 2014;111:E546 [and the references therein].

8 Growth Kinetics: Basics of Crystal Growth Mechanisms

Makio Uwaha

DEPARTMENT OF PHYSICS, NAGOYA UNIVERSITY, CHIKUSA-KU, NAGOYA, JAPAN

CHAPTER OUTLINE

8.1 Introduction: Purpose of This Chapter 360
8.2 Crystal Growth as a Process of Phase Transition 361
8.2.1 Equilibrium and Transition to Solid Phase 361
8.2.2 Linear Kinetics 363
8.2.3 Transport and Chemical Potential 363
8.3 Growth from Various Mother Phases 364
8.3.1 Growth from the Vapor 364
8.3.2 Growth from the Melt 365
8.3.2.1 Sharp Interface Picture 366
8.3.2.2 Diffuse Interface Picture 368
8.3.3 Growth from a Solution 369
8.4 Normal Growth and Lateral Growth 370
8.4.1 Adhesive Growth on Rough Faces 370
8.4.2 Growth via Two-Dimensional Nucleation on a Facet 371
8.4.2.1 Single-Nucleation versus Multinucleation 373
8.4.2.2 Growth of a 2D Circular Island 374
8.4.2.3 Change of Coverage by Nucleation and Spread 375
8.4.2.4 Velocity of Multinucleation Growth 376
8.4.3 Spiral Growth with Screw Dislocations 377
8.4.3.1 Velocity of Spiral Growth 377
8.4.3.2 Growth of a Facet with Many Dislocations 378
8.4.3.3 Change of Growth Mode and Kinetic Roughening 379
8.4.3.4 Formation of Mounds 380
8.5 Growth of Vicinal Faces 380
8.5.1 A Model of Solution Growth: Bulk Diffusion 381

Handbook of Crystal Growth. http://dx.doi.org/10.1016/B978-0-444-56369-9.00008-3

8.5.2 Vapor Growth and Surface Diffusion 382
8.5.2.1 The BCF Model 383
8.5.2.2 Asymmetric Kinetics at the Step 384
8.5.2.3 A Simple Example of Growth Kinetics 385
8.5.3 Growth of Vicinal Faces and Morphology of the Stepped Surface 387
8.5.3.1 Wandering of Steps 388
8.5.3.2 Bunching of Steps 390
8.6 Crystal Growth in a Diffusion Field 391
8.6.1 Melt Growth and Solution Growth 391
8.6.1.1 Evacuation of Latent Heat in Melt Growth 391
8.6.1.2 Supply of Matter in Solution Growth 394
8.6.2 Growth of a Planar Interface in a Diffusion Field 394
8.7 Various Simulation Models of Crystal Growth 395
8.7.1 Atomistic Model and Simulation 395
8.7.2 Continuum Models 396
References 397

8.1 Introduction: Purpose of This Chapter

We know that the ground state of many identical particles (e.g., atoms and molecules) is a state of a crystal in which particles form a periodic array. Figure 8.1 shows an example: a tiny 1-μm crystal of Pb shows a perfect array of atoms [1]. If a Pb atom were as large as

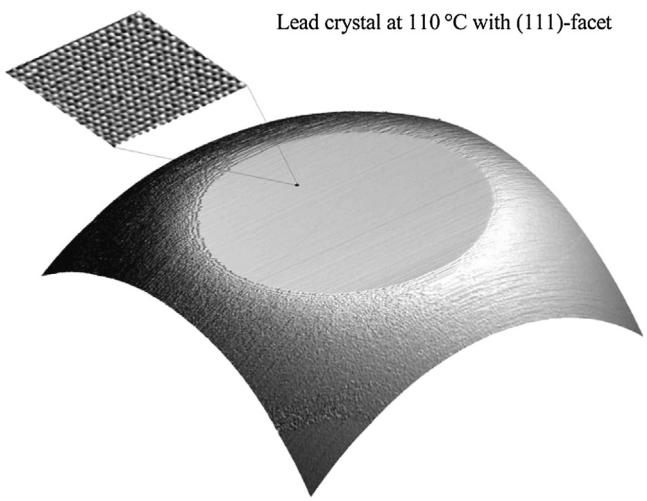


FIGURE 8.1 A 750 nm × 750 nm scanning tunneling microscope image of a Pb crystal acquired at a temperature of 383 K. The inset shows the atomically resolved (111) surface zooming into the top facet. *Reprinted from Ref. [1], with permission from AIP.*

a golf ball, the size of the facet would be several tens of meters. It is not evident at all, however, that the particles can really form such a structure spontaneously. Merely by reducing temperature or increasing pressure, atoms or molecules in a liquid/gas aggregate and organize themselves to form a crystal. The purpose of this chapter is to explain the basic mechanism of this self-organization of atoms for beginners who have some basic knowledge of physics.[1]

The present chapter is not intended to give a comprehensive view of growth mechanisms but rather to explain the simple typical pathways of crystal growth on such a level that beginners can understand the physical mechanism. As a result, topics are limited and a comprehensive biography is not supplied.

8.2 Crystal Growth as a Process of Phase Transition

Crystallization is the process of the phase transition from a liquid or a gas to a solid. Crystal growth is a synonym of crystallization, and the word suggests that the crystal is becoming larger and larger. The birth of a crystal is called *nucleation,* which is explained in Chapter 7 of this book. This chapter discusses how small crystals grow.

8.2.1 Equilibrium and Transition to Solid Phase

The mother phase is the uniform isotropic phase and the solid phase has a periodic atomic structure and therefore is anisotropic (no spherical symmetry). Because of the difference in symmetry, the two phases can be distinguished clearly. In Figure 8.2, a

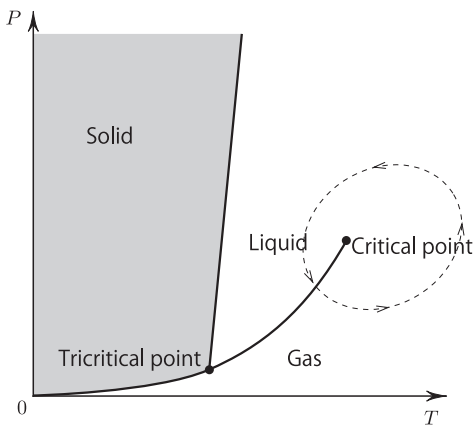


FIGURE 8.2 A typical form of the phase diagram of simple material. Continuous transition from the gas to the liquid is realized along the dashed curve.

[1]There are already many textbooks available in bookstores. Some of the books the author noted and consulted to write the present chapter are Refs [3–9].

typical phase diagram is shown. In contrast to the solid–liquid (or gas) transition, a continuous transition between the gas and the liquid is possible, as indicated by the dashed curve in Figure 8.2, because there is not any change in symmetry. Liquid and gas can be distinguished clearly only when they exists at the same time with an interface.

If one changes the temperature T and/or pressure P of the system, the state of matter may change so as to lower the Gibbs free energy $G(T, P)$. The most important quantity is the chemical potential $\mu(T, P)$, which is the Gibbs free energy per particle[2]: $\mu = G/N$. Figure 8.3 shows chemical potentials of solid and liquid (or gas). Figure 8.3(A) corresponds to the change of μ along a horizontal cut in Figure 8.2, and Figure 8.3(B) is along a vertical cut. If one decreases the temperature or increases the pressure beyond the coexisting point, the chemical potential μ_L of the liquid becomes higher than that of the solid, μ_S. Their difference, $\Delta\mu = \mu_L - \mu_S$ is the *driving force of solidification*. Note that we have assumed the entropy in solid s_S, which is the down slope in Figure 8.3(A), is smaller than that in liquid s_L. Otherwise, the solid phase appears at the high-temperature side.[3] Also, we have assumed the atomic volume in solid v_S, which is the slope in Figure 8.3(B), is smaller than that in liquid v_S. A well-known counterexample is water (H_2O): ordinary ice melts when pressure is applied. For simplicity, we always assume the standard behavior, as in Figure 8.3. Similarly, if the mother phase is a gas, $\Delta\mu = \mu_G - \mu_S$ is the corresponding driving force.

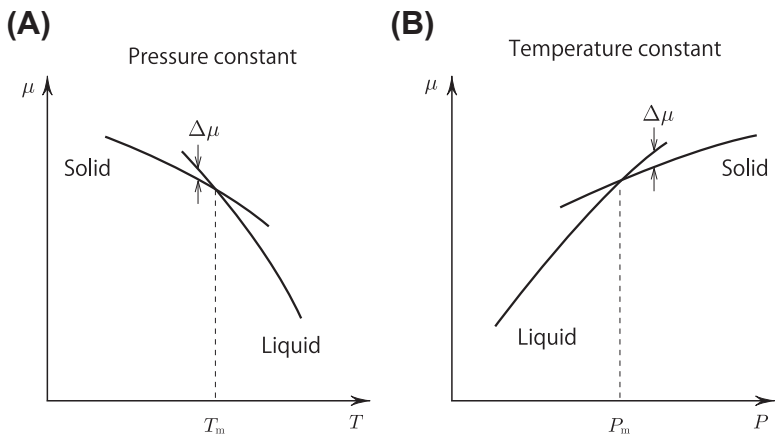


FIGURE 8.3 Change of chemical potentials of the solid and the liquid (gas) as a function of temperature T (A) and pressure P (B).

[2]Unless explicitly mentioned, we always deal with the simplest case—that is, one-component simple atoms/molecules.

[3]This case is rare but is seen in He [6].

8.2.2 Linear Kinetics

If a solid and a liquid are in equilibrium, temperature T, pressure P, and chemical potential μ are equal in both phases. When these quantities are shifted from equilibrium, the pressure balance is related to mechanical equilibrium and restored at once. An imbalance of temperature produces heat flow, and an imbalance of chemical potential produces mass flow. When the shift is not too large, the corresponding response is proportional to the amount of shift. Such a linear response relationship usually holds true. In crystal growth, however, nonlinear behavior sometimes controls the system.

8.2.3 Transport and Chemical Potential

The process of crystallization is like constructing a brick building. Bricks that make up the building are transported from other places by various ways to the construction site. There, bricks are piled up one by one, and the building becomes larger and larger. Waste from construction is in turn transported by trucks from the site and discarded somewhere in the environment. Crystallization processes are categorized into two types: transport of materials and waste (transport processes) from/to remote places and assembling at the construction site (kinetics at the interface). In both processes, free energy is consumed. In crystallization at given temperature and pressure, chemical potential is consumed in such processes.

If convection is neglected, transport is controlled by *diffusion*, and the local mass current $\boldsymbol{j}$ at the position $\boldsymbol{r}$ is determined by

$$\begin{aligned} \boldsymbol{j}(\boldsymbol{r}) &= -D_\mu \nabla \mu(\boldsymbol{r}) \\ &= -D \nabla n(\boldsymbol{r}), \end{aligned} \tag{8.1}$$

where D_μ and D are the diffusion coefficients and $n(\boldsymbol{r})$ is the local number density of atoms. In a dilute system, chemical potential is proportional to density n and the second line holds.[4] Material is transported to the interface by the gradient of chemical potential. At the interface, with consumption of chemical potential $\Delta\mu$, transformation from liquid (gas) to solid—crystallization—occurs. Such process is called *interface kinetics* (surface kinetics), and it consists of several atomic processes. Macroscopically, the speed of liquid–solid transformation is measured by the velocity of the solid surface V and can be written as

$$V = K\Delta\mu, \tag{8.2}$$

where $\Delta\mu$ is the difference of the chemical potential of the liquid at the interface and that of the solid. The proportionality coefficient K is called the kinetic coefficient. The proportionality is expected, except for a singular surface called a facet. A distinction between the facet and nonsingular faces is essential to understand the crystal growth mechanism, as explained in detail in Chapters 5 and 6 of this book.

[4]Because $\mu = k_\mathrm{B}T \ln n + \text{const.}$, for a small change $\delta\mu = (k_\mathrm{B}T/n)\delta n \propto \delta n$, and $D = D_\mu k_\mathrm{B}T/n$.

8.3 Growth from Various Mother Phases

Crystallization occurs not only from the liquid phase but also from various other phases: from the vapor (from a gas, pure or mixture), from the solution (from a liquid mixture), and from other solid or amorphous phases. This section explains the characteristic features of several basic cases.

8.3.1 Growth from the Vapor

Growth of a solid directly from a gas is called *vapor growth*. It is the reverse process of evaporation (sublimation, more specifically). Growth of snow crystals in the sky and growth of frost on the window glass are examples from everyday life. Because observation of the surface is relatively easy, one can see what is happening there to find the growth mechanism better than in other cases. In a dilute gas, each elementary process contains only a few atoms (molecules), which also helps our understanding. As a result, we have relatively good control of the system.

Features of the vapor growth include the following:

1. Because the crystal grows from a dilute phase, the growth velocity is low. It is convenient for the surface manipulation (e.g., of semiconductor crystals) to make a device but is inefficient for making a big bulk crystal.
2. Unless the crystal is very large, the growth environment can be regarded as uniform, and local difference of the growth conditions is usually negligible.
3. The difference of density in solid and in gas is large, and the anisotropy of the surface is strong. As a consequence, large facets appear on the surface of the crystal.
4. Because most of the surface of a crystal consists of facets, where immediate solidification is usually difficult, various surface processes limit the growth velocity and the growth velocity is slow. These surface processes are very important and will be explained in the following sections.

Let us estimate the *maximum* growth velocity expected for growth from the vapor. If all atoms (or molecules) coming onto the crystal surface are taken into the crystal and solidify, the incoming flux will give the growth rate when evaporation is absent. At finite temperatures, evaporation occurs, and the evaporating flux should be subtracted from the incoming flux to obtain the net flux of solidification. If we denote the incoming current (flux density: number of atoms per unit area per unit time) by j_{in}, the outgoing current by j_{out} and atomic volume in the solid by v_{S}, the growth velocity, which is the advancing velocity of the surface, is given by

$$V = v_{\mathrm{S}}\left(j_{\mathrm{in}} - j_{\mathrm{out}}\right). \tag{8.3}$$

For a classical ideal gas, which in an equilibrium state takes the Maxwell–Boltzmann distribution function,

$$f(\boldsymbol{p})d^3p = \frac{n_{\mathrm{G}}}{\left(2\pi m k_{\mathrm{B}}T\right)^{3/2}} e^{-p^2/2mk_{\mathrm{B}}T} d^3p, \tag{8.4}$$

where n_G is the number density of the gas, j_{in} is easily calculated. The number of atoms incident to a unit surface area in the $x-y$ plane within a time Δt from the above[5] is expressed as

$$\begin{aligned} j_{in}(T,P) &= \int_{-\infty}^{0} dp_z \int_{-\infty}^{\infty} dp_y \int_{-\infty}^{\infty} dp_x \left(-\frac{p_z}{m}\right) f(\boldsymbol{p}) \\ &= \frac{n_G k_B T}{\sqrt{2\pi m k_B T}} = \frac{P}{\sqrt{2\pi m k_B T}} \end{aligned} \tag{8.5}$$

because atoms of the velocity $v_z = p_z/m$ (<0) within the distance $|v_z|\Delta t$ collide with the surface irrespective of the velocity v_x and v_y. The rate of evaporation, in contrast, must be a function of temperature T and is independent of the atmosphere. Calculating the evaporation current $j_{out}(T)$ is difficult, but we know that under the equilibrium vapor pressure $P_{eq}(T)$, with which the solid neither grows nor sublimates, the two currents must balance: $j_{out}(T) = j_{in}(T, P_{eq}(T))$. Therefore, we can write the growth velocity (Eqn (8.3)) as

$$V_{max} = \frac{1}{\sqrt{2\pi m k_B T}} v_S \left(P - P_{eq}(T)\right). \tag{8.6}$$

Because we have assumed that all incoming atoms solidify, this equation, called the *Hertz-Knudsen formula,* gives the maximum growth velocity from the vapor [10,11]. The sticking coefficient—that is, the ratio of atoms captured by the solid surface to all incoming atoms—is less than unity (only at a very rough surface is it close to unity), and the growth is considerably slower.

If supersaturation is not too high, we can obtain a linear relation, as in Eqn (8.2). From the relationship between chemical potential and pressure for an ideal gas:

$$\mu = k_B T \ln P + (\text{function of } T). \tag{8.7}$$

$\Delta\mu$ is related to the pressure difference as

$$\Delta\mu = k_B T \ln\left(\frac{P}{P_{eq}}\right) \approx k_B T \frac{P - P_{eq}}{P_{eq}} = \left(P - P_{eq}\right) v_G, \tag{8.8}$$

where v_G is the volume per atom in gas. Equation (8.6) can be rewritten as

$$V_{max} = \frac{v_S}{v_G \sqrt{2\pi m k_B T}} \Delta\mu = \frac{n_G}{n_S} \frac{1}{\sqrt{2\pi m k_B T}} \Delta\mu. \tag{8.9}$$

Therefore, the maximum value of the kinetic coefficient is

$$K_{max}^{vapor} \approx \frac{v_S}{v_G} \frac{1}{m v_{therm}}. \tag{8.10}$$

8.3.2 Growth from the Melt

When a crystal grows from a liquid phase of the same material, it is called *melt growth.* This is the reverse process of melting.

[5]The space $z \leq 0$ is assumed to be the solid.

Features of melt growth include the following:

1. Because the density difference between the two phases is small, the interfacial free energy and its anisotropy is usually small, and the interface tends to be rough.
2. There is no need for material transport, and a rough interface brings about rapid growth.
3. The velocity of growth is limited by transport of the latent heat. The transport of heat is carried out by heat conduction in the material and/or convection in the liquid. As a result, nonuniformity of the system is usually relevant, and stability of the growth form should be always taken care of.

8.3.2.1 Sharp Interface Picture

One can estimate the growth velocity from the melt as in the vapor growth. When the interface is rough, the growth velocity is determined by the balance of the solidifying current j_{sol} and the melting current j_{mel}. For a rough surface, kink sites, where solidification and melting take place, are everywhere, and the solidifying current is

$$j_{\mathrm{sol}} \approx N_{\mathrm{kink}} \nu_0 W_{\mathrm{S}} e^{-E_{\mathrm{b}}/k_{\mathrm{B}}T}, \tag{8.11}$$

where $N_{\mathrm{kink}} \sim a^{-2}$ (a: lattice constant) is the number of kinks in a unit area, ν_0 is a characteristic frequency of molecular motion (such as the Debye frequency), W_{S} is the number of possible molecular configurations in a solid, and E_{b} is the height of the energy barrier to enter a kink site of the crystal. The energy barrier E_{b} is about the same magnitude as the energy barrier E_{d} for atomic diffusion in the liquid. W_{S} is related to the configuration entropy as $W_{\mathrm{S}} = e^{s_{\mathrm{S}}/k_{\mathrm{B}}}$ and is unity for an isotropic molecule, such as a rare gas atom. For molecules whose orientation in solid is limited, the difference of the number of allowed molecular orientations in liquid, $W_{\mathrm{L}} = e^{s_{\mathrm{L}}/k_{\mathrm{B}}}$, and in solid may be significant. The current of the inverse process is

$$j_{\mathrm{mel}} \approx N_{\mathrm{kink}} \nu_0 W_{\mathrm{L}} e^{-(E_{\mathrm{b}}+\Delta h)/k_{\mathrm{B}}T}, \tag{8.12}$$

where $\Delta h = h_{\mathrm{L}} - h_{\mathrm{S}}$ is the difference of enthalpy (note that the two processes occur under a given pressure).

At the equilibrium temperature T_{m}, the two currents j_{sol} and j_{mel} must be equal and, from Eqns (8.11) and (8.12), we have $h_{\mathrm{L}} - T_{\mathrm{m}} s_{\mathrm{L}} = h_{\mathrm{S}} - T_{\mathrm{m}} s_{\mathrm{S}}$—that is, $\mu_{\mathrm{L}} = \mu_{\mathrm{S}}$. If temperature deviates from T_{m}, the net solidifying current is

$$\begin{aligned} j_{\mathrm{sol}} - j_{\mathrm{mel}} &\approx N_{\mathrm{kink}} \nu_0 W_{\mathrm{S}} e^{-E_{\mathrm{b}}/k_{\mathrm{B}}T} \left(1 - e^{\Delta s/k_{\mathrm{B}}} e^{-\Delta h/k_{\mathrm{B}}T}\right) \\ &= N_{\mathrm{kink}} \nu e^{-\Delta s/k_{\mathrm{B}}} e^{-E_{\mathrm{b}}/k_{\mathrm{B}}T} \left(1 - e^{-\Delta\mu/k_{\mathrm{B}}T}\right), \end{aligned} \tag{8.13}$$

where $\nu = \nu_0 W_L$ and $\Delta s = s_L - s_S$. In the last line, we take the motion in liquid as the standard and write the equation in a form that emphasizes reduction of the growth rate by the difference of configurational entropy Δs. If supersaturation is not too large, the growth velocity is written as

$$\begin{aligned} V &\approx \nu_S (j_{sol} - j_{mel}) \\ &\approx \nu_S N_{kink} \nu e^{-\Delta s/k_B} e^{-E_b/k_B T} \frac{\Delta\mu}{k_B T}, \end{aligned} \tag{8.14}$$

and is proportional to the driving force $\Delta\mu$. The temperature dependence of the kinetic coefficient

$$K^{melt} \approx \frac{\nu_S N_{kink} \nu}{k_B T} e^{-\Delta s/k_B} e^{-E_b/k_B T} \tag{8.15}$$

$$\approx \frac{a\nu}{k_B T} e^{-\Delta s/k_B} e^{-E_b/k_B T} \tag{8.16}$$

is controlled by the energy barrier E_b. Note that, in the last line, we have assumed that $N_{kink} \approx a^{-2}$: kinks cover the rough surface. Otherwise, a reduction factor $N_{kink} a^2$ is necessary.

In terms of supercooling $\delta T = T_m - T$, the linear relation is written as

$$V = K\Delta\mu = K_T \delta T, \tag{8.17}$$

$$K_T = a\nu e^{-E_b/k_B T} e^{-\Delta s/k_B} \frac{\Delta s}{k_B T}, \tag{8.18}$$

where we have used $\Delta\mu = (s_L - s_S)\delta T = \Delta s \delta T$. The entropy gap is related to the latent heat per molecule, l, as $\Delta s = l/T_m$. Because the diffusion coefficient is given by $D \approx a^2 \nu e^{-E_b/k_B T}$, the first three factors of Eqn (8.18) are simply D/a, and $K^{melt} \approx (D/k_B T a) e^{-\Delta s/k_B}$. The equation for the growth velocity from the melt is called the *Wilson-Frenkel formula* [12,13]. With the Einstein–Stokes relation[6] $D = k_B T/(3\pi\eta a)$, the diffusion coefficient is related to the viscosity η, which is easily measured. Thus, the kinetic coefficient K_T is essentially determined by the viscosity of the liquid, but reduced by a factor $e^{-\Delta s/k_B}$ if the entropy gap is not negligible. By lowering the temperature, the growth velocity first increases linearly with δT. When the temperature becomes too low, however, the viscosity increases exponentially ($\eta \propto e^{E_b/k_B T}$) and solidification stops: in extreme cases, amorphous material may appear instead of crystal.

For metals, the values of kinetic coefficients are estimated as $K_T \approx$ 1–50 cm/s K ($K_T \approx 10^{15-17}$ s/g cm) [3]. In real systems, because the transport of latent heat is the rate-limiting process, the apparent kinetic coefficient is much smaller if one uses the values of $\Delta\mu$ or δT measured far[7] from the interface.

[6] We put the molecular radius as $a/2$, half the lattice constant.
[7] Macroscopically close sometimes means microscopically far.

It is known from molecular dynamics simulations that the activation type behavior of the kinetic coefficient (Eqn (8.18)) is not true for simple materials, such as molecules with van der Waals interactions [14]. It seems that crystallization proceeds without an energy barrier and $av e^{-E_b/k_B T}$ is replaced by $a v_{\mathrm{therm}}/l_{\mathrm{mf}}$, where $v_{\mathrm{therm}} = \sqrt{3k_B T/m}$ is the thermal velocity and $l_{\mathrm{mf}}(\sim a)$ is the mean free path of an atom. To understand such behavior, the image of melt growth must be modified.

8.3.2.2 Diffuse Interface Picture

The kinetic coefficient given above is derived with the assumption that the incorporation of an individual atom to the solid is the elementary process of crystallization. In this viewpoint, each atom belongs to the solid or to the liquid, and the boundary between solid and liquid is sharply defined, although it may be very rough. In reality, as indicated by many molecular dynamics simulations, it is not always possible to divide liquid and solid sharply at a rough interface, and collective motion in the liquid is significant. The periodic arrangement of atoms in the solid penetrates into the liquid phase with rather slow decay. Then, growth of crystal is nothing but advancement of such an order to the region previously considered as liquid. In this process, many atoms shift their average positions only slightly, and the crystalline order grows simultaneously in several atomic layers. Therefore, the growth velocity can be much faster than that in the growth of single atoms.

Historically, a continuum model of a diffuse interface was introduced by van der Waals [15] and has been studied in the context of a first-order phase transition as the Ginzburg-Landau model [16]. The important conclusion derived from the continuum model is that the kinetic coefficient is essentially given by the interface width divided by the relaxation time of the local order. (A modern continuum model is the phase field model. Various versions of it are used in many other fields, and the models in crystal growth are explained in Chapter 15 in Volume IB.) Cahn studied the growth of a periodic structure in a diffuse interface with the use of a lattice model [17]. More recently, an estimation of the real kinetic coefficient based on a picture of the diffuse interface was given by Mikheev and Chernov [18]. In this picture, the relevant quantity to describe the system is the order parameter η, which changes continuously at the interface. The local number density may be represented as follows (Figure 8.4):

$$n(\boldsymbol{r}, t) = n_C + \sum_{\boldsymbol{G}} \eta_{\boldsymbol{G}}(z - Rt) e^{i\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{8.19}$$

where $\boldsymbol{G}$ is the reciprocal lattice vector and R is the velocity of the interface. With a detailed theoretical analysis, the following formula for the kinetic coefficient is proposed [19,20]:

$$K \approx \frac{n_L^2 S(G_1)}{k_B T \tau_{G_1}} \left[\sum_{|\boldsymbol{G}|=G_1} \int \left(\frac{d\eta_{\boldsymbol{G}}}{dz} \right)^2 dz \right]^{-1}, \tag{8.20}$$

where $S(G_1)$ is the structure factor of the liquid for the smallest $\boldsymbol{G}$ (G_1 is assumed to be approximately equal to the wave number at the peak of $S(\boldsymbol{k})$ in the liquid). The relaxation time of the density fluctuation, τ_{G_1}, is determined from the half-width of the peak $S(\boldsymbol{k},\omega)$.

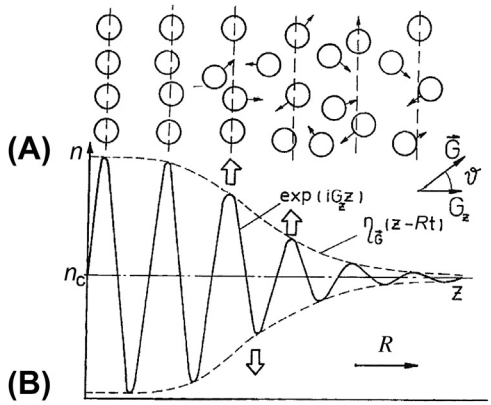


FIGURE 8.4 Diffuse interface between solid (left) and liquid (right) propagating during growth [18,19]. (A) Atoms are ordered in the crystal and disordered in the melt. (B) Profile of the number density is assumed to have the form $n(z) = n_C + \eta_G(z - Rt)e^{i\boldsymbol{G}\cdot\boldsymbol{r}}$, with $n_L \approx n_S \equiv n_C$. *Reprinted from Ref. [19], with permission from Elsevier.*

The last factor with n_L^2 roughly corresponds to the interface width, which is larger than the atomic distance and contributes to the increase in K as expected from the simple continuum theory [16]. The Eqn (8.20) can be estimated with measurable quantities,[8] and gives a good values of K for simple metals [19]. It provides the correct hierarchy of K for different faces. The order of magnitude of the estimated values $K \sim (mv_{\text{therm}})^{-1}$ agrees reasonably with the experimental data. The analysis also suggests that the growth is limited by formation of the lateral order in the interface plane.

8.3.3 Growth from a Solution

Although the major pathways of crystal growth are from the melt and from the vapor, growth from a solution is also very important. A crystal, simple substance or compound, can grow from a liquid phase with additional components, which is called *solution growth*.[9] Many kinds of crystals, such as salt and quartz, are grown from solutions. Salt (NaCl) crystals are easily produced from seawater at room temperature despite that the melting temperature is 800 °C. The melting temperature of quartz (SiO_2) is very high as 1610 °C, but industrial quartz is produced from an alkaline solution under high pressure at approximately 350 °C (hydrothermal synthesis).

Solution growth is similar to vapor growth in the respect that growth occurs with an excess of solute concentration c. Solidification proceeds via incorporation of an atom, which exists at the probability $ca^3 \approx cv_S$ in an atomic volume of the solution, into a kink site by passing over an energy barrier. In solution growth, it is sometimes necessary for an atom (or a molecule) to break bonds with solvent molecules in solidification, which is

[8]The interface width is identified as the half width of the peak of $S(\boldsymbol{k})$ in the liquid.
[9]Sometimes, it is called flux growth when inorganic compounds are used as solvents.

called desolvation. Therefore, by a similar argument that led to Eqn (8.14), we obtain the growth velocity from the solution

$$\begin{aligned} V &\approx v_S(j_{sol} - j_{mel}) \\ &\approx v_S^2 N_{kink} \nu e^{-\Delta s/k_B} e^{-(E_b+E_{des})/k_B T}(c - c_{eq}), \end{aligned} \tag{8.21}$$

where E_{des} is the extra energy barrier for desolvation. In a dilute solution, the chemical potential μ_{sol} is related to the concentration as $\mu_{sol} = k_B T \ln c +$ (function of T), and chemical potential μ_S in the solid is the equilibrium value μ_{eq}. Then, $\Delta\mu = \mu_{sol} - \mu_S = \mu_{sol} - \mu_{eq} = k_B T \ln(c/c_{eq})$, and we obtain

$$\begin{aligned} V &\approx N_{kink} \nu v_S^2 c_{eq} e^{-\Delta s/k_B} e^{-(E_b+E_{des})/k_B T}\left(e^{\Delta\mu/k_B T} - 1\right) \\ &\approx N_{kink} \nu v_S^2 \frac{c_{eq}}{k_B T} e^{-\Delta s/k_B} e^{-(E_b+E_{des})/k_B T} \Delta\mu. \end{aligned} \tag{8.22}$$

The linear relation in the last line, with the kinetic coefficient

$$K \approx N_{kink} \nu v_S^2 \frac{c_{eq}}{k_B T} e^{-\Delta s/k_B} e^{-(E_b+E_{des})/k_B T} \tag{8.23}$$

is valid for small supersaturation. Compared with the melt growth in Eqn (8.16), there is an extra factor $v_S c_{eq} e^{-E_{des}/k_B T}$, which is very small for a dilute solution and for a high desolvation energy barrier. Solution growth is much slower than melt growth.

8.4 Normal Growth and Lateral Growth

Except for faceted faces, the surface of a crystal is rough and there are kinks everywhere.[10] The surface can grow in the direction normal to the local surface orientation, on a microscopic scale, by incorporating surrounding atoms or molecules. The growth is called *normal growth* or *adhesive growth*. On the other hand, if the surface is faceted, an atom or a molecule needs to find a kink site to crystallize stably. Kinks are present along a step on the faceted surface, and steps on the facet move forward along the surface so that the surface advances. Because the crystal grows by the lateral motion of steps, the growth is called *lateral growth* or *layer growth*.

8.4.1 Adhesive Growth on Rough Faces

For adhesive growth to occur, the surface must be rough—that is, temperature must be higher than the roughening temperature of the face. As we have seen in Eqn (8.3), the normal growth velocity V of the surface is proportional to the supersaturation $\Delta\mu$ near

[10] Meanings of rough and smooth faceted faces are explained in Chapters 5 and 6 and depicted schematically in Figure 8.5.

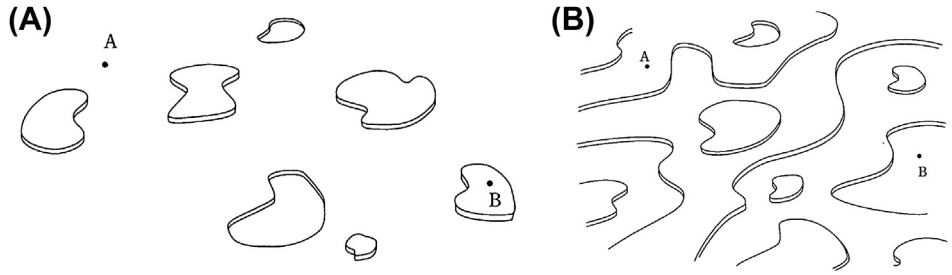


FIGURE 8.5 Smooth (A) and rough (B) faces. The height difference of A and B is within the lattice constant a_z in (A). The height difference is expected to increase logarithmically with the distance between A and B in (B).

the kink sites. Because kinks are everywhere, $\Delta\mu$ is practically the same all over the surface. Such a linear relationship between the growth velocity and the supersaturation, $V = K\Delta\mu$, is characteristic for rough surfaces. Thus, the kinetic coefficient K is a well-defined quantity for rough surfaces.

The linearity does not hold for a facet because supersaturation is not uniform on the surface and the local supersaturation at the growth site differs depending on the configuration of kinks (or steps). To understand the growth of a faceted surface, we should consider the generation mechanism of steps and the motion of steps, of which kinks are present at the edge.

8.4.2 Growth via Two-Dimensional Nucleation on a Facet

On a facet of a perfect crystal at very low temperatures, there are no steps, so the facet cannot grow at all. At finite temperatures, there are small two-dimensional (2D) islands or holes that are thermally excited, as shown in Figure 8.5(A). For a 2D island of typical radius R, the step edge costs energy typically $2\pi R\beta$, where β is the step free energy per unit length.[11] Near equilibrium, $\Delta\mu \approx 0$, as a result of thermal fluctuation, islands whose energy is of the order of k_BT or smaller may be created. Thus, the typical size of the thermally created islands is $\xi \sim k_BT/(2\pi\beta)$, which is the correlation length of the height. If such a thermally excited island grows, the energy will increase. Most islands tend to lower their energy and disappear soon.

If the environment becomes supersaturated, the free energy of a circular 2D island is given by (Figure 8.6):

$$G_2(R) = -\frac{\pi R^2}{\Omega_2}\Delta\mu + 2\pi R\beta, \tag{8.24}$$

[11]Roughly speaking, the free energy density of a step, β, vanishes above the roughening transition temperature, and the formation of an island does not require any extra free energy on a rough surface.

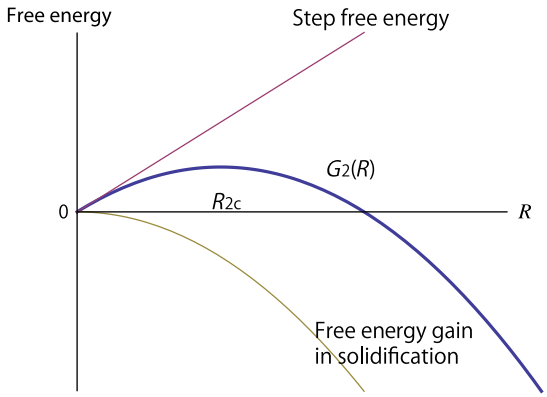


FIGURE 8.6 Free energy of a circular two-dimensional nucleus as a function of the radius.

where $\Omega_2 = 1/(a_z n_S) = v_S/a_z$ is the atomic area and a_z is the atomic height. The first term is the free energy gain by the solidification of $\pi R^2/\Omega_2$ atoms. In a supersaturated state, $\Delta\mu > 0$, G_2 increases with the size and takes a maximum value

$$G_2(R_{2c}) = \pi \frac{\Omega_2 \beta^2}{\Delta\mu} \equiv G_{2c} \tag{8.25}$$

at the radius

$$R_{2c} = \frac{\Omega_2 \beta}{\Delta\mu}, \tag{8.26}$$

and decreases beyond this point. The 2D island of the size R_{2c} is at unstable equilibrium with the environment, called the *critical 2D nucleus*. R_{2c} is called the radius of a critical nucleus, and G_{2c} is its free energy. According to the nucleation theory described in Chapter 7, if an island larger than the critical radius appears as a result of thermal fluctuation, it most likely grows forever. This growth of an island is *nucleation* of the new layer of the crystal.

The frequency of the appearance of a critical nucleus by thermal fluctuation [20] is determined by the critical free energy G_{2c}. From the nucleation theory, the steady-state nucleation rate per unit area per unit time is given by the product of three quantities [4,5,21]:

$$j_{nuc} \approx Z w_c^+ n_c^{eq}. \tag{8.27}$$

The last factor, n_c^{eq}, is the equilibrium density of critical islands expressed as

$$n_c^{eq} \approx n_1 e^{-G_{2c}/k_B T} \sim \frac{e^{-G_{2c}/k_B T}}{\Omega_2}, \tag{8.28}$$

where n_1 is the density of atoms on the surface and is $\sim\Omega_2^{-1}$ for melt growth. This quantity is controlled by the critical free energy G_{2c} and the most sensitive factor to $\Delta\mu$ in j_{nuc}. The second factor, w_c^+, is the attachment rate of atoms to the critical island; it is proportional to the edge length $2\pi R_c = 2\pi\Omega_2\beta/\Delta\mu$ and to the incoming current j_{in}. Its

explicit expression depends on systems we consider. For surface diffusion with diffusion length x_s, it is the number of impinging atoms onto a zone near the step edge of the width $2x_s$ (assuming $R_c \gtrsim x_s$):

$$w_c^+ \approx 4\pi x_s R_c f = \frac{4\pi x_s \Omega_2 \beta}{\Delta\mu} f, \tag{8.29}$$

where f is the impingement rate of atoms onto the surface per unit area. For simple melt growth,

$$w_c^+ \approx \frac{a\nu_0 e^{-E_b/k_B T}}{\Omega_2} 2\pi R_c = 2\pi a \nu_0 e^{-E_b/k_B T} \frac{\beta}{\Delta\mu}. \tag{8.30}$$

The first factor Z is the Zeldovich factor, which appears as a result of stochastic nature of nucleation (see Chapter 7), representing broadness of the peak of the free energy (Eqn (8.24)):

$$Z = \sqrt{\frac{(\Delta\mu)^3}{4\pi^2 \Omega_2 k_B T \beta^2}}. \tag{8.31}$$

From Eqns (8.28), (8.30), and (8.31), the nucleation rate (Eqn (8.27)) is estimated as [21,22]

$$j_{nuc} \sim \frac{\overline{\nu}}{\Omega_2} \left(\frac{\Delta\mu}{k_B T}\right)^{1/2} e^{-\pi\Omega_2\beta^2/(\Delta\mu k_B T)}, \tag{8.32}$$

where $\Omega_2^{-1} = a^{-2}$ is the number of sites where nucleation is possible, and $\overline{\nu} = \nu_0 e^{-E_b/k_B T}$ is a typical frequency of the atomic attachment to the island edge per site.

8.4.2.1 *Single-Nucleation versus Multinucleation*

When supersaturation is weak, G_{2c} is large so that $e^{-G_{2c}/k_B T}$ is extremely small: the nucleation rate is practically zero. If supersaturation is increased, nucleation may occur within the observation period, and the facet starts to grow. For a very small facet, the nucleated supercritical 2D island spreads to cover the whole facet, which means that the growth stops until a new supercritical island is born on the new layer. Such intermittent growth is called *single-nucleation growth* (Figure 8.7(A)). The growth velocity of a facet of the size R_f is the lattice constant times the nucleation rate:

$$V \approx \pi a R_f^2 j_{nuc} = a\overline{\nu} \frac{\pi R_f^2}{\Omega_2} \left(\frac{\Delta\mu}{k_B T}\right)^{1/2} e^{-\pi\Omega_2\beta^2/\Delta\mu k_B T}. \tag{8.33}$$

For a large facet, many nucleation events occur before a single island covers the whole facet. Many islands coalesce to form a new atomic layer. The growth is continuous and called *multinucleation growth* (Figure 8.7(B)).

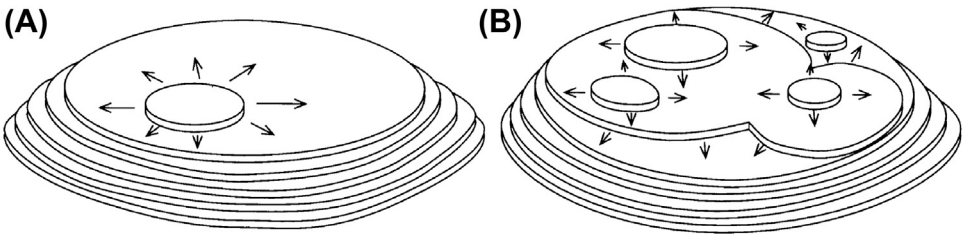


FIGURE 8.7 (A) Single-nucleation growth. (B) Multinucleation growth.

8.4.2.2 Growth of a 2D Circular Island

To find the growth velocity in multinucleation growth, we need to know the growth of a 2D island under constant supersaturation. For a circular island bound by an isotropic step, the effective driving force of the step edge is

$$\Delta\mu_{\text{eff}} = \Delta\mu - \frac{\Omega_2\tilde{\beta}}{R}, \tag{8.34}$$

where the second term is the retraction force due to the *step stiffness*[12] $\tilde{\beta}$. Because the step is always rough at a finite temperature, the velocity is proportional to the driving force:

$$\begin{aligned} V_{\text{st}} &= K_{\text{st}}\left(\Delta\mu - \frac{\Omega_2\tilde{\beta}}{R}\right) \\ &= K_{\text{st}}\Delta\mu\left(1 - \frac{R_{2c}}{R}\right), \end{aligned} \tag{8.35}$$

where K_{st} is the kinetic coefficient of the step. The second line is expressed in terms of the critical radius R_{2c}, and it is valid if the crystal anisotropy in the plane can be neglected.[13] The island is then circular, and the radius at time t is obtained by integrating Eqn (8.35) as

$$\frac{K_{\text{st}}\Delta\mu}{R_{2c}}t = \frac{R}{R_{2c}} + \ln\left|\frac{R}{R_{2c}} - 1\right| + \text{const.} \tag{8.36}$$

If the initial radius is smaller than the critical radius $R < R_{2c}$, the island will disappear. When $R \ll R_{2c}$, it shrinks as

$$R(t) \approx R_{2c}\left[\frac{2K_{\text{st}}\Delta\mu}{R_{2c}}(t_0 - t)\right]^{1/2} = \left[2K_{\text{st}}\Omega_2\tilde{\beta}(t_0 - t)\right]^{1/2}, \tag{8.37}$$

where t_0 is the time it vanishes. If the initial radius is larger than the critical radius $R > R_{2c}$, the island expands. When the island becomes much larger than the critical radius, $R \gg R_{2c}$, the step stiffness may be neglected and move at a constant

[12]The stiffness of a step, $\tilde{\beta}(\phi)$, is the most important quantity to characterize the step when crystal anisotropy is relevant. It is the sum of $\beta(\phi)$, which is now a function of the azimuthal angle ϕ, and its second derivative $\beta''(\phi)$: $\tilde{\beta} = \beta + \beta''$.

[13]For an isotropic step, the stiffness is constant and the same as the free energy density: $\tilde{\beta} = \beta$.

speed: $R(t) = K_{st}\Delta\mu\, t$. Except for a very early stage, the radius of an island created by 2D nucleation grows at a constant speed:

$$V_{st} = K_{st}\Delta\mu. \tag{8.38}$$

8.4.2.3 Change of Coverage by Nucleation and Spread

To estimate the velocity of multinucleation growth, let us first consider the change of coverage of a new layer on a flat facet by nucleation and spread of 2D islands [16,23,24]. For simplicity, we consider the growth of a single layer. We assume that nucleation occurs randomly at constant probability and circular islands spread at the constant step velocity V_{st}. At time t, an arbitrary point P located at $\boldsymbol{x}$ is still on the initial facet surface with a probability p—that is, the ratio p of the facet area is not covered by the new layer. The point P is covered by the new layer if there was a nucleation event at the location $\boldsymbol{x}'$ and the time t' that satisfies the following:

$$V_{st}(t - t') > |\boldsymbol{x} - \boldsymbol{x}'|.$$

If we denote $t - t' = \bar{t}$ and $\boldsymbol{x}' - \boldsymbol{x} = \bar{\boldsymbol{x}}$, the probability p is given by the product of probabilities that nucleation does not occur at the point satisfying the condition $V_{st}\bar{t} > |\bar{\boldsymbol{x}}|$. That is,

$$p = \prod_{0<|\bar{\boldsymbol{x}}|<V_{st}\bar{t},\ \ 0<\bar{t}<t} \left(1 - j_{nuc} d^2\bar{x}\, d\bar{t}\right). \tag{8.39}$$

Because the second term in the bracket is infinitesimal, this equation can be witten as

$$\begin{aligned} p &= \exp\left[\ln\prod(1 - j_{nuc} d^2\bar{x}\, d\bar{t})\right] \\ &= \exp\left(-\sum j_{nuc} d^2\bar{x}\, d\bar{t}\right) \\ &= \exp\left(-j_{nuc}\int_0^t d\bar{t}\int_0^{V_{st}\bar{t}} 2\pi\bar{x}\, d\bar{x}\right) \\ &= \exp\left(-\frac{\pi}{3} j_{nuc} V_{st}^2 t^3\right). \end{aligned}$$

Therefore, the coverage of the new layer Θ_{2D} changes as

$$\Theta_{2D}(t) = 1 - p = 1 - \exp\left(-\frac{\pi}{3} j_{nuc} V_{st}^2 t^3\right). \tag{8.40}$$

As shown in Figure 8.8, it initially increases as

$$\Theta_{2D}(t) \approx \frac{\pi}{3} j_{nuc} V_{st}^2 t^3, \tag{8.41}$$

and approaches unity at about $j_{nuc}^{1/3} V_{st}^{2/3} t \approx 1.5$. A similar calculation in one and three spatial dimensions gives the following results for the advancement of a straight step and volume occupation of a new phase:

$$\Theta_{1D}(t) = 1 - \exp\left(-\frac{1}{2} j_{nuc} V_k t^2\right), \tag{8.42}$$

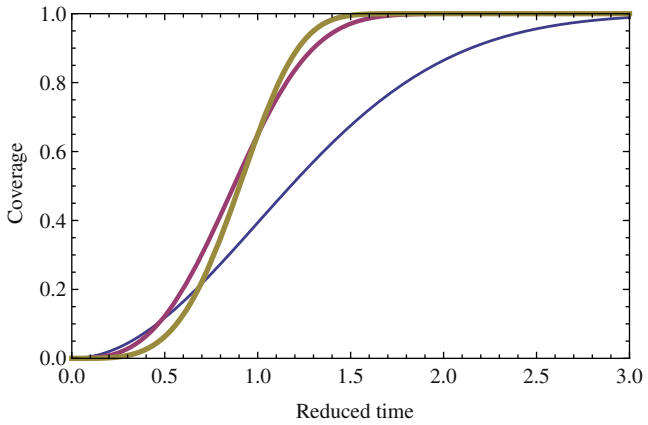


FIGURE 8.8 Coverage of the new phase, in one, two, and three dimensions (thin line to thick line). The reduced time is t multiplied by $(j_{\mathrm{nuc}}V_{\mathrm{k}})^{1/2}$, $(j_{\mathrm{nuc}}V_{\mathrm{st}}^2)^{1/3}$ and $(j_{\mathrm{nuc}}V^3)^{1/4}$, respectively.

$$\Theta_{3\mathrm{D}}(t) = 1 - \exp\left(-\frac{\pi}{3} j_{\mathrm{nuc}} V^3 t^4\right), \tag{8.43}$$

where V_{k} is the velocity of a kink and V is the velocity of the interface.

8.4.2.4 *Velocity of Multinucleation Growth*

Because one atomic layer is covered in a period $\tau \sim (j_{\mathrm{nuc}} V_{\mathrm{st}}^2)^{-1/3}$, the velocity of a facet growing by multinucleation growth is

$$\begin{aligned} V = \frac{a}{\tau} &\approx a j_{\mathrm{nuc}}^{1/3} V_{\mathrm{st}}^{2/3} \\ &= a\left(\frac{\bar{\nu}}{\Omega_2}\right)^{1/3}\left(\frac{\Delta\mu}{k_{\mathrm{B}}T}\right)^{1/6} e^{-G_{2c}/3k_{\mathrm{B}}T_{(K_{\mathrm{st}}\Delta\mu)^{2/3}}} \\ &\sim (\Delta\mu)^{5/6} e^{-\pi\Omega_2\beta^2/\left(3k_{\mathrm{B}}T\Delta\mu\right)}. \end{aligned} \tag{8.44}$$

Note that the energy of a critical nucleus in the exponential factor is divided by three compared with Eqn (8.32). In the case of step growth limited by surface diffusion, the step velocity is proportional to $c - c_{\mathrm{eq}} = c_{\mathrm{eq}}(e^{\Delta\mu/k_{\mathrm{B}}T} - 1) \approx \Delta\mu/k_{\mathrm{B}}T$, and the prefactor becomes proportional to $(\Delta\mu/k_{\mathrm{B}}T)^{5/6}$. The result is consistent with the Monte Carlo simulation [25], which demonstrates the sharp contrast of the $\Delta\mu$ dependence of growth velocity below and above the roughening transition. In the experiment, the growth velocity of a ^{4}He crystal below the roughening transition temperature was measured with changing[14] $\Delta\mu$ [26]. The data were analyzed with a simpler formula[15]: $V \sim V_{\mathrm{st}} e^{-G_{2c}/3k_{\mathrm{B}}T} = K_{\mathrm{st}}\Delta\mu e^{-\pi\Omega_2\beta^2/(3k_{\mathrm{B}}T\Delta\mu)}$. The plots $\ln(V/\Delta\mu)$ versus $(\Delta\mu)^{-1}$ at various temperatures show clear linear dependence, and from their slopes the step free energy density $\beta(T)$ was determined. The result showed a singular temperature dependence and supported the theory of roughening transition (see Chapter 6) [6].

[14]Although this is a melt growth experiment, the temperature is fixed and $\Delta\mu$ is varied by the change of pressure.

[15]The present system is a quantum system, and the pre-exponential factor is different from Eqn (8.44).

In real growth of a macroscopic facet, there are steps originating from dislocations exposed on the facet, and the growth proceeds via spiral growth, which is explained in the following section. The 2D nucleation growth becomes important only under high supersaturation.

8.4.3 Spiral Growth with Screw Dislocations

If a crystal is perfect without any defect, its facets cannot grow under weak supersaturation because the free energy of a critical 2D island is much larger than the thermal energy, $\delta G_{2c} \gg k_B T$, and nucleation never occurs during the observation period. In real crystals, however, facets certainly grow under weak supersaturation. This puzzle was solved by Frank's proposal that facets grow with the help of screw dislocations that always exist in a macroscopic crystal [27–29].

A screw dislocation is a topological defect of a crystal lattice. If one moves around the dislocation, the lattice plane shifts by one layer (or more layers), like a spiral staircase. The Burgers vector of a screw dislocation is parallel to the dislocation line. If the screw dislocation is exposed on a facet, a step emerges at the end of the dislocation and stretches to an edge of the facet (Figure 8.9(B)) When the chemical potential of the environment is changed, the step advances or recedes according to the sign of $\Delta\mu$, as shown in Figure 8.9(A) and (C). Because the end of the step is pinned at the dislocation,[16] the shape of the step becomes spiral. Such growth helped by a dislocation is called *spiral growth*. The advantage of spiral growth is that the step does not disappear, and the growth continues without the creation of a new step.

8.4.3.1 Velocity of Spiral Growth

The velocity of a step is proportional to supersaturation, $V_{st} = K_{st}\Delta\mu_{eff}$, and growth is possible under small supersaturation. If the distance between steps is l (Figure 8.9(C)), the growth velocity V perpendicular to the facet is $V_{st}(a_z/l)$. Because, far apart from the dislocation, the curvature of the step is very small, the step velocity is $V_{st}^{\infty} = K_{st}\Delta\mu$, and the growth velocity is given by

$$V = \frac{a_z}{l} K_{st}\Delta\mu. \tag{8.45}$$

FIGURE 8.9 Spiral step with a dislocation: (A) in melting, $\Delta\mu < 0$; (B) in equilibrium, $\Delta\mu = 0$; (C) in growth, $\Delta\mu > 0$.

[16]Dislocations may be moved with elastic stress, but not with $\Delta\mu$.

The shape of the spiral step is similar to an Archimedean spiral, whose expression in the polar coordinates is $r(\phi) = A\phi$. In the polar coordinates, the curvature of the Archimedean spiral is

$$\frac{1}{R} = \frac{r^2 + 2(r')^2 - rr''}{\left(r^2 + (r')^2\right)^{3/2}} = \frac{A^2\phi^2 + 2A^2}{\left(A^2\phi^2 + A^2\right)^{3/2}}. \tag{8.46}$$

Near the origin (dislocation), where $\phi \to 0$, the curvature is $R^{-1} = 2/A$. Because the step advances little at the center of the spiral, the curvature there must be R_c, which means $A = 2R_{2c}$. Thus, with one rotation, the distance from the center increases by $2\pi A = 4\pi R_{2c} = l$. For a more accurate calculation, one must solve Eqn (8.35) with Eqn (8.46) numerically. The result is $l \approx 19R_{2c}$ [30], and the growth velocity is

$$V \approx \frac{a_z}{19R_{2c}} K_{st}\Delta\mu = K_{st}\frac{a_z}{19\Omega_2\tilde{\beta}}(\Delta\mu)^2. \tag{8.47}$$

Because the step distance is inversely proportional to $\Delta\mu$, the growth velocity is proportional to $(\Delta\mu)^2$, which is the characteristic feature of spiral growth.

Here, we have assumed that the step is isotropic. If a step is anisotropic, the spiral is also anisotropic and the coefficient differs from 1/19. For example, if the equilibrium shape of the 2D island is square, the coefficient becomes $(4L_{2c})^{-1}$ instead of $(19R_{2c})^{-1}$, where $L_{2c} = 2\tilde{\beta}/\Delta\mu$ is the side length of the square critical island. This relationship is derived by assuming the straight step can grow only when the length exceeds L_{2c} [5].

8.4.3.2 *Growth of a Facet with Many Dislocations*

It is important to note that the growth velocity of a facet does not depend on the number of dislocations as long as they are nonzero [28]. Figure 8.10(A) shows spiral growth with a pair of screw dislocations of the opposite sign. If the Burgers vector of the dislocation is opposite, the spiral winds in the opposite direction. The two spirals originating from the opposite dislocations collide and merge to form concentric circles asymptotically. The distance of successive steps is the same as the original spiral, and the growth velocity remains the same as that with a single step. For two dislocations of the same sign (the same Burgers vector), a similar recombination of steps occur if the centers of the spirals

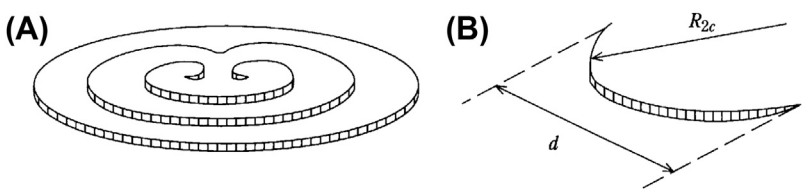


FIGURE 8.10 (A) Spiral growth with a pair of screw dislocations. (B) A step pinned by a pair of screw dislocations.

are not too close. If their distance d is smaller than the period of the spiral l, they operate additively, and the efficiency as a step source can be twice as that of a single dislocation, at the best.[17] Therefore, the growth velocity of the whole facet is not much larger than that with a single dislocation.

Although we find the growth velocity is proportional to $(\Delta\mu)^2$, there is a small threshold value of supersaturation for spiral growth to work. If there is a pair of dislocations of the opposite sign as shown in Figure 8.10(B), the distance d should be larger than $2R_{2c}$ for the pair to operate as a step source. Otherwise, the advancement of the step is blocked. Therefore, for growth of a facet, supersaturation should be stronger than the threshold value

$$\Delta\mu_{\mathrm{th}} = \frac{2\Omega_2\tilde{\beta}}{d}, \tag{8.48}$$

where d is the narrowest channel for the step on the facet.[18] For a small facet, growth is sensitive to the configuration of dislocations: their number and location, the magnitude of Burgers vector, etc. [31].

8.4.3.3 Change of Growth Mode and Kinetic Roughening

In thermal equilibrium, only small islands of the size $\xi \sim k_{\mathrm{B}}T/\tilde{\beta}$ appear temporarily as fluctuations, and the position of a facet is locked at a single atomic layer: the surface is atomically smooth (Figure 8.5(A)). Under weak supersaturation, the 2D nucleation rate is so low that the facet never grows without dislocations. A facet under weak supersaturation is thermodynamically in a *metastable equilibrium* state. With screw dislocations, the facet grows above the threshold supersaturation $\Delta\mu_{\mathrm{th}}$. If $\Delta\mu \gg \Delta\mu_{\mathrm{th}}$, the growth velocity is proportional to $(\Delta\mu)^2$. By increasing supersaturation, 2D nucleation growth occurs in parallel to the spiral growth and eventually dominates the growth (Figure 8.11). The successive nucleation of islands blurs the flat facet. Under high supersaturation, the free energy of the critical nucleus, as in Eqn (8.25), becomes small and comparable to the thermal energy $k_{\mathrm{B}}T$. The condition $\delta G_{2c} \sim k_{\mathrm{B}}T$ is equivalent to the condition $R_{2c} \sim \xi$, which means that the critical radius is of the same order of the characteristic length of step fluctuation. There, so many steps appear on the facet and distinction between steps originating from dislocations versus 2D nucleation is no longer possible. The facet is atomically rough and its growth velocity becomes as fast as rough faces. The gradual crossover phenomenon, with increasing supersaturation, from a facet to a rough face is called *kinetic roughening*.[19]

[17] In fact, a dislocation with a large Burgers vector works very efficiently as a step source.

[18] Like dislocations, impurities may pin down steps on a facet. As a result, scattered impurities on a facet can retard the motion of steps and the growth. If the concentration of impurities on a facet becomes too high, they may completely block the growth of the facet and hysteresis may be observed in the relationship between the supersaturation and the growth velocity [32,33].

[19] The concept of kinetic roughening has been used in a wider field of study than crystal growth. Some of the reviews describing crystal growth are Refs [34–36].

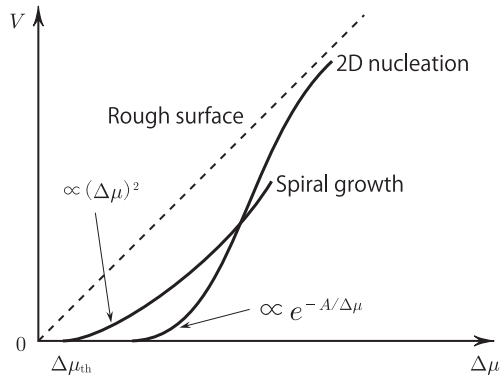


FIGURE 8.11 Change of growth mode by increasing supersaturation.

8.4.3.4 *Formation of Mounds*

We estimated the characteristic time τ necessary to cover one layer by considering a single layer system. In reality, nucleation occurs on a new layer before the new layer covers the whole surface: growth proceeds on many layers simultaneously. During growth from a vapor, transport along the surface–surface diffusion is important.[20] On a growing 2D island, nucleation of the new layer is likely to occur near the center of the island because atoms landing near the step edge are consumed by the step growth and the density of atoms near the center is higher. By successive nucleation in the central area of new layers, a nearly concentric multilayer island is formed when supersaturation is sufficiently high. Such a multilayer island is called a mound. Note that in growth at low temperatures, evaporation is weak and nonuniformity of the height can be produced only when the transport of material along the surface occurs. Thus, surface diffusion is necessary for the formation of mounds. Mounds are formed by several ways; detailed explanations are given in Refs [34,37,38].

8.5 Growth of Vicinal Faces

A face slightly tilted from a facet consists of an array of parallel steps and terraces bound by the steps. Such a face is called a *vicinal face*. A vicinal face grows by layer growth—that is, lateral advancement of parallel steps. Unlike the growth of a facet, the average distance l of parallel steps is determined by external conditions, such as the tilt angle in cutting the crystal.[21]

The growth velocity in the normal direction to the terrace is

$$V_z = \frac{a_z}{l} V_{\mathrm{st}} = a_z n_{\mathrm{st}} V_{\mathrm{st}}, \tag{8.49}$$

[20]The role of surface diffusion in vapor growth is explained in Section 8.5.2.
[21]If a cone formed by a spiral step is large, its side face can be regarded as a vicinal face.

where l is the distance between steps and $V_{st} = K_{st}\Delta\mu_{eff}$ is the velocity of a step. The growth velocity in the direction normal to the face is

$$V = V_z \cos\theta = V_{st} \sin\theta. \tag{8.50}$$

The tilt angle θ (or the slope p) and the step density n_{st} are related as $a_z n_{st} = \tan\theta$ $(=p)$. If θ is very small, the distance between steps is very large and each step can move independently. If the distance of neighboring steps becomes shorter, the steps compete each other for obtaining material or evacuating heat via overlapping diffusion fields. Here, we consider two typical cases in which the step velocity changes with the distance l depending on the transport mechanism.

8.5.1 A Model of Solution Growth: Bulk Diffusion

In solution growth, flow of the solution (either artificial or natural) makes the solute concentration and the temperature uniform. Because of the viscosity of the liquid, however, the solution near the crystal surface is usually stagnant and diffusion in the bulk liquid is important. A simple model representing this situation is shown in Figure 8.12 [29]. At a distance λ_0 from the surface, the solute is regarded as static and uniform with the bulk concentration c_∞. This hypothetical layer of the thickness λ_0 is called the *diffusion boundary layer*. Diffusion of the solute to the equidistant parallel steps conveys the material for growth of the crystal. A growing step is a sink of the solute, and the diffusion flow brings up the advancement of the step. Approximating the step of the height a with a half cylinder with the radius a/π, the step velocity is related to the diffusion flow as

$$\frac{aV_{st}}{\nu_S} = D\left.\frac{\partial c}{\partial r}\right|_{r=a/\pi} a. \tag{8.51}$$

Solving the diffusion equation $\nabla^2 c = 0$ with the boundary conditions (Eqn (8.51)) and $c(x, y = \lambda_0) = c_\infty$, the step velocity is calculated as [3,29]:

$$V_{st} = \nu_S \frac{\delta c_\infty}{\frac{1}{\kappa_{st}} + \frac{a}{\pi D}\ln\left[\frac{l}{a}\sinh\left(\frac{\pi\lambda_D}{l}\right)\right]}, \tag{8.52}$$

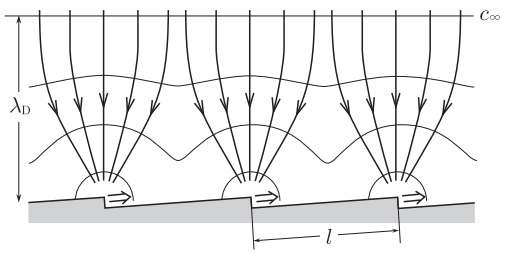


FIGURE 8.12 Contours of the solute concentration and flow-lines to the steps of the solute [29].

where $\delta c_\infty \equiv c_\infty - c_{\rm eq}$, and the kinetic coefficient of the step is defined as[22]

$$\mathcal{K}_{\rm st} = \frac{V_{\rm st}}{\upsilon_{\rm S}\delta c_{\rm st}}, \tag{8.53}$$

where $\delta c_{\rm st} = c(\boldsymbol{r}_{\rm step}) - c_{\rm eq}$ is the supersaturation at the step position. The structure within the boundary layer is difficult to observe, and the experimentally observable kinetic coefficient of the step, $\mathcal{K}_{\rm st}^{\rm eff} = V_{\rm st}/(\upsilon_{\rm S}\delta c_\infty)$, is related to $\mathcal{K}_{\rm st}$ as

$$\frac{1}{\mathcal{K}_{\rm st}^{\rm eff}} = \frac{1}{\mathcal{K}_{\rm st}} + \frac{a}{\pi D}\ln\left[\frac{l}{a}\sinh\left(\frac{\pi\lambda_{\rm D}}{l}\right)\right]. \tag{8.54}$$

The first term represents the resistance of attachment and detachment at the step, and the second term represents the growth resistance of bulk diffusion. If the step distance is larger than the thickness of the boundary layer, the second term becomes independent of the step distance. Then, the growth velocity of the vicinal face is proportional to the slope of the face, $p = a/l$, and $K \sim \mathcal{K}_{\rm st}^{\rm eff}(p \to 0)p$. On the other hand, if the step distance is small, the diffusion field interferes strongly and the growth velocity becomes insensitive to the slope p. If the step distance is close to the lattice constant $l \sim a$, the second term of Eqn (8.54) is $\lambda_{\rm D}/D$, which means the growth resistance of the diffusion is proportional to the thickness of the boundary layer: the face we consider is no longer a vicinal face but practically a rough face.

The kinetic coefficient defined by Eqn (8.53) is intrinsic to the step and does not depend on the surrounding conditions. Its magnitude can be estimated with the use of Eqn (8.23). $N_{\rm kink}a^2$ in the surface kinetic coefficient is replaced by the corresponding kink density $a/l_{\rm kink}$ ($l_{\rm kink}$ is the average kink distance along the step) for a step, and $k_{\rm B}T/(\upsilon_{\rm S}c_{\rm eq}^0)$ is multiplied to convert K to $\mathcal{K}$:

$$\mathcal{K}_{\rm step} \approx \frac{a}{l_{\rm kink}} a\nu e^{-\Delta s/k_{\rm B}} e^{-(E_{\rm b}+E_{\rm des})/k_{\rm B}T}. \tag{8.55}$$

For simple molecules $\Delta s \approx 0$, and $E_{\rm des} = 0$ if desolvation is not necessary at the step. Then we obtain

$$\mathcal{K}_{\rm step} \approx \frac{a}{l_{\rm kink}} a\nu e^{-E_{\rm b}/k_{\rm B}T}. \tag{8.56}$$

To some extent, a similar argument will apply to melt growth if the concentration field of solute is replaced by a temperature field that evacuates the latent heat.

8.5.2 Vapor Growth and Surface Diffusion

In vapor growth, atoms landing on the surface migrate on a flat terrace and crystallize at a kink position along the step. It is important to make a proper model of the surface to understand the growth of a vicinal face.

[22]Here, $\mathcal{K}$ is defined in terms of the concentration c and the current j, and not in terms of the chemical potential μ and the velocity V. Because $\Delta\mu = k_{\rm B}T\ln(c/c_{\rm eq}^0) \approx k_{\rm B}T(\delta c/c_{\rm eq}^0)$, they are related as $K = (\upsilon_{\rm S}c_{\rm eq}^0/k_{\rm B}T)\mathcal{K}$.

8.5.2.1 The BCF Model

The classical model of vapor growth is the BCF model studied by Burton, Cabrera, and Frank [28]. Atoms enter from a uniform gas environment onto the crystal surface at a constant rate: f atoms per unit area per unit time. Those atoms adsorbed on the surface are called *adatoms*. An adatom migrates on the crystal surface and evaporates into the gas with a lifetime τ. Therefore, the number density c of adatoms at the position $\boldsymbol{x}$ ($\boldsymbol{x}$ is a two-dimensional vector) obeys the diffusion equation

$$\frac{\partial c(\boldsymbol{x})}{\partial t} = D_s \nabla^2 c(\boldsymbol{x}) - \frac{c(\boldsymbol{x})}{\tau} + f. \tag{8.57}$$

If adatoms are uniformly distributed, then the balance of evaporation and impingement, the last two terms, determines the adatom density as $c_\infty = f\tau$. In the model, atoms (molecules) consisting of the crystal and atoms migrating on the surface are distinguished. Adatoms cannot solidify without steps, where kink sites are found.

If there is a step on the surface, the adatoms join the crystal there or the solid atoms "melt" from the step onto the surface. Exchange of atoms between the solid and the adatom layer occurs at the step. If the exchange is very fast, the adatom density at the step becomes the equilibrium density:

$$c_{st} \equiv c\left(\boldsymbol{x}_{step}\right) = c_{eq}. \tag{8.58}$$

The equilibrium density at the step is given by

$$c_{eq} = c_{eq}^0 \exp\left(-\frac{f_{st}\Omega_2}{k_B T}\right) \approx c_{eq}^0\left(1 - \frac{f_{st}\Omega_2}{k_B T}\right), \tag{8.59}$$

where c_{eq}^0 is the equilibrium density for a straight step. The exponential factor represents the Boltzmann factor for the extra work necessary if a force f_{st} is acting on the step[23] when an atom of the area Ω_2 solidifies. If $c_\infty > c_{st}$, adatoms flow into the step and the step becomes a sink of adatoms: solidification occurs. If $c_\infty < c_{st}$, adatoms flow out from the step and the step becomes a source of adatoms: melting occurs.

As for the force f_{st} acting on the step, two origins are well known. One is the capillary force due to the step stiffness: $-\tilde{\beta}/R$ corresponding to the second term in Eqn (8.34).[24] The other is a force due to elastic stress, which drives the step to the direction of lower elastic energy [7]. If there are two parallel steps of the same sign (both up or both down steps) at a distance d, a repulsive force proportional to d^{-3} arises to make them move apart [39].

The simple BCF model assumes fast step kinetics—that is, the adatom density takes the equilibrium value c_{eq} at the step. Then, Eqn (8.58) is the boundary condition for the diffusion Eqn (8.57). Usually the density of adatoms is low, $c_{eq}\Omega_2 \ll 1$, and the movement of steps is slower than relaxation of the adatom density. Under such conditions, one may use the static approximation of Eqn (8.57), in which the left-hand side is neglected and

[23] f_{st} is defined as positive in the direction of the outward normal of the step.

[24] As the driving force of solidification, the extra work $\delta\mu = f_{st}\Omega_2$ is written as $\delta\mu = (\partial\mu/\partial c)\delta c = (k_B T/c)\delta c$ because $\mu = k_B T \ln c$ for an adatom. Then, the shift δc corresponds to $f_{st}\Omega_2 c/k_B T$ as in Eqn (8.59).

the solution with a given boundary condition at each instance is used. The diffusion equation is cast into the form

$$\left(\nabla^2 - \frac{1}{x_s^2}\right)(c(\boldsymbol{x}) - c_\infty) = 0, \tag{8.60}$$

where

$$x_s = \sqrt{D_s \tau}$$

is the *surface diffusion length*. It represents the characteristic length scale that an atom entering on a large facet can move within the lifetime τ.

8.5.2.2 Asymmetric Kinetics at the Step

In general, solidification at the step proceeds at a finite rate as on a rough surface. Supersaturation at the step is finite, and the solidification current is

$$j_s = \mathcal{K}_{st} \delta c_{st}, \tag{8.61}$$

from Eqn (8.53). Adatoms may come to the step either from the upper or from the lower side of the step. We may write the solidification current separately for each side of the step (Figure 8.13(A)):

$$j_{s-} = \mathcal{K}_{st}^{-} \delta c_{st}^{-}, \quad j_{s+} = \mathcal{K}_{st}^{+} \delta c_{st}^{+}, \tag{8.62}$$

where j_{s-} (j_{s+}) indicates the solidification current on the left (right) side terrace of the step, and $c_{st}^{\pm} = c(\boldsymbol{x}_{step} \pm 0)$ is the adatom density on each side of the step. The corresponding kinetic coefficients for the upper terrace $\mathcal{K}_{st}^{-}$ and for the lower terrace $\mathcal{K}_{st}^{+}$ may be different

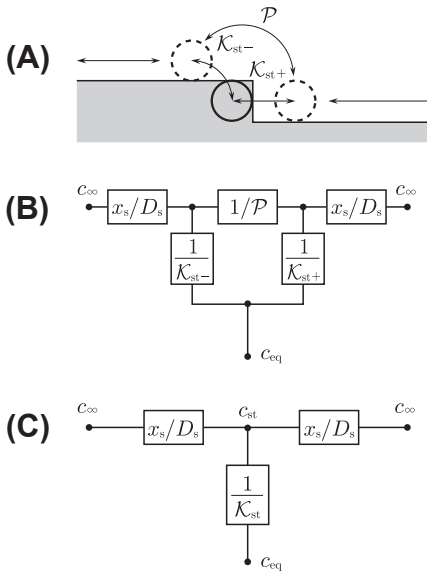


FIGURE 8.13 (A) Transport processes around the step and their resistances. (B, C) Paths of solidification current from far on the terrace, c_∞, to the solid, c_{eq}.

because there is an extra energy barrier for hopping over the step edge: $\mathcal{K}_{st}^{+}$ is usually larger than $\mathcal{K}_{st}^{-}$. The extra energy barrier is called the *Ehrlich-Schwoebel barrier* (ES barrier) and may have significant influence on the morphology of stepped surfaces in growth [40,41]. In addition, if the kink density along the step is very low, adatoms may pass through the step without solidification or melting. Such a step is *permeable* (or transparent).

Figure 8.13(A) and (B) shows schematically the general boundary condition at the step for the diffusion equations, Eqns (8.57) or (8.60). The adatom density at the step site (in the solid), c_{eq}, is related to that at the upper side of the step, c_{st}^{+}, and that at the lower side of the step, c_{st}^{-}, by the resistance $(\mathcal{K}_{st}^{\pm})^{-1}$. The latter two sites are connected by the resistance $\mathcal{P}^{-1}$ ($\mathcal{P}$: permeability of the step), and also with the environment away from the step, c_∞, by the resistance of the surface diffusion on the terrace x_s/D_s. Like an electric circuit, the current between these positions are related to satisfy the conservation law:

$$j_{-} = j_{s-} + j_{p}, \quad j_{+} = -j_{s+} + j_{p}, \tag{8.63}$$

where j_{-}, and j_{+} are the diffusion current on both terraces, and j_p is the current that passes through the step. The step velocity is given by the solidification current from both terraces as

$$V_{st} = \Omega_2 (j_{s-} + j_{s+}) = \Omega_2 (j_{-} - j_{+}), \tag{8.64}$$

which is also expressed by the difference of the diffusion currents. These currents are proportional to the differences of the corresponding adatom densities (Eqn (8.62); equivalently, proportional to the chemical potentials) and [42]

$$j_p = \mathcal{P}(c(\boldsymbol{x}_{st-}) - c(\boldsymbol{x}_{st+})). \tag{8.65}$$

8.5.2.3 A Simple Example of Growth Kinetics

As an example for the growth of a vicinal face, let us consider the simple BCF model—that is, the case with $\mathcal{K}_{st+} = \mathcal{K}_{st-}$, $\mathcal{P} = \infty$ (Figure 8.13(C)). Because $c(\boldsymbol{x}_{st+}) = c(\boldsymbol{x}_{st-}) = c_{st}$, the solution of the diffusion Eqn (8.60) for an array of equidistant steps at $y = 0$ and $y = l$ is Figure 8.14

$$c(y) = \frac{e^{-y/x_s} + e^{(y-l)/x_s}}{1 + e^{-l/x_s}} (c_{st} - c_\infty) + c_\infty \tag{8.66}$$

for $0 < y < l$. From the density profile in Eqn (8.66), the diffusion current that flows into the step at the origin is

$$j_{-} - j_{+} = -2D_s \left.\frac{\partial c}{\partial y}\right|_{y=0} = 2\frac{D_s}{x_s}(c_\infty - c_{st}) \tanh\left(\frac{l}{2x_s}\right), \tag{8.67}$$

which should be the same as the solidification current $j_{s-} + j_{s+}$ obtained from Eqn (8.62). This condition determines the density at the step c_{st}. With c_{st} so determined, the velocity of the step is easily calculated as

$$V_{st} = \Omega_2 \frac{\delta c_\infty}{\frac{1}{\mathcal{K}_{st}} + \frac{1}{\frac{2D_s}{x_s}\tanh\left(\frac{l}{2x_s}\right)}}, \tag{8.68}$$

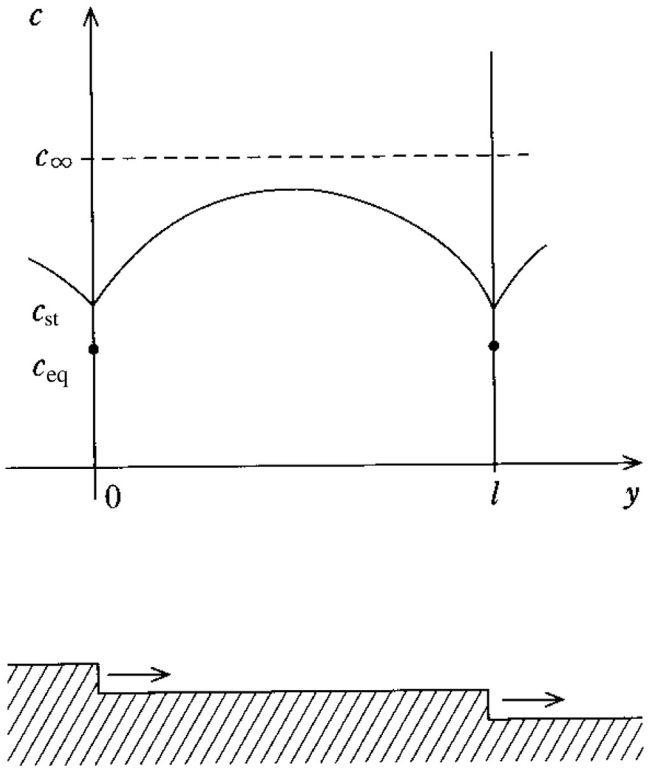


FIGURE 8.14 Adatom density on a terrace of the width l in a growing vicinal face.

where the kinetic coefficient is $\mathcal{K}_{st} = 2\mathcal{K}_{st+} = 2\mathcal{K}_{st-}$. This equation corresponds to the step velocity in Eqn (8.52) in solution growth.

If the step distance l is much larger than the diffusion length ($l \gg x_s$), the step incorporates adatoms within distance x_s of its both sides. From Eqn (8.68) with $\tanh(l/x_s) \to 1$, the effective kinetic coefficient, defined by $\mathcal{K}_{st}^{eff} = V_{st}/(\Omega_2 \delta c_\infty)$,

$$\frac{1}{\mathcal{K}_{st}^{eff}} = \frac{1}{\mathcal{K}_{st}} + \frac{x_s}{2D_s} \tag{8.69}$$

is constant. If the step distance is small, $l \ll x_s$, the territory for a step decreases by a factor l/x_s; the second term increases as

$$\frac{1}{\mathcal{K}_{st}^{eff}} = \frac{1}{\mathcal{K}_{st}} + \frac{x_s^2}{D_s l} = \frac{1}{\mathcal{K}_{st}} + \frac{\tau}{l}. \tag{8.70}$$

If the exchange of atoms at the step is fast (relatively to the diffusion), we may put $\mathcal{K}_{st} \to \infty$ in Eqn (8.68), and the growth velocity of the surface becomes

$$V_z = \frac{a_z}{l}\,\Omega_2 \frac{2D_s}{x_s} \tanh\left(\frac{l}{2x_s}\right)\delta c_\infty . \tag{8.71}$$

For a vicinal face with a very small tilt, $n_{\rm st}x_{\rm s} = x_{\rm s}/l \ll 1$, the velocity is proportional to the step density $n_{\rm st}$:

$$V_z = n_{\rm st} a_z \Omega_2 \frac{2D_{\rm s}}{x_{\rm s}}\, \delta c_\infty \,. \tag{8.72}$$

With a large tilt, $n_{\rm st}x_{\rm s} = x_{\rm s}/l \gg 1$, since $x_{\rm s}^2 = D\tau$, it becomes

$$V_z = a_z \Omega_2 \frac{\delta c_\infty}{\tau} = a_z \Omega_2 \left(f - f_{\rm eq}^0\right) = \upsilon_{\rm S} \delta f. \tag{8.73}$$

The growth velocity is given by the impingement rate that exceeds the evaporation rate. When steps are closely distributed, atoms entering onto the terrace always reach the step, and the growth rate is given by the balance of impingement and evaporation as on the rough surface (Eqn (8.6)).

For general cases, the static diffusion equation,[25] Eqn (8.60), is solved with the boundary conditions for a given configuration of steps. Then, the solution $c(\boldsymbol{x})$ determines the diffusion currents at the step

$$j_\pm = -D_{\rm s} \widehat{\boldsymbol{n}} \cdot \boldsymbol{\nabla} c(\boldsymbol{x}_{\rm st\pm}), \tag{8.74}$$

where $\widehat{\boldsymbol{n}}$ is the unit normal vector of the step pointing downward. The difference of the diffusion current gives the local step velocity from Eqn (8.64), and that determines the change of the step configuration. Thus, the time evolution of the step-adatom system can be calculated (in principle, numerically).

8.5.3 Growth of Vicinal Faces and Morphology of the Stepped Surface

In equilibrium, a vicinal face consists of equidistant parallel steps, whose configuration is stabilized by the stiffness of each step and by repulsion between the steps. For a face with a small tilt $\boldsymbol{p} = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right)$, the free energy per unit projected area $f(p)$ can be written as[26]

$$f(p) = \alpha(0,0) + \frac{\beta}{a_z}\left|p\right| + \frac{\phi}{a_z^3}\left|p^3\right| + \cdots. \tag{8.75}$$

where $\alpha(0, 0)$ is the surface free energy per unit area without steps, and isotropy in the xy-plane has been assumed (therefore, $\tilde{\beta} = \beta$). The second term is the free energy of noninteracting steps, and the third term is the interaction energy, which is proportional

[25] The full time-dependent diffusion in Eqn (8.57) is seldom used because of the difficulty of solving it.
[26] Note the relationship between $\alpha(p)$ and $f(p)$: $\alpha(p) = f(p)/\sqrt{1+p^2}$.

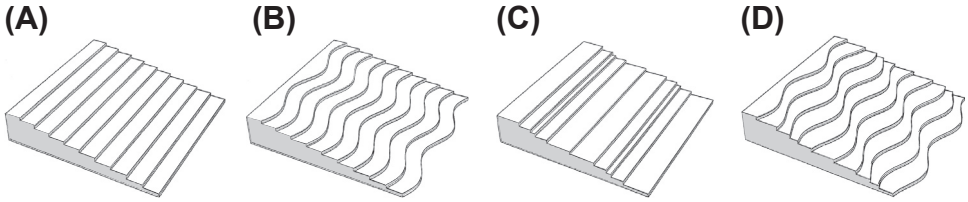


FIGURE 8.15 Deformation of a vicinal face: (A) ideal vicinal face, (B) in-phase wandering of steps, (C) bunching of steps, and (D) wandering with a phase shift. *Reprinted from Ref. [43], with permission from Elsevier.*

to the repulsion energy[27] ϕl^{-2} times the step density $p = a_z/l$. The surface stiffness of the two orthogonal directions are calculated from Eqn (8.75) as follows [4]:

$$\tilde{\alpha}_{\perp}(p) \approx \frac{\beta}{a_z}\frac{1}{p}, \tag{8.76}$$

$$\tilde{\alpha}_{\parallel}(p) \approx 6\frac{\phi}{a_z^3}p. \tag{8.77}$$

Equation (8.76) represents stiffness of the surface against deformation, such as Figure 8.15(B), and is determined by the stiffness of each step. Equation (8.77) represents stiffness against deformation, such as Figure 8.15(C), and is determined by the step repulsion. By approaching the facet, a vicinal surface becomes stiffer in one direction and softer in the orthogonal direction because the number density of steps decreases: the surface deformation becomes more difficult and the step repulsion becomes weaker.

When a vicinal surface is growing or sublimating, surface diffusion tends to produce deformation of the surface against the stabilizing effect of the surface stiffness. If the destabilizing effect wins, the vicinal face becomes unstable and sinusoidal deformation develops [45–47].

8.5.3.1 Wandering of Steps

If step kinetics is not symmetric (i.e., kinetic coefficients for the upper and the lower terraces differ), a step may become unstable. If $\mathcal{K}_{\mathrm{st}+} > \mathcal{K}_{\mathrm{st}-}$, the step incorporates more atoms from the lower terrace than from the upper terrace during growth. If a small bump is formed along the straight step as a result of fluctuation, the bump has an advantage for further growth because it has moved into an area of higher adatom density. The bump grows faster when the step stiffness is not strong enough to pull it back. The initially straight step becomes wavy, and the instability is called *step wandering* (or meandering). This is a kind of two-dimensional Mullin-Sekerka instability (see Chapter 14 of Volume IB) [48]. The stability of the step is characterized by the amplification rate of sinusoidal step fluctuation of wave number q, $\delta y_q(t) = \delta y_0 e^{-iqx+\omega_q t}$. It is given by[28] [49,50]

[27]Two types of repulsive interactions exist rather generally: the elastic repulsion originated from elastic deformation near the steps [40] mentioned before, and an entropic repulsion due to thermal wandering fluctuation of steps with the coefficient $\phi = (\pi^2/6)((k_{\mathrm{B}}T)^2/\beta)$ [44].

[28]The growth rate of fluctuation, ω_q, given here is the formula for $l \to \infty$.

$$\begin{aligned}\omega_q &= -D_s\Omega_2\left[\frac{c_\infty - c_{eq}^0}{x_s^2} - \Lambda_q\left(\frac{c_\infty - c_{eq}^0}{x_s} - \Gamma q^2\right)\right] \\ &= V_{st}^0(\Lambda_q - x_s^{-1}) - D_s\Omega_2\Gamma\Lambda_q q^2,\end{aligned} \tag{8.78}$$

where V_{st}^0 is the velocity of the straight step, $\Gamma = \Omega_2 c_{eq}^0 \tilde{\beta}/k_B T$ and $\Lambda_q \equiv \sqrt{q^2 + x_s^{-2}}$. The coefficient of q^2 in the series expansion of Eqn (8.78) becomes positive when $V_{st}^0 \geq 2D_s\Omega_2\Gamma/x_s^2$,[29] and long wavelength fluctuations grow. When evaporation is negligible, $x_s \to \infty$ $\Lambda_q \to q$, and the amplification rate becomes

$$\omega_q = V_{st}^0 q - D_s\Omega_2\Gamma q^3. \tag{8.79}$$

Now, the step is always unstable under growth ($V_{st}^0 > 0$). The fastest-growing mode is given by the largest ω_q in Eqn (8.79), with the wave number

$$q_{max} = \sqrt{\frac{V_{st}^0}{3D_s\Omega_2\Gamma}}. \tag{8.80}$$

The characteristic wavelength $\lambda_{max} = 2\pi/q_{max}$ is inversely proportional to the square root of the step velocity. Once the instability sets in, the pattern of the step is controlled by the nonlinearity of the system. The time evolution of the destabilized step with evaporation is found to be chaotic [49,50]: bumps of the wavy pattern of the wavelength λ_{max} arise, move, and annihilate randomly. In a vicinal face, the motion of neighboring steps is correlated, and in-phase step wandering (Figure 8.15(B)) is observed [51]. Troughs perpendicular to the steps appear.

The cause of the wandering instability is the front and back asymmetry in step kinetics resulting from the step edge energy barrier. Other such asymmetry may induce a similar instability. When an Si(111) surface is heated by direct electric current perpendicular to the step, wandering instability is observed [46]. The current produces drift of adatoms in the same direction as the current. Theoretically, the drift of adatoms modifies the effective diffusion length in the upper terrace and that in the lower terrace differently, and the instability may occur when the drift direction is opposite to the step motion [52]. Also, in the growth of Si(111), wandering instability is seen near the structural phase transition from the low-temperature 7×7 surface structure to the high-temperature 1×1 structure [53]. The phase transition starts at the step; near the transition temperature, the structures of the upper terrace and the lower terrace of a step are different, which makes strong asymmetry to induce step wandering.

[29] In terms of the impingement rate, this condition is $f \geq f_c \equiv f_{eq}^0(1 + (2\tilde{\beta}\Omega_2)/(x_s k_B T))$.

8.5.3.2 Bunching of Steps

In equilibrium, the distance between parallel steps is kept equal by the step repulsion. If an ES barrier is present (suppose $\mathcal{K}_{st+} > \mathcal{K}_{st-}$), equidistant steps become unstable during sublimation. When a step recedes slightly faster than other steps by fluctuation, its lower side terrace is wider than that of other steps, so it recedes even faster. If this effect is stronger than the step repulsion that stabilizes the equidistant steps, the steps tend to bunch and the vicinal face deforms like Figure 8.15(C). The instability is called *step bunching*. If the step repulsion is short-range or very weak, pairing of steps will occur [54]. With the existence of the power law repulsive interaction, bunching instability occurs at a long wavelength; that is, a density wave of steps is seen (Figure 8.16(A)).

Bunching of steps occurs with various causes. In Figure 8.17, several examples are shown. The velocity of a step is, in many cases, a function of the widths of its upper and lower terraces, and the relationship determines the stability of the vicinal face [8]. If the contribution of the lower terrace to the step velocity is dominant, the face is unstable during growth. The impurity effect shown in Figure 8.17(A), in addition to the ES barrier effect (Figures 8.17(B) and (C)), is an example. A considerable amount of impurities exist in many systems, and they accumulate on terraces during growth to reduce the velocity of steps. The density of impurities on the lower terrace is proportional to the exposure time of the surface, which is proportional to the width of the terrace. Thus, impurity accumulation causes instability of equidistant steps as shown in Figure 8.17(A) and the formation of step bunches [56,57]. An external field may play a similar role to induce step bunching. In Si(111) vicinal faces, direct electric current perpendicular to the steps for heating the crystal induces step bunching [46,58,59] (Figure 8.17(D)). In solution growth,

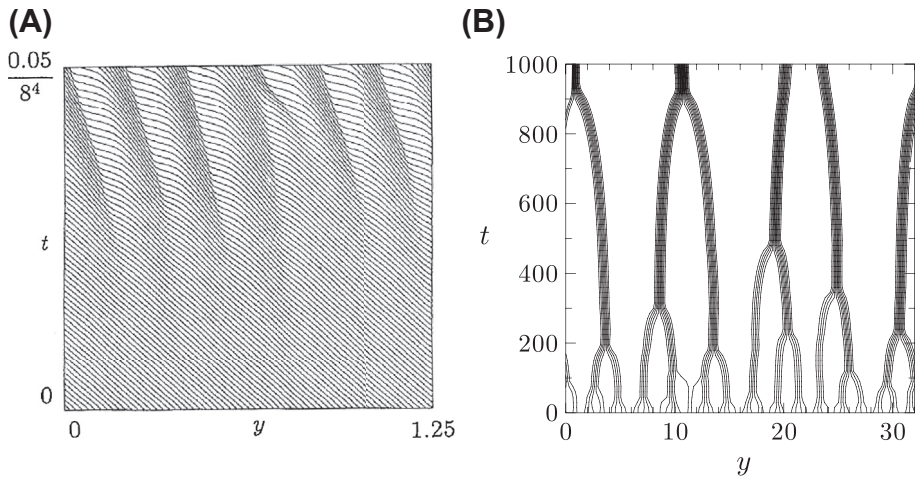


FIGURE 8.16 Evolution of step bunching (A) in sublimation and (B) with drift of adatoms. (A) Step bunching develops as a density wave of steps. [32]. (B) Hierarchical pairing results in large bunches. *Reprinted from Ref. [53], with permission from AIP; and Ref. [55], with permission from Elsevier.*

like the drift of adatoms, external flow of solution induces step bunching in growth, as shown in Figures 8.17(E) and (F) [60,61]. In this case, the velocity of a step is not locally determined but is influenced by more distant conditions: growth/dissolution of a part of the vicinal face is correlated with growth/dissolution of an upstream area.

8.6 Crystal Growth in a Diffusion Field

As discussed in the previous section, the growth of a vicinal face is possible through the transport of atoms from the environment by bulk and/or surface diffusion followed by incorporation of atoms at the steps. The series of transport processes is a general feature of crystal growth. The flow of matter is controlled by the difference (or spatial gradient) of chemical potential. The difference of chemical potential between the solid and the environment, $\delta\mu_\infty$, is consumed by each step of the transport: several processes work in series (or sometimes in parallel) as resistances in an electric circuit (as in Figures 8.13(B) and (C)). The diffusion current at the interface must be the same as the solidification current. These conditions give relationships that determine the growth velocity.

Also, in general, heat released in solidification needs to be evacuated to the environment. Heat transport should work in parallel with the transport of matter. Usually in melt growth, heat transport is the rate-limiting process; in solution growth, material transport is the rate-limiting process. We may neglect the faster process—material transport in melt growth and heat transport in solution growth, respectively.

8.6.1 Melt Growth and Solution Growth

For simplicity, we neglect convection in the following discussion. Then, we can formulate the two problems—evacuation of the latent heat by heat diffusion (synonym of heat conduction) in melt growth and supply of matter by diffusion in solution growth—in a similar way.

8.6.1.1 Evacuation of Latent Heat in Melt Growth

In growth from the melt, the growth velocity is proportional to supercooling at the interface:

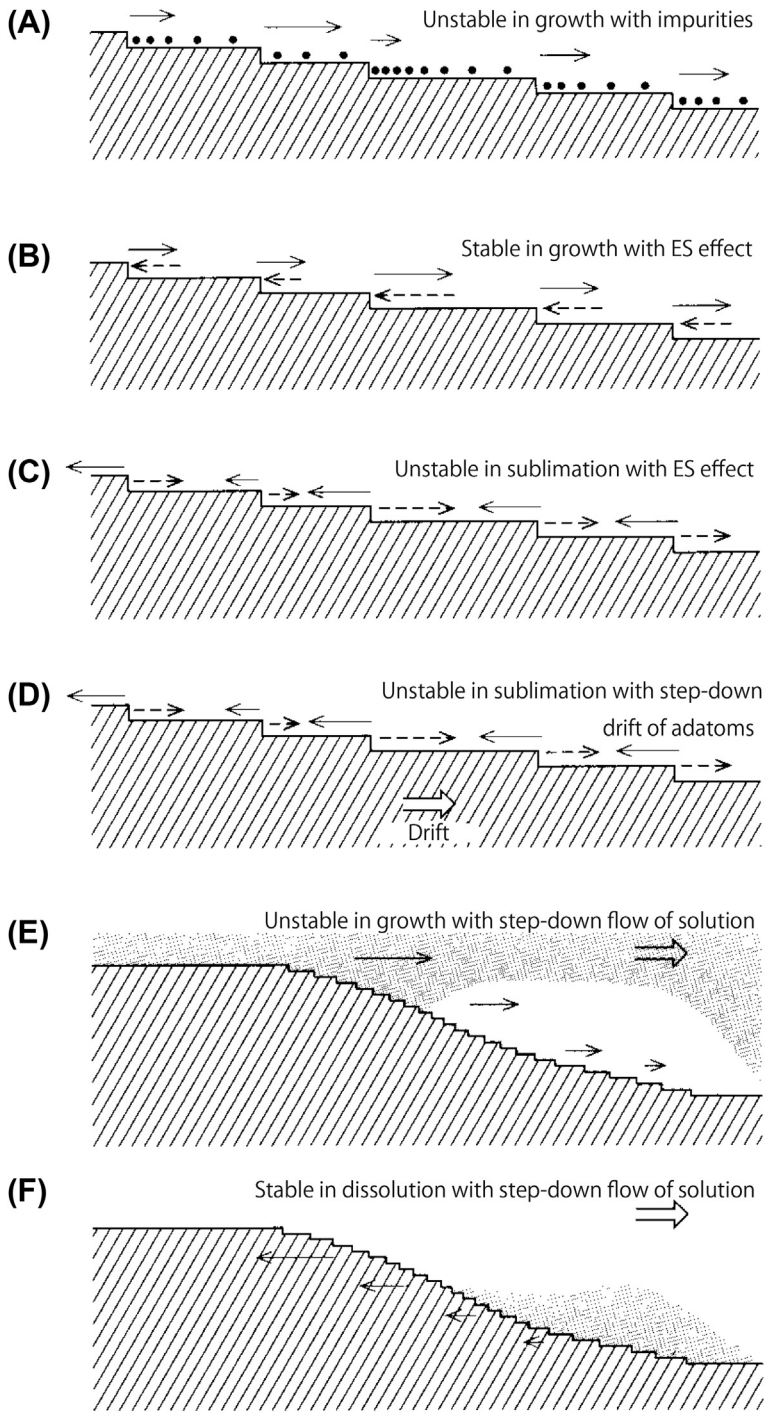
$$V = K_T(T_\mathrm{m} - T_\mathrm{i}), \tag{8.81}$$

where T_m is the melting temperature (the equilibrium temperature), T_i is the temperature at the interface,[30] and K_T is a kinetic coefficient. Because the latent heat (L per unit volume of a solid) should be evacuated from the interface by heat diffusion,

$$LV = -\kappa_T \hat{\boldsymbol{n}} \cdot \nabla T \tag{8.82}$$

must hold. Here, κ is heat conductivity in the melt, and we have assumed that the heat is evacuated into the bulk liquid phase. In general, heat goes to both phases depending on

[30]Here, we assume a solid is the low-temperature phase.


(A)
Unstable in growth with impurities
(B)
Stable in growth with ES effect
(C)
Unstable in sublimation with ES effect
(D)
Unstable in sublimation with step-down
drift of adatoms
Drift
(E)
Unstable in growth with step-down flow of solution
(F)
Stable in dissolution with step-down flow of solution

boundary conditions. The simple model here is called a one-sided model. The heat diffusion is described by the diffusion equation

$$\frac{\partial T}{\partial t} = D_T \nabla^2 T, \tag{8.83}$$

where the thermal diffusion coefficient (diffusivity) D_T is related to the thermal conductivity κ_T and the heat capacity per unit volume c_P as $D_T = \kappa_T / c_P$.

It is convenient to use a dimensionless supercooling, in which temperature increase from the far environment is measured by the temperature increase by the latent heat, L/c_P,

$$u = \frac{T - T_\infty}{L/c_P}. \tag{8.84}$$

With the dimensionless supercooling u, Eqns (8.83) and (8.82) are cast into the simple forms

$$\frac{\partial u}{\partial t} = D_T \nabla^2 u, \tag{8.85}$$

$$V = -D_T \hat{\boldsymbol{n}} \cdot \boldsymbol{\nabla} u. \tag{8.86}$$

The other expression of the growth velocity, Eqn (8.81), with the Gibbs–Thomson effect, is written as

$$V = \tilde{K}_T \left(\Delta - u_\mathrm{i} - d_T \kappa \right), \tag{8.87}$$

where $\tilde{K}_T = KL/c_P$, and

$$\Delta = u_\mathrm{m}^0 = \frac{\left(T_\mathrm{m}^0 - T_\infty \right) c_P}{L} \tag{8.88}$$

is the value of u for a flat equilibrium interface (i.e., the dimensionless supercooling of the environment). In Eqn (8.87), $\kappa = 1/R_1 + 1/R_2$ is the curvature of the interface and d_T is the *capillary length,* which expresses the strength of the surface stiffness $\tilde{\alpha}$

$$d_T = \frac{\tilde{\alpha} c_P T_\mathrm{m}^0}{L^2}. \tag{8.89}$$

The capillary length is a length scale at which temperature increase by the latent heat, L/c_P, and the change of the equilibrium temperature due to the Gibbs–Thomson effect, $(\alpha/d)(T_\mathrm{m}^0/L)$, become comparable.

FIGURE 8.17 Various causes of step bunching. (A) Accumulation of impurities on terraces decreases the velocity of steps. The step behind the wider terrace moves slower, and the step behind the narrower terrace moves faster: a step pair is formed. (B) With Schwoebel barrier in growth, the retarded step has a wide lower terrace and moves faster to recover the delay. (C) With Schwoebel barrier in sublimation, the advanced step has a wide lower terrace and moves faster to amplify the advantage. (D) With step-down drift in sublimation, the advanced step has a wide lower terrace and moves faster to amplify the advantage. (E) With step-down flow in growth, the diluted part of solution moves to the lower area and hiders growth of the lower steps: the slope becomes steeper. (F) With step-down flow in dissolution, the concentrated part of solution moves to the lower area and decelerates dissolution of the lower steps: the slope becomes gentler.